Fuzzy Sets and Systems 36 (1990) 245–258
North-Holland

THE FUZZY GMDH ALGORITHM BY POSSIBILITY MODELS AND ITS APPLICATION

Isao HAYASHI
Central Research Laboratories, Matsushita Electric Industrial Co., Ltd., Moriguchi, 570 Japan

Hideo TANAKA
Department of Industrial Engineering, College of Engineering, University of Osaka Prefecture, Sakai, Osaka, 591 Japan

Received March 1988
Revised January 1989

Abstract: It is the purpose of this paper to formulate the Group Method of Data Handling (GMDH) by the possibility model. The GMDH formulated by the possibility model is called the fuzzy GMDH. To formulate the fuzzy GMDH, we discuss possibilistic linear regression by the LP problem, called Min Problem. In the fuzzy GMDH, the estimated model by using possibilistic linear regression with a multilayer procedure is a non-linear system with parameters in the form of fuzzy numbers. The estimation by the fuzzy GMDH can be obtained in fuzzy numbers. Thus, this approach can be regarded as fuzzy interval analysis.

Keywords: GMDH; possibility model; possibilistic linear regression; Min Problem; fuzzy interval analysis.

1. Introduction

Linear regression analysis is a well-known technique for obtaining mathematical models which represent the relationship between input and output data. However, if the relationship between input and output data is complex and non-linear, it is not easy to determine a set of input variables for identifying models and the number of powers of input variables, since we can consider many kinds of combinations of input variables. Thus, the GMDH [4, 5] (Group Method of Data Handling) has been developed to obtain the optimal combination of input variables of non-linear models. In the GMDH, the estimated non-linear model is obtained by combining second-order functions of two variables with a multilayer procedure. In each layer, parameters of the function of two variables are determined by linear regression analysis. Hence, the GMDH can be said to be based on linear regression analysis.

With the above view, we propose the fuzzy GMDH based on fuzzy linear regression analysis in this paper. Fuzzy linear regression analysis, called also possibilistic linear regression has been proposed by Tanaka et al. [1] and discussed in detail in [2, 3]. We can represent fuzziness by using possibilistic linear systems. In possibilistic linear regression, deviations between the observed data and the estimated values are assumed to depend on the possibility of the system structure. Thus, we regard these deviations as the possibility of system parameters. However, in possibilistic linear regression, we have to determine a structure of the model in advance, as in conventional linear regression analysis. Thus, in this paper, we propose a new method by which the structure of the model can be automatically found. Following the concept of the GMDH, we formulate the fuzzy GMDH based on possibilistic linear systems. In our formulation of the fuzzy GMDH, our characteristic is that parameters in the model are obtained as fuzzy numbers by possibilistic linear regression. The concept of fitting index of the estimated model to the data and a stopping rule in the multilayer procedure are introduced to the algorithm of the fuzzy GMDH in order to obtain a good estimation and terminate the algorithm. The fuzzy GMDH is a powerful method for obtaining the non-linear model for identifying a fuzzy phenomenon.

To explain the applicability of the algorithm of the fuzzy GMDH, we apply the fuzzy GMDH to structural identification of the amount of production of computers in Japan and to analysis and prediction of water temperatures in a dam reservoir. The fuzzy GMDH is a new method proposed from the view-point of possibility. This method might be called the fuzzy interval analysis and its results can be obtained in fuzzy numbers.

2. Possibilistic linear systems

A linear system whose parameters are fuzzy numbers, is called a possibilistic linear system. It is assumed that a fuzzy number M, $\mu_M : \mathbb{R} \to [0, 1]$, satisfies:

$$M_\lambda = \{x \mid \mu_M(x) \geq \lambda\} \text{ is a closed interval.} \tag{1}$$

$$\exists x \text{ such that } \mu_M(x) = 1. \tag{2}$$

$$\mu_M(\lambda x_1 + (1-\lambda)x_2) \geq \mu_M(x_1) \wedge \mu_M(x_2) \quad \text{for } \lambda \in [0, 1]. \tag{3}$$

A general representation of a fuzzy number is given as the L–R type of fuzzy numbers by Dubois and Prade [6]. Since we regard only a symmetric fuzzy number in this paper, a fuzzy number is defined as follows.

Definition 1. A symmetric fuzzy number A denoted $A = (\alpha, c)_L$ is defined by

$$\mu_A(x) = L((x-\alpha)/c), \quad c \geq 0, \tag{4}$$

where the reference function $L(x)$ satisfies (i) $L(x) = L(-x)$, (ii) $L(0) = 1$ and (iii) $L(x)$ is strictly decreasing on $x > 0$.

