

# Construction of a Model for Water Purification Mechanisms in a River by Using a Neural Network Approach

Tomoo AOYAMA<sup>a</sup>, Junko KAMBE<sup>b,c</sup>, Aiko YAMAUCHI<sup>d</sup> and Umpei NAGASHIMA<sup>c,e,\*</sup>

<sup>a</sup>Faculty of Technology, Miyazaki University  
Gakuen Kihanadai Nishi, Miyazaki, 889-2192, Japan,

<sup>b</sup>Faculty of Foreign Language, Daito Bunka University  
1-9-1 Takashimadaita, Itabashi, Tokyo 175-8571, Japan

<sup>c</sup>Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST)  
Kawaguchi Center Building, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012 Japan

<sup>d</sup>Graduate School of Pharmaceutical Sciences, University of Tokushima  
1-78 Sho-machi, Tokushima 770-8505, Japan

<sup>e</sup>Research Institute of Computational Science, National Institute of Advanced Industrial Science and Technology  
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

\*e-mail: [u.nagashima@aist.go.jp](mailto:u.nagashima@aist.go.jp)

(Received: November 9, 2006; Accepted for publication: February 27, 2007; Published on Web: May 15, 2007)

A model for purification mechanisms in a river was proposed to express changes in the values of BOD (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), T-N (Total Nitrogen), and T-P (Total Phosphorous) as the combination of inflows, streams, and weirs. Because expressions for these functions are unknown, the model first constructs multi-layer neural-network functions based on observations, and then uses the derivatives to evaluate the cause and effect of pollution.

For data of the Tamagawa through Tokyo, Japan in 2002, the model suggested that the cause of pollution is inflows from sewage, the stream of the Tamagawa has purification functions for COD and T-P, but has little ability for T-N, and the weirs have purification for COD, but have no purification for T-N and T-P. The results from the model were consistent with common sense for the water quality, and thus, there was no failure in the model.

**Keywords:** River Model, Water Purification, Neural Network, Derivatives

## 1 Introduction

Sustaining the natural environment is a critical problem, especially water quality of rivers [1] because they are our major source of water for everyday life. Rivers worldwide are currently polluted, due to human waste being dumped into them. Despite attempts to mitigate such pollution, such as constructing sewers and waste treatment plants, discharge water was at first considered pure, but is now known to be polluted as well [2].

We did an extensive search of published observation data for expressions that can be used as indicators of changes in water quality in rivers. Each observation of a water quality index, however, is represented by a vector, and therefore discrete, thus making it difficult to extract information directly from the vector itself. If a conser-

vation function could be fitted to the vector, the derivative can be obtained, which indicates a change in water quality, which in turn represents a mechanism in the river that purifies the water. The approach was derived and tested for ideal data generated using uniform random numbers [3]. In this approach, there is no observation error and no abnormal value, the water purification descriptors are inflows, streams, and weirs, and neither infiltration nor transpiration is considered. In a previous study [4], we proposed this approach; however, the discussion was based on an idealized situation. The approach should be tested for a river based on observations. In the study, we used a multi-layer neural network to emulate the purification mechanisms. The learning is back-propagation learning with “erasing” of connections among neurons [5]. This is a simulated annealing method to avoid local

minima [6] and to eliminate inactive connections between output neurons and invalid input neurons. This type of learning is called “reconstruction learning.”

## 2 Indexes for water quality

There are 10 to 20 observation points along each river in Japan [2]. BOD (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), T-N (Total Nitrogen), and T-P (Total Phosphorous) are among the measured indexes at each of these observation points. These indexes indicate different aspects of the quality of sewage processing on the order of parts per million (ppm).

BOD indicates the level of decomposition of organic compounds by microbes. Different types of BOD indexes are used throughout the world. In Japan, C-BOD5 is used [7], where C indicates carbon and 5 indicates 5-day fermentation. Current BOD levels indicate that sewage in the Tamagawa is almost as clean as “natural” river branches, i.e., 2-3 [mg/L] [2]. This is acceptable pollution for human beings. COD is the amount of organic compounds oxidized by  $\text{MnO}_4^-$  ion or  $\text{Cr}_2\text{O}_7^{2-}$  ions under acidic conditions. The difference between BOD and COD indicates the substances not yet decomposed by microbes. Current COD levels show that sewage is often about 10 times more polluted than natural river water. Current T-N and T-P levels also indicate pollution. Despite advanced processing in sewage plants, however, sewage water is not as clean as natural river water.

At the [ppm] level of pollution, we concluded that changes in BOD, COD, T-N, and T-P would be sufficient to evaluate the purification mechanisms functions of a river. We have not yet developed an approach for the [ppb] level. If rivers can decompose these contaminants at the [ppm] level, there is a less severe pollution problem. If they cannot, then the remaining sewage is discharged directly into the sea, thus the entire water ecological system will be affected.

## 3 Expression of river purification mechanisms

### 3.1 River representation

To express changes in water quality, we define a river purification mechanism model as,

$$S_i = S_{i-1} + f_i p_i + K d_i + L h_i, \quad (1)$$

where  $S_i$  is the amount of pollution substances at the  $i$ -th observation point  $X_i$ , subscript “ $i-1$ ” indicates upstream for “ $i$ ”,  $d_i$  is the distance between  $X_{i-1}$  and  $X_i$ ,  $f_i p_i$  is the product of the flow and the substance concentration of a branch plus sewage that is discharged between  $X_{i-1}$  and  $X_i$  (if there are multiple branches or sewage, then  $f_i p_i$  is

the sum of the products),  $K$  and  $L$  are constants of the stream- and the weir-purification functions, respectively, and  $h_i$  is a dummy variable to express the existence of weirs only. A weir is a small dam used in agriculture in Japan, and generates stagnant water where organic substances might be decomposed.

