In Proceedings of the 39th SICE Annual Conference International Session Papers, 112A-4, pp.1-6, 2000

GMDH-type Neural Networks with a Feedback Loop and their Application to the Identification of Large-spatial Air Pollution Patterns.

Tadashi Kondo[1] and Abhijit S.Pandya[2]

[1]School of Medical Sci.,The Univ.of Tokushima,3-18-15 Kuramoto-cho Tokushima 770-8509,Japan
[2]Comp.Sci.& Eng.,Florida Atlantic Univ.,Boca Raton, FL 33431,U.S.A.

Abstract: The GMDH (Group Method of Data Handling)-type neural networks with a feedback loop have been proposed in our early work. The architecture of these networks is generated by using the heuristic self-organization method that is the basic theory of the GMDH method. The number of hidden layers and the number of neurons in the hidden layers are determined so as to minimize the error criterion defined by Akaike's Information Criterion (AIC). Furthermore, the optimum neurons that can handle the complexity of the nonlinear system are selected from a variety of prototype functions, such as the sigmoid function, the radial basis function, the high order polynomial and the linear function. In this study, the GMDH-type neural networks with a feedback loop is applied to the identification of large-spatial air pollution patterns. The source-receptor matrix that represents a relationship between the multiple air pollution sources and the air pollution concentration at the multiple monitoring stations is accurately identified by using the GMDH-type neural networks with a feedback loop. The identification results of the GMDH-type neural networks are compared with those identified by other identification methods.

Keywords: GMDH, Neural networks, Identification, Air pollution

1. Introduction

The GMDH-type neural networks with a feedback loop have been proposed in our early works [1]. The GMDH-type neural networks can organize the optimum multilayered neural network architecture by using the heuristic self-organization method [2],[3]. In each layer, numerous intermediate variables are combined with each other to generate the optimum neuron architectures. The intermediate variables are the outputs of the previous layer. They describe the partial characteristics of the nonlinear system and each combination successively increases the complexity of the neural networks. In the conventional GMDH-type neural networks [4], the outputs of the neurons, which are the intermediate variables, are used as the input variables of the next layer and are combined each other, so the complexity of the neural networks increases rapidly. In the GMDH-type neural networks with a feedback loop in this study, the outputs of the neurons are not combined with each other but they are combined with the input variables of the system. Therefore, the complexity of the neural network increases gradually and a more accurate structural identification of the nonlinear system can be carried out through the layers.

In this paper, the GMDH-type neural networks with a feedback loop are applied to the identification problem of large-spatial air pollution patterns. Numerous air pollution models have been derived using various methods for the identification problem of large-spatial air pollution patterns. The GMDH model[2],[3], the multilayered neural network model [5],[6], the combined model of the source-receptor matrix and the GMDH, the combined model of the source-receptor matrix and the neural networks, and the physical models, have been proposed [7]. Here, the source-receptor matrix was usually estimated by the physical model or a linear regression analysis with limited accuracy [8]. In this paper, the source-receptor matrix that represents a relationship between the multiple air pollution sources and the air pollution concentration at the multiple monitoring stations is accurately estimated by using the GMDH-type neural networks. Then the air pollution concentration patterns at the large area are identified by using estimated source-receptor matrix. The identification results of the GMDH-type neural networks are compared with those identified by other identification methods.

2. GMDH-type neural network algorithm with a feedback loop [1]

The GMDH-type neural networks which have multilayered neural network architectures are automatically organized by using the heuristic self-organization method [2],[3] and the structural parameters such as the useful input variables, the number of layers, the number of neurons in a hidden layer and the optimum architectures of the neurons in a hidden layer, are automatically determined in this algorithm. It is very easy to organize the optimum neural network architectures which fit the complexity of the nonlinear system.

The GMDH-type neural networks in this paper has a feedback loop. Figure 1 shows the architecture of the GMDH-type neural networks. By using the loop calculation, the neural network architectures are organized by degrees.