As an example of $L(x)$, $\max(0, 1 - |x|)$ is used in this paper. Thus, a fuzzy number A becomes a triangular fuzzy number with the center value α and the width c.

An arithmetic with fuzzy numbers is defined by the extension principle [6] as follows.

Definition 2. Given a function $y = f(x_1, x_2, \ldots, x_n)$, a fuzzy output $Y = f(A_1, A_2, \ldots, A_n)$ whose inputs are fuzzy numbers $A_1, A_2, \ldots, A_n$ in exchange for inputs $x_1, x_2, \ldots, x_n$ is defined as

$$\mu_Y(y) = \sup_{\{x \mid y = f(x)\}} \mu_{A_1}(x_1) \wedge \cdots \wedge \mu_{A_n}(x_n). \tag{5}$$

Now, a possibilistic linear system whose parameters are symmetric fuzzy numbers $A_j = (\alpha_j, c_j)_L$ is denoted by

$$Y = A_1 x_1 + A_2 x_2 + \cdots + A_n x_n = \mathbf{A}\boldsymbol{x} \tag{6}$$

where the membership function of a fuzzy parameter A_j is

$$\mu_{A_j}(a_j) = L((a_j - \alpha_j)/c_j). \tag{7}$$

A membership function $\mu_Y(y)$ of a possibilistic linear system (6) is rewritten using the extension principle in Definition 2 as follows.

Theorem 1. *The membership function of a fuzzy output Y is*

$$\mu_Y(y) = L((y - \boldsymbol{\alpha x})/\boldsymbol{c}\,|\boldsymbol{x}|) \tag{8}$$

where $|\boldsymbol{x}| = (|x_1|, |x_2|, \ldots, |x_n|)^{\mathrm{T}}$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ *and* $\boldsymbol{c} = (c_1, c_2, \ldots, c_n)$.

The proof is shown in [7]. Given a fuzzy parameter A_j, the membership function of a fuzzy output Y can be calculated easily. A similar discussion is given by Nahmias [8].

3. Possibilistic linear regression

To formulate possibilistic linear regression, let us consider the following definition $b \in_h A$, where A is an L–R fuzzy number and b is a real number.

Definition 3. The containment $b \in_h A$ with a threshold h is defined as

$$b \in_h A \Leftrightarrow b \in [A]_h$$

where

$$[A]_h = \{x \mid \mu_A(x) \geq h\}. \tag{9}$$

We assume that deviations between the observed values and the estimated values depend on the possibility of the system parameters. Thus, the possibility of estimated values can be regarded as a fuzzy number.

Let us consider the following possibilistic linear system as a linear regression model:

$$Y_i^* = A_0^* + A_1^* x_{i1} + \cdots + A_n^* x_{in}, \quad i = 1, 2, \ldots, N, \tag{10}$$

where the input vector $x_i = (x_{i1}, \ldots, x_{in})^{\mathrm{T}}$ and the output y_i are non-fuzzy. From relations between outputs y_i and fuzzy estimated values Y_i^*, we consider the

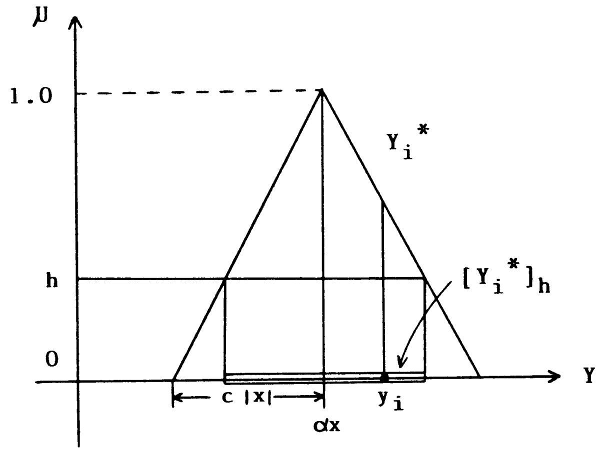

Fig. 1. The concept of Min Problem.

following problem which is to determine fuzzy parameter $\boldsymbol{A}^*$ when a threshold h is given:

$$y_i \in_h Y_i^* = A_0^* + A_1^* x_{i1} + \cdots + A_n^* x_{in} = \boldsymbol{A}^* \boldsymbol{x}_i. \tag{11}$$

By Definition 3 and (11), our requirement is that the output data y_i should be contained in the estimated fuzzy outputs Y_i^* in the sense of h-level set, i.e.

$$y_i \in_h Y_i^* \Leftrightarrow y_i \in [Y_i^*]_h, \quad i = 1, 2, \ldots, N. \tag{12}$$

In possibilistic linear regression, the problem to determine fuzzy parameter $\boldsymbol{A}^*$ is called Min Problem [7]. The concept of Min Problem is shown in Figure 1. The output y_i is included in the h-level set $[Y_i^*]_h$ of the fuzzy estimated value Y_i^*.