If numerous observations are available to determine the characteristics of the stream between  $X_{i-1}$  and  $X_i$ , then  $K$  can be written as  $\{K_i\}$ . However, when the number of observation points is between 10 and 20, then there is insufficient information to evaluate  $\{K_i\}$ . Under such conditions, an approximation to reduce  $\{K_i\}$  to  $K$  is needed. The approximation is valid when the river has the same status as the region of interest. The status is based on the existence of weirs, rapids, flooded fields, and vegetation along the shores.  $K$  is expressed in units of [mg/L]/[distance]. If the river has purification mechanisms, then  $K$  is negative. Ideally,  $L$  should equal  $\{L_i\}$ . However, similar to  $\{K_i\}$ , determining the characteristics of weirs is difficult. An approximation to reduce  $\{L_i\}$  is therefore needed.

In the model, transpiration and infiltration are neglected. In temperate zones, transpiration is negligible. Although infiltration is unknown, it is considered negligible in the model because there are clay layers in the Tamagawa basin.

### 3.2 Representation of purification mechanisms

The  $\{S_{i-1}\}$ ,  $\{f_i p_i\}$ ,  $\{d_i\}$  and  $\{h_i\}$  are used as the causes, and the  $\{S_i\}$  is used as the effects. We can learn a multi-layer neural network based on these parameters. The neural network is a consecutive function. Thus, a continuous function can be derived from discrete observations of 10-20 points. The function is differential. The purification mechanisms of a river can be determined by using the derivatives [8]. The amplitude of the mechanism is an average because of the approximation for  $K$ .

### 3.3 Reconstruction learning for neural networks

The reconstruction learning for the multi-layer neural networks used in the model was set up as follows. For the first layer, the number of neurons was the number of descriptors plus a bias. For the second layer, the number was initially the number of descriptors, but was then decreased by optimization during erasing operations in the learning. A bias neuron was set in the second layer and was fixed. The neuron emulation function in the second layer was the sigmoid function. In the third layer, there was one neuron, and the emulation function was a linear function.

Thus, the function was a linear combination of the sigmoid collectively. The learning constants ranged from

0.1 to 0.2 for the second- and third-layer neurons. The erasing operations for the reconstruction were 0.04/50-iterations for the back-propagation learning (BP). The learning was continued until the BP-error reached 0.03.

We evaluated the precision of the neural networks by using model data [4]. The results revealed a 2% precision for both  $K$  and  $L$  in Eq. (1).

## 4 Characteristics of the Tamagawa

### 4.1 Observation points and the flows

There are 15 observation points in the Tamagawa (Table 1), where water quality indexes and in/out-flows are observed. The data set for 2002 is complete and is the latest data [2, 10]. However, some data are defective; namely, the inflows at Hamura and the brackish water area near the river mouth. The inflows at Hamura are important in the analysis of the Tamagawa. The reported diversion ratio is 78-80% of the inflow at the Hamura dam site, and therefore all inflows at observation points downstream of Hamura are affected. Observed and calculated values showed good agreement with each other.

### 4.2 Sewage population ratio

The pollution of rivers in Japan apparently peaked in the 1970s. To decrease this pollution, more sewage plants were constructed. In 1999, an average of 62% of the population in Japan was serviced by a sewer system [9]. In metropolitan areas, this level is 95-100% [10]. Due to

such countermeasures, rivers appear to be regaining their natural beauty. Water indexes for nitrogen and phosphorous contamination reveal, however, that pollution is still a problem [11].

The source of river pollution today is sewage [10]. Chemical reactions and biological metabolic mechanisms have been developed to decompose sewage, and thus rivers are not watercourses but rather decomposition sites via rapids, stagnant, weirs, and microbes in the water or on the surface of gravel.

The Tamagawa is a small river that is 138 km long and flows at 20 [m<sup>3</sup>/s] in the Tokyo metropolitan area. The population in the river basin is 4.25 million. The river has been a water resource for the past 300 years. Today, 78-80% of the water transported from the river into the water supply for the surrounding population occurs at the Kosaku and the Hamura Dams located upstream. Also, 40% of the water flow in the middle and lower streams is sewage [10]. In the Tamagawa basin, there are 6 sewage plants (Table 2), which use state-of-the-art technology in disposal water processing. With respect to the BOD index, the resultant water quality in the Tamagawa is the same as that for natural water. However, with respect to T-N and T-P indexes, the water quality is low [10]. Therefore, estimating whether the Tamagawa can decompose nitrogen and phosphorous compounds dumped into it is crucial. Unlike European rivers, there is no flood plain in the Tamagawa; therefore, nitrogen and phosphorous flow directly into the sea. The purification ability of the sea is not infinite, and one nitrogen compound, namely, ammonia, is poisonous to fish.

Table 1. Observation points, distance from river mouth, inflow, and outflow.

| #  | Observation points         | Distance [km] | Inflow [m <sup>3</sup> /s] |             |            | Outflow [m <sup>3</sup> /s] |             |
|----|----------------------------|---------------|----------------------------|-------------|------------|-----------------------------|-------------|
|    |                            |               | Main stream                | Sewage      | Branches   | Calculated                  | Observed    |
| 0  | Wada Bridge                | 61.8          | 15                         | none        | none       | 15                          | 15          |
| 1  | Chofu Bridge               | 59.8          | 13.6                       | none        | none       | 13.6                        | 13.6        |
| 2  | Hamura Dam                 | 53.2          | 14                         | none        | none       | 3                           | unpublished |
| 3  | Nagata Bridge              | 51.7          | 3                          | none        | none       | 3                           | 3.84        |
| 4  | Haijima-Bara Reservoir     | 47.8          | 3                          | none        | 0.70, 3.65 | 7.35                        | unpublished |
| 5  | Haijima Bridge             | 46.2          | 7.35                       | none        | none       | 7.35                        | 5.82        |
| 6  | Hino Bridge                | 39.8          | 7.35                       | 1.62, 0.81* | 0.33, 0.24 | 10.35                       | 9.54        |
| 7  | Sekito Bridge              | 34.6          | 10.35                      | 0.58, 0.69  | 3.13       | 14.75                       | 13.7        |
| 8  | Koremasa Bridge            | 31.5          | 14.75                      | 1.16        | 0.56       | 16.47                       | 15.9        |
| 9  | TamaGawara Bridge          | 27.9          | 16.48                      | 2.31        | none       | 18.79                       | 17.5        |
| 10 | TamaSuido Bridge           | 23.2          | 18.79                      | none        | 0.29       | 19.08                       | 18.9        |
| 11 | 3rd Keihin Tamagawa Bridge | 16.5          | 19.08                      | none        | 0.48       | 19.56                       | 20          |
| 12 | DenenChufu Reservoir       | 13.1          | 19.56                      | none        | none       | 19.56                       | 23.4        |
| 13 | Rokugo Bridge              | 5.5           | 19.56                      | none        | none       | 19.56                       | unpublished |
| 14 | Daishi Bridge              | 2.4           | 19.56                      | none        | none       | 19.56                       | unpublished |

At Rokugo and Daishi Bridges, the volume is not published because of the brackish area.