[1] The first loop calculation.

First, the original data are separated into training and testing sets. Then the architecture of the input layer is organized.

< The input layer>

$$u_j=x_j \quad (j=1,2,\cdots,p) \tag{1}$$

where x_j (j=1,2,⋯,p) are the input variables of the system, and p is the number of input variables. In the first layer, input variables are set to the output variables.

<The hidden layer>

Many combinations of the input variables are generated. For each combination, optimum neuron architectures which fit the characteristics of the nonlinear system are automatically selected by using AIC [9] out of the following seven types of neurons. Figure 2 shows the architectures of the neurons.

(1) The first type neuron.

Σ: (Nonlinear function)

$$z_k= w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j +w_8u_iu_j^2+w_9u_j^3 - w_0\,\theta_1 \tag{2}$$

f : (Nonlinear function)

$$y_k= 1 / (1+exp (- z_k)) \tag{3}$$

(2) The second type neuron.

Σ: (Nonlinear function)

$$z_k= w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j +w_8u_iu_j^2+w_9u_j^3 - w_0\,\theta_1 \tag{4}$$

f : (Linear function)

$$y_k= z_k \tag{5}$$

(3) The third type neuron.

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r - w_0\,\theta_1 \quad (r<p) \tag{6}$$

f : (Nonlinear function)

$$y_k= 1 / (1+exp (- z_k)) \tag{7}$$

(4) The fourth type neuron.

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r - w_0\,\theta_1 \quad (r<p) \tag{8}$$

f : (Linear function)

$$y_k= z_k \tag{9}$$

(5) The fifth type neuron.

Σ: (Nonlinear function)

$$z_k= w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j +w_8u_iu_j^2+w_9u_j^3 - w_0\,\theta_1 \tag{10}$$

f : (Nonlinear function)

$$y_k= exp (- z_k^2) \tag{11}$$

(6) The sixth type neuron.

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r - w_0\,\theta_1 \quad (r<p) \tag{12}$$

f : (Nonlinear function)

$$y_k= exp (- z_k^2) \tag{13}$$

(7) The seventh type neuron.

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r - w_0\,\theta_1 \quad (r<p) \tag{14}$$

f : (Nonlinear function)

$$y_k= a_0+a_1z_k+a_2z_k^2+\cdots +a_mz_k^m \tag{15}$$

Here, θ_1=1and w_i (i=1,2,…) are weights between the neurons and estimated by applying the stepwise regression analysis [10] to the training data. The output variables y_k of the optimum neurons are called intermediate variables.

<The output layer >

In the output layer, the outputs of the neurons in the hidden layer are combined by the following linear function.

$$\Phi^*= a_0 + \sum_{k=1}^{L} a_k y_k \tag{16}$$

Here, L is the number of combinations of the input variables and y_k is the intermediate variables. The useful intermediate variables y_k are selected by using the stepwise regression analysis [10] in which AIC [9] is used as the variable selection criterion. Then, the estimated output values Φ^* is used as the feedback value and it is combined with the input variables in the next loop calculation.

[2] The second and succeeding loop calculations.

In the second and succeeding loop calculations, the estimated output value Φ^* is combined with the input variables. First, many combinations between the estimated output value Φ^* and the input variables are generated. The same calculation as the first loop is carried out for each combination. If the value of AIC in the linear function (16) is increased, the loop calculation is terminated and the complete neural network architecture is organized.

By using the above procedures, the GMDH-type neural networks with a feedback loop can be organized.

3. Application to the identification of large-spatial air pollution patterns.

In the case of regional environmental planning and impact assessment, it is necessary to estimate the spatial distribution pattern of each air pollution source at a large area. For this purpose, the source receptor matrix presenting the relationship between each air pollution source and the air pollution concentration at each monitoring station is very useful [8]. For estimating the source-receptor matrix, it has been proposed to use physical model such as plume model, puff model and box model and furthermore non physical model such as a linear regression model. On the other hand, the combined model of the source-receptor matrix and the GMDH, and the combined model of the source-receptor matrix and neural networks have been proposed for monthly or yearly average prediction of the air pollution concentration at a large area [7].