The width of the estimated value Y_i^*, i.e. $\boldsymbol{c}\,|\boldsymbol{x}_i|$, can be considered as fuzziness based on the possibilistic linear system. The smaller $\boldsymbol{c}\,|\boldsymbol{x}_i|$ is, the better the estimated model is. Thus, we consider the following objective function concerning the sum of the widths of Y_i^*. The Min Problem is reduced to an LP problem under (12):

$$\min_{\boldsymbol{\alpha}, \boldsymbol{c}} \; J(\boldsymbol{c}) = \sum_{i=1}^{N} \boldsymbol{c}\,|\boldsymbol{x}_i| \tag{13}$$

$$\begin{aligned} \text{subject to} \quad & y_i \leqslant \boldsymbol{\alpha}\boldsymbol{x}_i + |L^{-1}(h)|\, \boldsymbol{c}\,|\boldsymbol{x}_i|, \\ & y_i \geqslant \boldsymbol{\alpha}\boldsymbol{x}_i - |L^{-1}(h)|\, \boldsymbol{c}\,|\boldsymbol{x}_i|, \\ & \boldsymbol{c} \geqslant 0, \quad i = 1, 2, \ldots, N, \end{aligned}$$

where $J(\boldsymbol{c})$ is called the index of fuzziness of the Min Problem.

4. The fuzzy GMDH

The GMDH based on the principles of heuristic self-organization is developed to solve complex problems with large dimensionality. We formulate the GMDH by the possibilistic model, which is called the fuzzy GMDH. We assume that input–output relations $(\boldsymbol{x}, y)$ can be specified by the following polynomial whose parameters are fuzzy numbers:

$$Y^* = A_0 + \sum_{k_1} A_{k_1} x_{k_1} + \sum_{k_1} \sum_{k_2} A_{k_1 k_2} x_{k_1} x_{k_2} + \cdots, \quad y \in_h Y^* \tag{14}$$

where parameters $A_0, A_{k_1}, \ldots$ are fuzzy numbers. As the conventional GMDH, the model Y^* is obtained by combining the following second-order possibilistic regression models of two variables in multilayers:

$$Y_k = A_{0k} + A_{1k} x_i + A_{2k} x_j + A_{3k}(x_i)^2 + A_{4k}(x_j)^2 + A_{5k} x_i x_j, \quad k = 1, 2, \ldots, \quad i, j = 1, 2, \ldots, i \neq j. \tag{15}$$

This possibilistic model is called the partial description.

In the each layer procedure of the fuzzy GMDH, the estimated values $Y_1, \ldots, Y_i, \ldots, Y_j, \ldots$ become input variables of the next layer. For example, we can obtain the following model by using input variables Y_i and Y_j:

$$Y_s = A_{0s} + A_{1s} Y_i + A_{2s} Y_j + A_{3s}(Y_i)^2 + A_{4s}(Y_j)^2 + A_{5s} Y_i Y_j, \quad s = 1, 2, \ldots, \quad i, j = 1, 2, \ldots, i \neq j. \tag{16}$$

However, the estimated value Y_s is vague since parameters and input variables are fuzzy numbers, and the width of the fuzzy numbers is large for multiplication of fuzzy numbers. Thus, we use the center values of fuzzy number Y_i and Y_j as input variables, i.e. the model (15) is used.

As the conventional GMDH, we select only good partial descriptions to fit the observed data by using the fitting index in each layer. The algorithm of the multilayer procedure is stopped by some stopping rule with which we cannot select best partial descriptions. Then, the estimated model Y^* is the last best partial description.

Now, let us discuss the fitting index. We propose two fitting indexes between the estimated fuzzy value Y_i^* and the data y_i.