\*The two inflows represent those for two plan

Table 2. Water quality of branches and sewage plants

| Source                   | Distance [km] | BOD [mg/L] | COD [mg/L] | T-N [mg/L] | T-P [mg/L] |
|--------------------------|---------------|------------|------------|------------|------------|
| Hirai River              | 50            | 0.54       | 1.48       | 3.42       | 0.02       |
| Aki River                | 49            | 0.5        | 1.15       | 1.42       | 0.015      |
| *Tama upper stream Plant | 43.7          | 3          | 10         | 12.0       | 0.9        |
| *Hachi'Oji Plant         | 42            | 1          | 9          | 15.2       | 1.6        |
| Yachi River              | 42            | 3.21       | 4.63       | 5.29       | 0.21       |
| Zanbori River            | 40.4          | 0.62       | 1.68       | 2.88       | 0.07       |
| *North Tama 2nd Plant    | 37            | 2          | 8          | 11.8       | 1.2        |
| *Asakawa Plant           | 37            | 1          | 9          | 10.8       | 0.9        |
| Asa River                | 37            | 1.56       | 3.65       | 4.97       | 0.32       |
| Ohguri River             | 34            | 1.17       | 3.37       | 2.13       | 0.08       |
| *South Tama Plant        | 32            | 1          | 9          | 11.1       | 1.0        |
| *North Tama 1st Plant    | 29.2          | 3          | 10         | 9.6        | 1.5        |
| Misawa River             | 26            | 3.62       | 6.3        | 4.16       | 0.44       |
| Noh River                | 17.8          | 2.23       | 5.28       | 7.54       | 0.14       |

\* indicates a sewage plant.

Table 3. Water quality of main stream and weirs.

| #  | Observation point          | BOD [mg/L] | COD [mg/L] | T-N [mg/L] | T-P [mg/L] | Weirs             |
|----|----------------------------|------------|------------|------------|------------|-------------------|
| 0  | Wada Bridge                | 0.5        | 1.21       | 0.75       | 0.015      |                   |
| 1  | Chofu Bridge               | 0.85       | 2.0        | 0.83       | 0.018      |                   |
| 2  | Hamura Dam                 | 0.5        | 1.2        | 0.78       | 0.015      | Hamura/Kosaku     |
| 3  | Nagata Bridge              | 0.77       | 1.8        | 0.93       | 0.018      |                   |
| 4  | Haijima-Bara Reservoir     | 0.53       | 1.2        | 1.5        | 0.016      |                   |
| 5  | Haijima Bridge             | 0.84       | 1.8        | 1.5        | 0.024      | Syowa             |
| 6  | Hino Bridge                | 2.05       | 4.3        | 4.7        | 0.38       | Hino              |
| 7  | Sekito Bridge              | 1.42       | 3.7        | 5.2        | 0.4        | Yotsuya Motosyuku |
| 8  | Koremasa Bridge            | 1.59       | 4.3        | 5.8        | 0.47       | Daimaru           |
| 9  | TamaGawara Bridge          | 2.35       | 5.6        | 6.7        | 0.61       |                   |
| 10 | TamaSuido Bridge           | 1.32       | 4.4        | 5.9        | 0.43       | Kamigawara        |
| 11 | 3rd Keihin Tamagawa Bridge | 1.14       | 3.5        | 5.99       | 0.36       | Syukugawara       |
| 12 | DenenChufu Reservoir       | 1.02       | 3.99       | 5.67       | 0.38       |                   |
| 13 | Rokugo Bridge              | 1.46       | 4.4        | 5.0        | 0.31       |                   |
| 14 | Daishi Bridge              | 1.39       | 4.4        | 4.2        | 0.26       |                   |

There are 8 weirs in the middle stream of the Tamagawa (Table 3). These weirs introduce agricultural water into inundation canals. Based on aerial photographs, all of the weirs have similar storage capacity. The water stagnates at weirs, where nitrogen and phosphorous compounds might be decomposed. Although this hypothesis has not yet been proven, research on the weir function is needed because decomposition of these compounds is urgently needed.

### 4.3 Water quality of branches and sewage

Table 2 lists the published water quality of the branches and sewage plants along the Tamagawa [2]. The COD, T-N, and T-P indexes for the sewage plants were about 10 times higher than those of the branches. Contaminants

from these sewage plants are discharged into the middle stream, which is a zone that ranges from 29 to 44 km from the river mouth.

Table 3 lists the measured water quality of the main stream of the Tamagawa. The water of the main stream was less contaminated than that of the sewage plants (Table 2).

To determine the cause for such a large difference in water quality, we calculated the BOD at each observation point (Table 4). In Table 4, Sums 1 and 2 are the BOD for the inflow and outflow, respectively, at the respective observation point,  $\Delta$  is the difference between Sums 1 and 2 divided by the maximum value of Sum 2, and P is a qualitative parameter indicating pollution ( $\times$ ) or purification ( $\circ$ ).