In this study, the source-receptor matrix is estimated by using the GMDH-type neural networks with a feedback loop and then the air pollution concentration pattern at a large area are identified by using estimated source-receptor matrix. The identification results are compared with those identified by other models such as the combined model of the source-receptor matrix and the GMDH, and the combined model of the source-receptor matrix and neural networks.

3.1 Source-receptor matrix [8]

It is assumed that an air pollution model used for monthly or yearly average prediction of air pollution concentration can be described by the following equation for single air pollution source

$$c = f \cdot q \tag{17}$$

where c is the air pollution concentration at the monitoring station, q is the emission intensity of pollution source and f is a coefficient determined by various conditions such as the pollution source, the diffusion field and the atmospheric stability. For multiple sources, the following equation is used.

$$c_i = \sum_{j=1}^{N} f_j(X_{ij},Y_{ij})\, q_j, \quad i=1,2,\ldots,M, \quad j=1,2,\ldots,N \tag{18}$$

where $X_{ij}=X_i^r-X_j^s$, $Y_{ij}=Y_i^r-Y_j^s$, c_i is the air pollution concentration at the i-th monitoring station, q_j is the emission intensity of the j-th pollution source, (X_i^r,Y_i^r) are the coordinates of the i-th monitoring station, (X_j^s,Y_j^s) are the coordinates of the j-th pollution source, M is number of monitoring stations, and N is number of pollution sources.

Eq. (18) can be written as

$$\mathbf{c} = F\,\mathbf{q} \tag{19}$$

where

$$\mathbf{c}^T=(c_1,c_2,\ldots,c_M) \tag{20}$$

$$\mathbf{q}^T=(q_1,q_2,\ldots,q_N) \tag{21}$$

$$F= (\,\mathbf{f}_1 , \mathbf{f}_2 , \ldots \;\; , \mathbf{f}_N) \tag{22}$$

$$\mathbf{f}_i^T= (\, f_{1i} , f_{2i} , \ldots , f_{Mi}) \tag{23}$$

Here F is called as source-receptor matrix.

Physical model such as plum model, puff model and box model has been proposed to use in order to estimate each element of the source-receptor matrix but physical model has limitations in practical applications because complex factors such as topography, down wash and down draught can not be easily taken into account in the model theoretically. So instead of using physical model, it has been proposed to use a regression analysis of the spatially distributed data obtained from wind tunnel experiments or the computer simulation of the air pollution diffusion for a single source.

In this paper, the GMDH-type neural networks with a feedback loop is organized by using synthetic data obtained by the computer simulation of the air pollution diffusion. Each element of the source receptor matrix is estimated by the GMDH-type neural networks.

3.2 Estimation of the source-receptor matrix by the GMDH-type neural networks with a feedback loop

If there is an air pollution source of intensity one at the coordinates (4,4) and the diffusion rate of the pollutant are 0.2, 0.2, 0.5 and 0.1, a steady-state pattern of the spatially distributed air pollution concentration is obtained by computer simulation as shown in Fig.3 (b). The GMDH-type neural networks with a feedback loop can estimate each element of the source receptor matrix by using the data such as the distribution pattern shown in Fig.3 (b).

Fig.4 shows the input and output variables of the GMDH-type neural networks. The input variables are the coordinates X_{i1}, Y_{i1}, X_{i2}, Y_{i2}, … ,X_{iN}, Y_{iN} at the monitoring station and the output variables are elements of the source receptor matrix $\mathbf{f}_1,\mathbf{f}_2,\ldots,\mathbf{f}_N$. The GMDH-type neural networks are organized in order to estimate the elements of source-receptor matrix for each air pollution source, therefore the number of neural networks is N.