Table 4. BOD sum at observation points in [(mg/L)×(m<sup>3</sup>/s)]

| #  | Observation point          | Sewage water |      | Branches |      | Sum 1 | Sum 2 | Δ     | P | Note                        |
|----|----------------------------|--------------|------|----------|------|-------|-------|-------|---|-----------------------------|
| 1  | Chofu Bridge               |              |      |          |      | 6.8   | 11.5  | 0.11  |   |                             |
| 2  | Hamura Dam                 |              |      |          |      | 11.9  | 7.0   | -0.11 |   | Hamura/Kosaku weirs         |
| 3  | Nagata Bridge              |              |      |          |      | 1.5   | 2.3   | 0.02  |   |                             |
| 4  | Haijima-Bara Reservoir     |              |      | 0.38     | 1.83 | 4.5   | 3.9   | -0.01 |   |                             |
| 5  | Haijima Bridge             |              |      |          |      | 3.9   | 6.2   | 0.05  |   | Syowa weir                  |
| 6  | Hino Bridge                | 4.86         | 0.81 | 1.06     | 0.15 | 13.1  | 21.2  | 0.19  | × | Hino weir                   |
| 7  | Sekito Bridge              | 1.16         | 0.69 | 4.88     |      | 27.9  | 20.9  | -0.16 | ○ | Yotsuya Moto-syuku weir     |
| 8  | Koremasa Bridge            | 1.16         |      | 0.66     |      | 22.8  | 26.2  | 0.08  |   | Daimaru weir                |
| 9  | TamaGawara Bridge          | 6.93         |      |          |      | 33.1  | 44.1  | 0.25  | × |                             |
| 10 | TamaSuido Bridge           |              |      | 1.05     |      | 45.2  | 25.2  | -0.45 | ○ | Futagaryo Kami-gawara weir  |
| 11 | 3rd Keihin Tamagawa Bridge |              |      | 1.07     |      | 26.2  | 22.3  | -0.09 |   | Futagaryo Syuku-gawara weir |
| 12 | DenenChufu Reservoir       |              |      |          |      | 22.3  | 19.9  | -0.05 |   |                             |
| 13 | Rokugo Bridge              |              |      |          |      | 19.9  | 28.5  | 0.20  | × | Brackish water              |
| 14 | Daishi Bridge              |              |      |          |      | 28.5  | 27.2  | -0.03 |   | Brackish water              |

“×” and “○” in P-column respectively show pollution and purification mechanisms for BOD.

Sum 1 is BOD multiplied by inflow, which is the amount of BOD.

Sum 2 is that of outflow. The Δ is calculated by (Sum2-Sum1)/max{Sum2}.

No clear relationship was found between the purification and weirs. However, this might be due to a limitation of a qualitative approach, where stream and weir purifications were not taken into account. We therefore introduced a neural network approach.

## 5 Neural Network Approach

### 5.1 Learning data

First, we discuss the flow. The following iteration was used in the learning process:

$$X_{i-1} + \sum_j Y_{ji} = X_i, \quad (2)$$

where  $i$  is an observation point ( $i=0,1,\dots,N$ ),  $X_i$  is the flow of the main stream at observation point  $i$ ,  $\sum_j Y_{ji}$  is the amount of inflow between observation points  $i-1$  and  $i$ , and  $\sum_j$  is the sum for all branches and sewage plants. Effects of transpiration and infiltration were not included in this relation. Equation (2) is the conservation at observation point  $i$  and corresponds to multiple vectors that express  $\{X_{i-1}\}$ ,  $\{\{Y_{ji}\}_j\}$ , and  $\{X_i\}$ . The  $\{\{Y_{ji}\}_j\}$  is the set that includes  $j$   $\{Y_{ji}\}$  individuals, where  $j$  is defined between  $i$  and  $i-1$ . Therefore, multiple preserved quantities are obtained based on the set of multiple vectors. The  $\{X_{i-1}\}$  and  $\{X_i\}$  are partial sets of  $\{X_i; i=0,1,2,\dots,N-1\}$ , where  $N$  is the number of observation points. The set  $\{X_i; i=0,1,2,\dots,N-1\}$  contains defects. We therefore interpolated these defects by using the iterative relations in

equation (2) and then minimizing the difference between the interpolated values and observed values.

Next, we discuss the contamination in the river. Discrete vectors were obtained as follows. These vectors are a kind of preservation quantity at observation points, and contain information to represent the river function. However, differential relations can not be extracted from these discrete vectors. Therefore, we projected the preservation quantities onto a continuous space. The learning in neural networks is a kind of projection. In this neural network projection, conservation is achieved, and thus the differential relations can be extracted. The amount of a particular water quality index  $B_i$  at the  $i$ -th observation point can be expressed as

$$B_i = X_{i-1}A_{i-1} + \sum_j Y_{ji}Z_j, \quad (3)$$

where  $A_i$  is the concentration of the water index at the  $i$ -th observation point along the main stream, and  $Z_i$  is that of a branch. Eq. (3) is invariable even if transpiration  $\rho$  is found between the  $i$  and  $i-1$  points. That is,

$$B_i - \rho = X_{i-1}A_{i-1} + \sum_j Y_{ji}Z_j - \rho.$$

The  $\rho$  has a bias that is eliminated by using the scaling process in the learning.

If the river is a water-course, then

$$C_i = X_iA_i. \quad (4)$$

Table 5. Learning data for neural network to express Tamagawa purification mechanisms.

| #  | Obs. point        | Distance/weir |       | BOD                          |       | COD                          |       | T-N                          |       | T-P                          |       |
|----|-------------------|---------------|-------|------------------------------|-------|------------------------------|-------|------------------------------|-------|------------------------------|-------|
|    |                   | $d_i$ [km]    | $h_i$ | [(mg/L)×(m <sup>3</sup> /s)] |       | [(mg/L)×(m <sup>3</sup> /s)] |       | [(mg/L)×(m <sup>3</sup> /s)] |       | [(mg/L)×(m <sup>3</sup> /s)] |       |
|    |                   |               |       | $B_i$                        | $C_i$ | $B_i$                        | $C_i$ | $B_i$                        | $C_i$ | $B_i$                        | $C_i$ |
| 1  | Chofu Bridge      | 2.0           | 0     | 6.8                          | 11.5  | 16.5                         | 27.2  | 10.2                         | 11.3  | 0.2                          | 0.2   |
| 2  | Hamura Dam        | 6.6           | 1     | 11.9                         | 7.0   | 28.0                         | 16.8  | 11.6                         | 10.9  | 0.3                          | 0.2   |
| 3  | Nagata Bridge     | 1.5           | 0     | 1.5                          | 2.3   | 3.6                          | 5.4   | 2.3                          | 2.8   | 0.0                          | 0.1   |
| 4  | Haijima-Bara      | 3.9           | 0     | 4.5                          | 3.9   | 10.6                         | 8.8   | 10.4                         | 11.0  | 0.1                          | 0.1   |
| 5  | Haijima Bridge    | 1.6           | 1     | 3.9                          | 6.2   | 8.8                          | 13.2  | 11.0                         | 11.0  | 0.1                          | 0.2   |
| 6  | Hino Bridge       | 6.4           | 1     | 13.1                         | 21.2  | 38.7                         | 44.5  | 45.2                         | 48.6  | 3.0                          | 3.9   |
| 7  | Sekito Bridge     | 5.2           | 1     | 27.9                         | 20.9  | 66.8                         | 54.6  | 78.5                         | 76.7  | 6.3                          | 5.9   |
| 8  | Koremasa Bridge   | 3.1           | 1     | 22.8                         | 26.2  | 66.9                         | 70.8  | 90.8                         | 95.5  | 7.1                          | 7.7   |
| 9  | TamaGawara Bridge | 3.6           | 0     | 33.1                         | 44.1  | 93.9                         | 105.2 | 117.7                        | 125.8 | 11.2                         | 11.4  |
| 10 | TamaSuido Bridge  | 4.7           | 1     | 45.2                         | 25.2  | 107.0                        | 83.9  | 127.0                        | 112.5 | 11.6                         | 8.2   |
| 11 | 3rd Keihin Tama.  | 6.7           | 1     | 26.2                         | 22.3  | 86.4                         | 68.4  | 116.1                        | 117.1 | 8.3                          | 7.0   |
| 12 | DenenChufu        | 3.4           | 0     | 22.3                         | 19.9  | 68.4                         | 78.0  | 117.1                        | 110.8 | 7.0                          | 7.4   |
| 13 | Rokugo Bridge     | 7.6           | 0     | 19.9                         | 28.5  | 78.0                         | 86.0  | 110.8                        | 97.8  | 7.4                          | 6.1   |
| 14 | Daishi Bridge     | 3.1           | 0     | 28.5                         | 27.2  | 86.0                         | 86.0  | 97.8                         | 82.1  | 6.0                          | 5.1   |

Distance is between adjacent observation points. "0/1" in the  $h_i$  (Weir) column indicates that observation point has a weir (1) or not (0); i.e., the dummy variable.  $B_i$  and  $C_i$  are those in Eqs. (3) and (4).

If the river has purification mechanisms, then  $B_i > C_i$ . If not, because there are various reactions in the flow, the following relation is often valid:

$$B_i < C_i. \quad (5)$$

Thus, a river function  $f(\cdot)$  is expressed in the  $\{B_i\}$  and  $\{C_i\}$  vectors; i.e.,

$$\{C_i\} = \{f(B_i)\}. \quad (6)$$

If  $f(\cdot)$  is unknown, it can be estimated by a neural network and descriptor vectors, such as the existence of weirs and the distance between adjacent observation points (i.e., between  $i-1$  and  $i$ ). Table 5 lists the learning data. After the learning is complete, constants  $K$  and  $L$  can be evaluated by using the following equation.

$$K = \frac{\partial f(\text{descriptors})}{\partial d_i}; L = \frac{\partial f(\text{descriptors})}{\partial h_i}, \quad (7)$$

using the values of  $d_i$  and  $h_i$  listed in Table 5.

In Eq. (1), we used a dummy variable to represent the weirs. The dummy variable was a vector of 0/1 elements, was scaled from  $[\varepsilon, 1-\varepsilon]$  to  $\varepsilon=0.05$ , and was used as a consecutive function in the neural network processing. The reliability of this dummy variable was determined, as discussed in section 5.3.

## 5.2 Calculated purification mechanisms

River purification mechanisms for BOD, COD, T-N, and T-P were calculated as follows. Figure 1a, b, c, and d, respectively, show the amplitudes of the derivatives between the respective index and descriptors at each observation point along the river from the upper stream to the

lower stream. A positive value indicates pollution and a negative indicates purification. Line A indicates the mechanism of the main stream and is  $K$  in Eq. (1), line B indicates the effect of weirs and is related to  $L$  in Eq. (1), and line C indicates the inflows. If the river has no purification mechanism in the inflows, then  $K=0$ ,  $L=0$ , and  $C=1$  (a straight line). If the discussion in section 3 and the learning data (including descriptors) are complete, then lines A, B, and C should be straight and parallel to the x-axis. However, due to some incompleteness in the descriptors and teaching data, all three lines are curved. The cause is the network's compensation for the incompleteness of the data. Therefore, we calculated the overall average for all the observation points.

For BOD (Figure 1a), the averaged values are  $A=0.03$ ,  $B=-0.08$ ,  $C=0.72$ , which reveal the following:

1. Pollution is caused by inflows from branches and sewage plants. Based on the analysis of the measured water quality (described in section 3.3), sewage is a dominant factor.
2. The main stream of the Tamagawa has no purification mechanism for BOD, as evidenced by the low value of  $A=0.03$ .
3. The weirs have a purification mechanism, as evidenced by  $B=-0.08$ .

For COD (Figure 1b), the averaged values are  $A=-0.1$ ,  $B=-0.1$ ,  $C=0.85$ , indicating that both the stream and weirs have purification ability for COD.