3.3 Identification of large-spatial air pollution patterns

In this study, a spatially distributed pattern of air pollutant emitted from three air pollution sources is identified by using the source-receptor matrix estimated by GMDH-type neural networks with a feedback loop. It is assumed that there exists three air pollution sources at the coordinates (2,4),(4,2) and (4,6) as shown in Fig.5 (a). A steady-state air pollution concentration pattern shown in Fig.5 (b) is obtained as the results of the computer simulation of the air pollution diffusion. The data underlined in Fig.5 (b) are assumed to have been observed at the monitoring stations. The diffusion rate of the pollutant are 0.2,0.2,0.5 and 0.1.

The GMDH–type neural networks with a feedback loop are organized by using the data of each air pollution source and each elements of the source receptor matrix are estimated by organized neural networks. The estimation results for the air pollution source (2,4) are shown as follows:

(1) The input variables.

Four variables $X_{i1}, Y_{i1}, X_{i1}^2, Y_{i1}^2$ are used as the input variables of the GMDH-type neural networks.

(2) The number of selected neurons in the hidden layers.

Four neurons were selected in the hidden layers.

(3) The architecture of the neural networks.

The calculation of the GMDH-type neural networks was terminated at the seventh loop calculation.

(4) Variation of AIC

Variation of AIC [9] in (16) is shown in Fig.6. The values of AIC decreased gradually by the feedback loop calculations and converged at the seventh feedback loop calculation.

For the other air pollution sources at the coordinates (4,2) and (4,6), the same estimation results were obtained. Three GMDH-type neural networks were organized and the source receptor matrix was estimated completely as follows.

$$F = \begin{bmatrix} 0.002 & 0.002 & 0.000 \\ 0.073 & 0.009 & 0.000 \\ 0.115 & 0.020 & -0.043 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0.2164 & 0.000 & 0.340 \end{bmatrix} \qquad (24)$$

Air pollution concentrations were calculated by using estimated source-receptor matrix (24) when the intensities of air pollution sources $\mathbf{q}^T$ was (1,2,3). Fig.7 shows the identified air pollution pattern.

(5) Estimation accuracy at observed points.

The estimation accuracy at observed points was evaluated by using the following equation

$$J_1 = \sum_{i=1}^{25} | \Phi_i^* - \Phi_i | / \Phi_i \qquad (25)$$

where Φ_i (i=1,2,…,25) were the observed values and Φ_i^* (i=1,2,…,25) were the estimated values by the source-receptor matrix. J_1 is shown in Table1.

(6) Estimation accuracy at unobserved points

The estimation accuracy at unobserved points was evaluated by using the following equation.

$$J_2 = \sum_{i=1}^{24} | \Phi_i^* - \Phi_i | / \Phi_i \qquad (26)$$

where Φ_i (i=1,2,…,24) were actual values and Φ_i^* (i=1,2,…,24) were the estimated values by the source-receptor matrix. J_2 is shown in Table1.

3.4 Comparison of the GMDH-type neural networks and other identification models.

The identification results were compared with those identified by other models [7] such as the combined model of the source-receptor matrix and the GMDH, the combined model of the source-receptor matrix and the neural networks, the neural networks model, and the source-receptor matrix estimated by the regression analysis. Table 1 shows the comparison of the GMDH-type neural networks and other identification methods. Fig.8 shows the block diagram of the prediction system using the combined model. Here, prediction procedures of the combined model are as follows. First, a source-receptor matrix was estimated by a regression analysis and this source-receptor matrix was used as a rough model of first-order approximation. Then, the difference between the output of the real system and the output of the rough model was identified by the GMDH [2],[3] or the neural networks [5],[6]. The estimated values of the source-receptor matrix were modified by these difference values. From Table1, the estimation accuracy J_1 and J_2 of the source-receptor matrix estimated by the GMDH-type neural networks are much smaller than those of the source-receptor matrix estimated by a regression analysis. Estimation accuracy J_2 of the source-receptor matrix estimated by the GMDH-type neural networks is smallest in the prediction models. We can see that good estimation accuracy can be obtained by using the source-receptor matrix estimated by the GMDH-type neural networks.

4. Conclusion

The GMDH-type neural networks with a feedback loop can automatically organize the optimum neural network architecture by using the heuristic self-organization method and so it is very easy to apply them to the practical complex problem. In this study, the GMDH-type neural networks with a feedback loop were applied to the identification problem of large-spatial air pollution patterns. The source-receptor matrix that presents a relationship between the multiple air pollution sources and the air pollution concentration at the multiple monitoring stations was accurately estimated by the GMDH-type neural networks with a feedback loop. Then, the large-spatial air pollution pattern was identified by using the source-receptor matrix estimated by the GMDH-type neural networks. The identification results of the GMDH-type neural networks are compared with those identified by other identification methods and it is shown that the GMDH-type neural networks are easy to apply for the identification problem of large-spatial air pollution patterns because the optimum neural network architecture is automatically organized so as to minimize AIC. It is also more accurate than other identification methods.

References

[1] Kondo T, Pandya A.S. and Zurada J.M., GMDH-type neural networks with a feedback loop and their application to nonlinear system identification, Smart Engineering System: Neural Networks, Fuzzy Logic, Evolutionary Programing, Data Mining, and Rough Sets, ASME Press, Vol.9,pp.117-124,1999.

[2] Ivakhnenko A.G., Heuristic self-organization in problems of engineering cybernetics, Automatica, Vol.6, No.2 No.2, pp.207-219,1970.

[3] Farlow S.J. ed., Self-organizing Methods in Modeling, GMDH-type Algorithms, Marcel Dekker, Inc., New York, 1984.

[4] Kondo T., Pandya A.S. and Zurada J.M., GMDH-type neural networks and their application to the medical image recognition of the lungs, Proc. of the 38th SICE Annual Conference, International Session Paper,pp.1181-1186,1999.

[5] Zurada J.M., Introduction to Artificial Neural Systems, PWS, Boston,1992.

[6] Pham D.T. and Liu X., Neural Networks for Identification, Prediction and Control, Springer-Verlag London, 1995.

[7] Kondo T., Identification of large-spatial air pollution patterns using a neural network, Trans. of the institute of systems, control and information engineers, Vol.7, No.2, pp.59-67, 1994.

[8] Tamura H. and Kondo T., Large-spatial pattern identification of air pollution by a combined model of source-receptor matrix and revised GMDH, Proc. of IFAC sympo. On environmental systems planning, design and control, pp.373-380,1977.

[9] Akaike H., A new look at the statistical model identification, IEEE Trans. Automatic Control, Vol.AC-19, No.6, pp.716-723, 1974.

[10] Draper N.R. and Smith H., Applied Regression Analysis, John Wiley and Sons, New York, 1981.

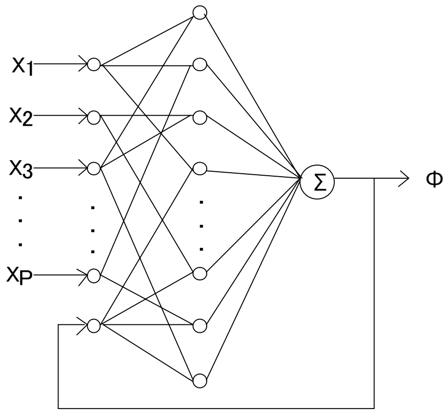
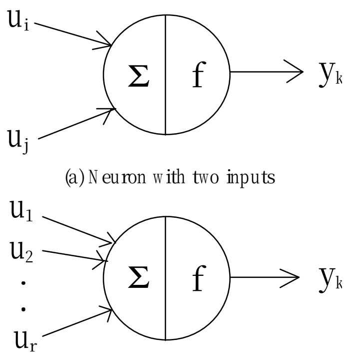