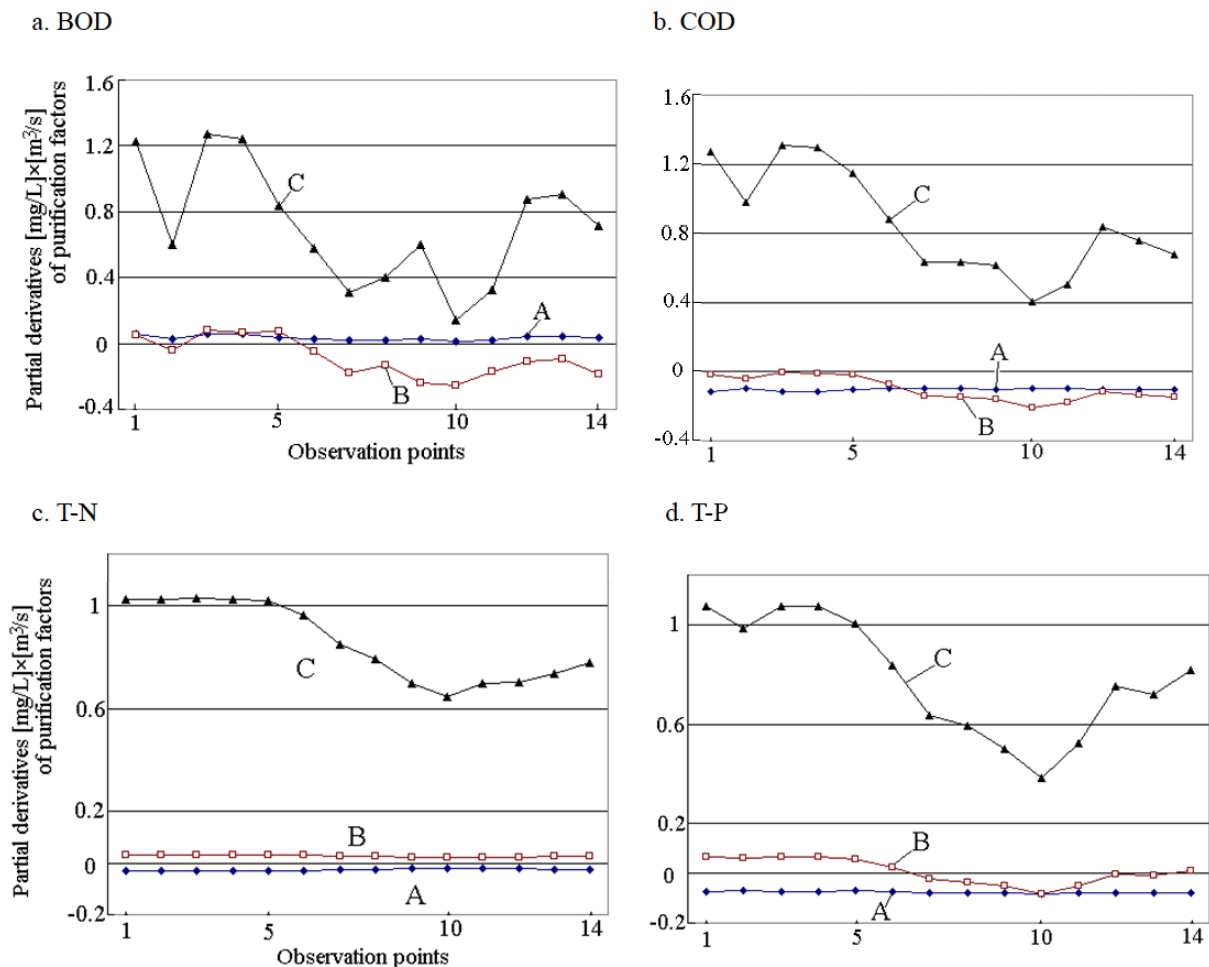


Figure 1. Purification mechanisms of stream, weirs, and inflow for (a) BOD, (b) COD, (c) T-N, and (d) T-P. Vertical and horizontal axis of Figure 1b and 1d are the same as Figure 1a and 1c: partial derivatives of purification factors and observation points, respectively. Line A indicates the mechanism of the main stream. Line B indicates the effect of weirs. Line C indicates the inflows.

For T-N (Figure 1c), the averaged values are  $A=-0.03$ ,  $B=0.03$ ,  $C=0.86$ , indicating that neither the stream nor weirs have purification ability for nitrogen compounds. The Tamagawa discharges the nitrogen compounds into the sea directly.

For T-P (Figure 1d), the averaged values are  $A=-0.08$ ,  $B=-0.007$ ,  $C=0.78$ , indicating that the weirs have no purification ability for phosphorus, whereas the stream has a small ability.

### 5.3 Tests for dummy variables

We introduce dummy variables without doing statistical tests. Here, we discuss the reliability of dummy variables in processing the data. The descriptors of dummy variables remain during the reconstruction learning. The existence of weirs has significance as a necessary condition.

If uniform random numbers are mixed in the dummy variable descriptor, the necessity is decreased according

to the mixing ratio. Neural networks connect input data to teaching data. Even if the input data are random numbers, the connections are realized. The non-significance is detected by the derivatives among input/output. Thus, using random number mixing and the derivatives, there is a possibility to test the dummy variables as a sufficient condition.

We introduce a meaningless vector  $\mathbf{Q}$ . Where, the meaningless is equal to that a descriptor gives no result. On neural network calculations, any descriptor gives a result surely. Therefore, we cannot represent the meaningless by a data set as the descriptor. The representation must be an infinity data set. As the total sum, the meaningless is represented.

However, we cannot express the representation in computing codes because of the infinity; therefore, as a secondary approach, we define the meaningless as a limit. The limit is as follows:

On a sufficiently large number of data set  $\{\mathbf{Q}'\}$ , when

the set are used a descriptor of a neural network, the derivative of the output of the neural network is “ $\lim(\text{the number} \rightarrow \infty) \sum \text{elements}[\partial \text{output} / \partial \text{elements of } \{\mathbf{Q}'\}] \rightarrow \mathbf{0}$ .” The idea is realized for each neural network, so, the definition is not accepted at all.

We introduced a condition that any correlation for each  $\mathbf{Q}'$ -vectors in  $\{\mathbf{Q}'\}$  was  $\mathbf{0}$ . No-correlation is equivalent to the set constructed by independent vectors. One  $\mathbf{Q}'$ -vector gives a result that is  $\xi_1$ ,  $0 < \xi_1 < 1$ . Another  $\mathbf{Q}'$ -vector gives  $\xi_2$ . Then, we get  $\{\xi\}$  from  $\{\mathbf{Q}'\}$ . Considering the distribution of  $\xi_i$  in  $\{\xi\}$ , the distribution is uniform, and the average is a constant, 0.5, which is a stationary point. Thus, “ $\lim(\text{the number} \rightarrow \infty) \sum \text{elements}[\partial \text{output} / \partial \text{elements of } \{\mathbf{Q}'\}] \rightarrow \mathbf{0}$ .” is satisfied approximately. This is our definition for the meaningless. Next, we consider the individual expressions.