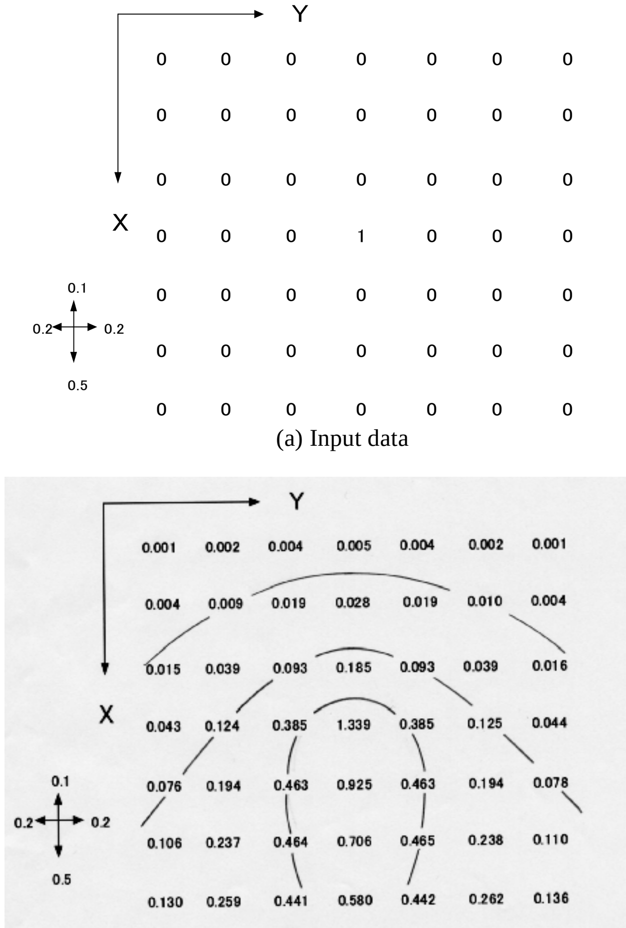
Fig.1 Architecture of the GMDH type neural networks [1]



(a) Neuron with two inputs

(b) Neuron with r inputs

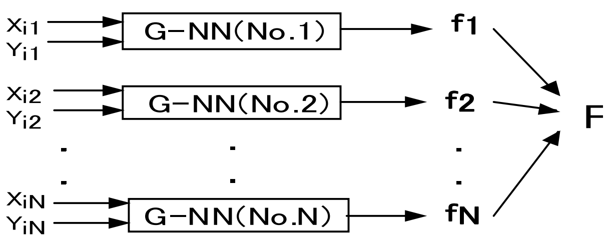
Fig.2 Architectures of the neurons



(a) Input data

(b) Output data

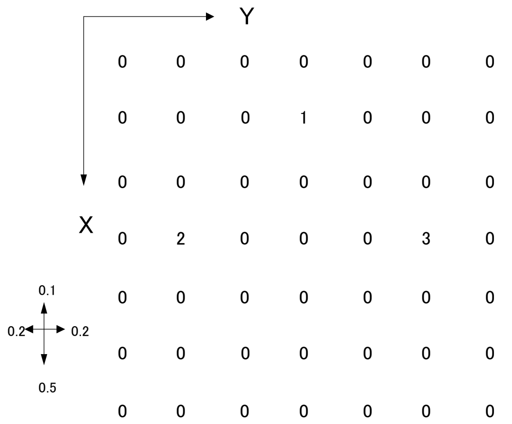
Fig.3 Input and output of the simulator for single source[8]