Elements of  $\mathbf{Q}$  are 0/1 and the order is random, and  $\{\mathbf{Q}\}$  is an infinite set. There is no dependency among the  $\mathbf{Q}_i$ -vectors. To complete the tests in a finite time, we use a finite  $\{\mathbf{q}\}$  instead of  $\{\mathbf{Q}\}$ . We define  $\{\mathbf{q}\}$  as follows:

1. An arithmetic series of interval  $[0,1]$ ,  $R_0$  is defined as

$$R_0 = \left\{ 0, \frac{1}{(N-1)}, \frac{2}{(N-1)}, \dots, \frac{(N-1)}{(N-1)} = 1 \right\}, \quad (8)$$

where  $N$  is the number of observation points in the river.

2. A number pair  $\{S, T\}$  with range  $[1, 2^{32}-1]$ , and their mod-values  $\{S_m, T_m\}$  are defined as

$$S_m = S_{(\text{mod } N)} + 1; T_m = T_{(\text{mod } N)} + 1, \quad (9)$$

3. The  $S_m$ - and  $T_m$ -elements of  $R_0$  are exchanged. The exchange is repeated 1000 times, yielding a new vector  $R_i$ .

4. The correlation coefficients are calculated as,

$$\{CO_{ij}\} = \frac{\sum_k [R_{jk} - \text{av}(R_{jk})][R_{ik} - \text{av}(R_{ik})]}{\sqrt{\sum_k [R_{jk} - \text{av}(R_{jk})]^2 \sum_k [R_{ik} - \text{av}(R_{ik})]^2}}, \quad j < k. \quad (10)$$

where  $\text{av}$  indicates an average. If the maximum of  $CO_{ij}$  is larger than 0.2, then  $R_i$  is discarded. If not, then  $R_i$  is added to  $\{\mathbf{q}\}$ . Because  $\{\mathbf{q}\}$  is finite, the accidental correlation must be less than 0.01.

5. Thus,  $\{\mathbf{q}_i; i=1,2,\dots,10\}$  is selected from  $\{R_0, R_1, \dots, R_{200000}\}$ .

For each finite  $\mathbf{q}_i$ -vector (that has sufficient meaningless character), the elements of the dummy variable vector are replaced by

$$Rh_i = \alpha \mathbf{q}_i + (1 - \alpha) h_i, \quad 0 < \alpha < 1. \quad (11)$$

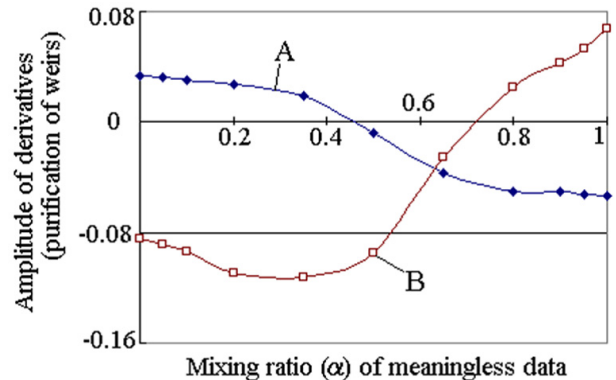


Figure 2. Calculated reliability of the BOD purification mechanisms of weirs

Using  $\{Rh_i\}$ , neural network calculations are executed, and the derivatives are evaluated. The result is Figure 2, which is the averaged  $\alpha$ -dependency curves for the A/B-curves in Figure 1a for the BOD.

If we used ideal learning data, then the amplitude of curve B in Figure 2 would be 0 at  $\alpha=1$ , and B would be a simple increasing curve with increasing  $\alpha$ . However, the amplitude is +0.07 and has a minimum, indicating the presence of unknown descriptors. If the partial character of the unknown descriptors is similar to  $\{Rh_i\}$ , the descriptors work as if they were  $\{Rh_i\}$ , by the replacement of descriptors.

Descriptor replacement occurs in the reverse case. The reverse case is that a  $Rh_i$  vector works as if it is an unknown descriptor. The phenomenon is virtual and the cause is that there is a small defect in our meaningless definition. We already express the defect as an approximation.

We introduce the meaningless vectors to the dummy variables; however, the introduction does not give a meaningless descriptor but a new one. Curve A indicates the effects of the stream, which are induced by the random mixing in the dummy variables. Curve B indicates the effects of the weirs. Both curves are averaged for all  $\mathbf{q}_i$ -vectors in Eq. (11).

Thus, the minimum is revealed. Because the descriptor replacement is partial, curve B simply increases after the minimum is reached. When the existence of an unknown descriptor is predicted, the derivative value of a neural network at  $\alpha=1$  is not zero but a finite value. To balance with the discussion of the meaningless vectors, we assume that the finite value is zero; namely, considering the absolute values, integer values less than the finite value are equal to zero. The purification ability of weirs is  $|-0.08| \approx 0.07$ ; thus, we cannot definitively conclude that weirs decompose the BOD.

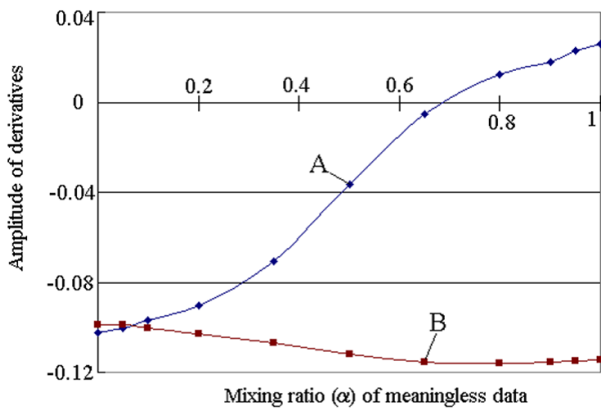


Figure 3a. Calculated reliability of the COD purification mechanisms of a stream

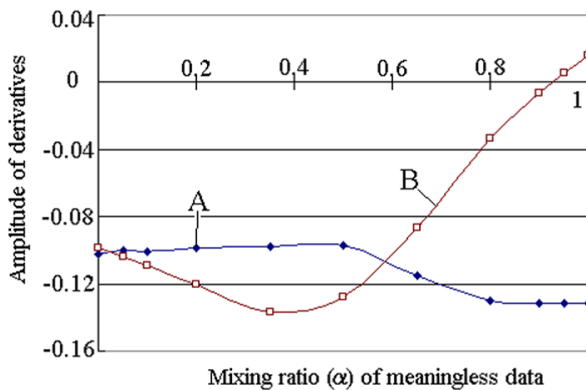


Figure 3b. Calculated reliability of the COD purification mechanisms of weirs

Curve A simply decreases with increasing  $\alpha$ . By eliminating the effect of weirs (i.e.,  $\alpha \rightarrow 1$ ), the weir descriptor is substituted by the stream descriptor. Although this substitution is inappropriate, purification of the stream is due to the strong non-linear fitting of neural networks. As a result of the substitution, the decrease in the curve is revealed. The purification is not realistic but virtual. The virtual purification yields no conclusion about the BOD. We therefore cannot draw any definitive conclusion about the BOD purification in the Tamagawa.

Figures 3a, 3b show the calculated reliability of the COD purifications of the stream and weirs. The derivative value at  $\alpha=1$  is  $+0.03$  for the stream and  $+0.02$  for the weirs. Thus, considering  $|-0.1| > 0.03$  and  $|-0.1| > 0.02$ , the COD of the stream and weirs is reliable. In conclusion, the Tamagawa purifies COD.

Figure 4 shows the calculated reliability of the T-P purification of the stream. The derivative value at  $\alpha=1$  is  $-0.015$ . Considering the difference of order in the inequality,  $|-0.08| > |-0.015|$ , the calculated purification of the stream is reliable. In conclusion, the main stream of the Tamagawa purifies T-P.

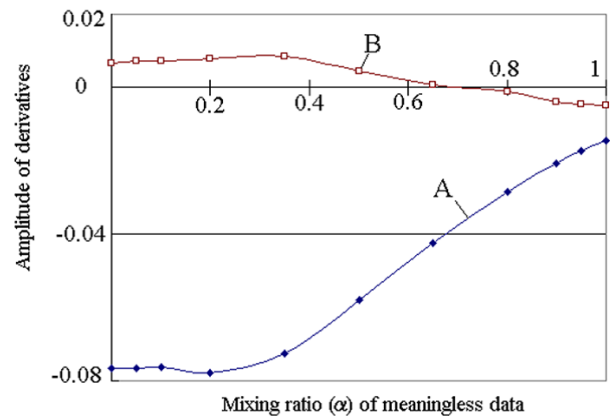


Figure 4. Calculated reliability of the T-P purification mechanisms of the stream

## 6 Conclusions

We developed a river model that includes the effects of inflows, stream, and weirs, but does not include transpiration and percolation. The model is discrete, and is suitable for estimating water purification in a river. The model can be applied to a river that has 10-20 observation points.

The discreteness of observations can be converted to a consecutive function by using a neural network. Because the network function is differential, unknown coefficients of descriptors in the model can be predicted from the partial derivatives of the network. We then applied the model and neural network calculations to four water quality indexes in the Tamagawa; BOD, COD, T-N, and T-P. The results were tested by introducing meaningless descriptors. The results yielded the following conclusions:

1. The cause of pollution is inflows from sewage.
2. The stream of the Tamagawa has purification functions for COD and T-P, but has little ability for T-N.
3. The weirs have purification for COD, but have no purification for T-N and T-P.
4. An unknown descriptor for BOD purification is a possibility.

Currently, the Tamagawa apparently has a purification function. However, the function is limited; the river showed no purification for T-N, and only a little purification for T-P. Because these indexes will increase as the population increases, we are not optimistic about the pollution situation.

We thank Professor H. Chuman of Tokushima University for numerous stimulating discussions.

## References

- [1] J. Nakanishi, “*Environmental strategy of water (in Japanese)*”, Iwanami Pub. Co. Ltd., Tokyo, Japan (1994).
- [2] Tokyo Metropolitan Government Bureau of Environment, “Kokyoyo suiiki suishitsu sokutei kekka (Reports of quality of water for supply) (in Japanese)”  
<http://www2.kankyo.metro.tokyo.jp/kansi/mizu/sokutei/sokuteikekka/kokyou.htm>  
(2002).
- [3] D. E. Rumelhart, J. L. McClelland, the PDP research group, “*Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1, 2*”, the MIT Press, Boston, USA (1986).
- [4] T. Aoyama, J. Kambe, U. Nagashima, *J. Comput. Chem. Jpn.*, **5**, 101-118 (2006).
- [5] T. Aoyama, H. Ichikawa, *Chem. Pharm. Bull.*, **39**, 1222-1228 (1991).
- [6] H. Hirao, Electronic publication  
<http://mikilab.doshisha.ac.jp/dia/research/SA/>  
(in Japanese) (2005).
- [7] Jpn. Soc. Hydrology and Water Resources ed., “*Hand book for hydrology and water resources (in Japanese)*”, Asakura Pub. Co. Ltd., Tokyo, Japan (1997).
- [8] T. Aoyama, H. Ichikawa, H., *Chem. Pharm. Bull.*, **39**, 372-378 (1991).
- [9] OECD, “OECD Environmental data, compendium 2002”, OECD, Paris, France (2002).
- [10] Tokyo Metropolitan Government Bureau of Sewerage, Summary of sewerage in Tokyo, Electronic publication,  
<http://www.gesui.metro.tokyo.jp/gijyutou/jg16/jg16.htm>  
or  
<http://www.gesui.metro.tokyo.jp/english/english.htm>  
(2004).
- [11] S. Ohgaki, ed., “*River and Nutrient Salt; foundation of river and watershed environmental management (in Japanese)*”, Giho-do Pub. Co. Ltd., Tokyo, Japan (2005).