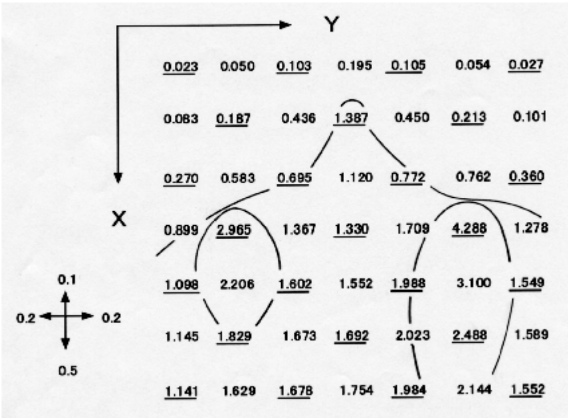
F : Source-receptor matrix

G-NN : GMDH-type neural networks

Fig.4 Input and output of the GMDH-type neural networks



(a) Input data



(b) Output data

Fig.5 Input and output of the simulator for multiple sources [8]

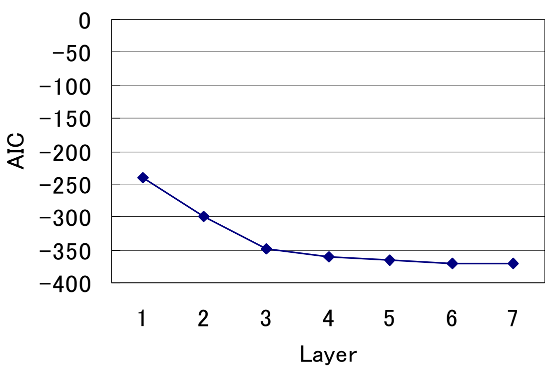


Fig.6 Variation of AIC

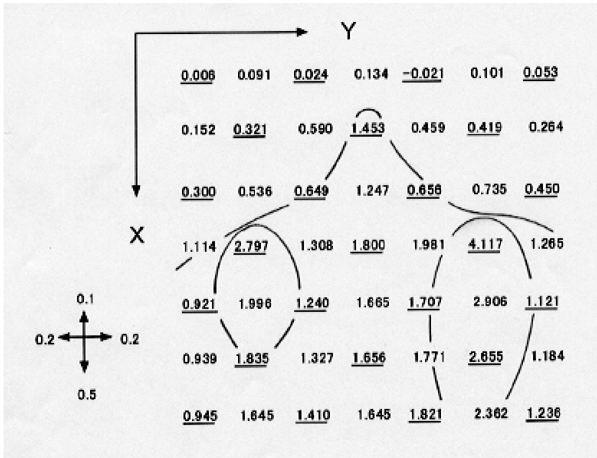


Fig.7 Identified air pollution pattern by using the source-receptor matrix

Table 1 Comparison of the GMDH-type neural networks and other identification methods

Prediction model	J_1(%)	J_2(%)
Source-receptor matrix estimated by GMDH-type neural networks	13.2	12.0
Source-receptor matrix estimated by a regression analysis	29.6	30.7
Combined model of source-receptor matrix and GMDH	—	19.7
Combined model of source-receptor matrix and neural networks	9.9	12.3
Neural networks	30.3	22.6

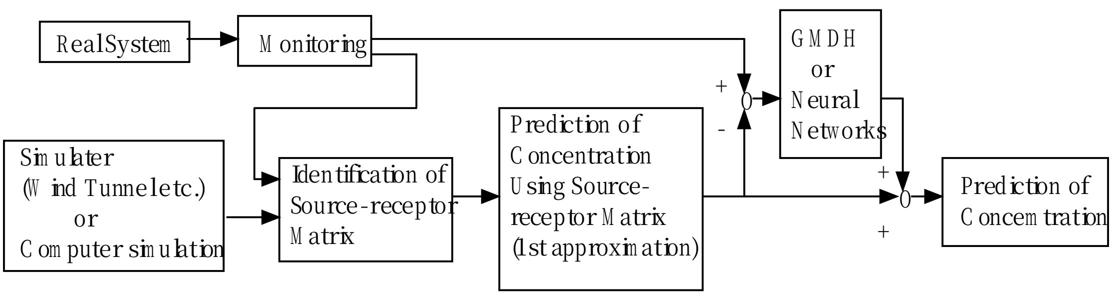


Fig.8 Block diagram of the prediction system using the combined model [7]