(i) $J_i^1 = \mu_{Y_i^*}(y_i),$ (17)

(ii) $J_i^2 = \boldsymbol{c}\,|\boldsymbol{x}_i|, \quad i = 1, 2, \ldots, N.$ (18)

The index J_i^1 means the degree to which the data is included by the estimated value Y_i^*. The index J_i^2 means the width of Y_i^*. The higher J_i^1 and the smaller J_i^2, the better the estimated value is. Thus, we propose the following performance index J which represents the goodness of fit of the partial description to the observed data.

$$J = \frac{\sum_{i=1}^{N} J_i^2}{\sum_{i=1}^{N} J_i^1} = \frac{\sum_{i=1}^{N} \boldsymbol{c}\,|\boldsymbol{x}_i|}{\sum_{i=1}^{N} \mu_{Y_i^*}(y_i)}. \tag{19}$$

If $J = 0$, i.e. $\boldsymbol{c}\,|\boldsymbol{x}_i| = 0$, the estimated value Y_i^*, $j = 1, 2, \ldots, N$, is equal to a real number y_i^*.

Now, let us explain the structure of the algorithm of the fuzzy GMDH.

Step 1: Choose inputs x_{ij}, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, m$, that are hypothesized to influence the output of interest y_i. If necessary, normalize the observed data (x_i, y_i).
Step 2: Compute the correlation coefficients between the input and output variables for m input variables, and preselect the best n input variables by the values of the correlation coefficients. These selected variables $(x_1, \ldots, x_n)$ are denoted as the input variables to the first layer.
Step 3: Separate data into two sets called the 'training data set' (hereafter TRD) and the 'checking data set' (hereafter CHD), respectively. Note that TRD is used to estimate fuzzy parameters of partial descriptions defined in Step 4 and CHD is used to independently evaluate the partial descriptions. Assume that the numbers of data in TRD and CHD are N_t and N_c, respectively.
Step 4: Form the following partial descriptions constructed by two inputs x_p and x_q:

$$Y_k = A_{0k} + A_{1k}x_p + A_{2k}x_q + A_{3k}(x_p)^2 + A_{4k}(x_q)^2 + A_{5k}x_px_q,$$
$$p, q = 1, 2, \ldots, n,\ p \neq q, \quad k = 1, 2, \ldots, \tfrac{1}{2}n(n-1). \tag{20}$$

The fuzzy parameters in (20) are determined by solving the Min Problem (13). Since the estimated value Y_k is a fuzzy number, the intermediate variables in the next layer are taken as

$$x_k = \{y_k^+ \mid \mu_{Y_k}(y_k^+) = 1\}, \quad k = 1, 2, \ldots, \tfrac{1}{2}n(n-1). \tag{21}$$

Note that the intermediate variable is non-fuzzy, as described before.
Step 5: Calculate the following performance index J_k of the partial description using CHD:

$$J_k = \sum_{i=1}^{N} c\,|x_i| \Big/ \sum_{i=1}^{N} \mu_{Y_{k,i}}(y_i), \tag{22}$$

which represents the goodness of fit of the partial description of CHD and is defined by (19).

By the index J_k, select the best r intermediate variables derived from (21) and (22). These selected x_k, $k = 1, 2, \ldots, r$, become inputs at the next layer. Calculate the threshold in this layer:

$$\theta = \min_k J_k. \tag{23}$$

Step 6: Go to Step 4. Repeat Step 4–6 until the threshold θ_{m+1} in the $(m+1)$-th layer becomes larger than θ_m in the m-th layer:

$$\theta_{m+1} = \min_s J_s \geqslant \theta_m = \min_k J_k, \quad m = 1, 2, \ldots. \tag{24}$$

By repeatedly substituting the intermediate variables into the partial descriptions in the next layer until the final layer, the estimated model Y^* is obtained.

5. Application

To explain the applicability of the algorithm of the fuzzy GMDH, we apply it to structure identification of the amount of production of computers in Japan and identification and prediction of water temperatures in a dam reservoir.

5.1. *Identification of amount of production of computers in Japan*

Table 1 shows input–output data among 1960–1981 [9]. According to the steps of algorithm, let us show the results of this example.

Step 1 and 2: y is an output variable giving the amount of production of computers (in hundred mil. yen). x_j, $j = 1, 2, \ldots, 11$, are input variables preselected by computing the correlation coefficients between 20 input variables and output variables. Input variables are obtained as follows:

x_1: National income (in 10 tril. yen).

x_2: Revenue (in tril. yen).

x_3: Value of shipments of chemicals, petroleum and allied products (in tril. yen).

x_4: Value of shipment of iron, steel and fabricated metal products (in tril. yen).

x_5: Value of shipments of machinery, exclusive electrical (in tril. yen).

x_6: Value of shipments of electrical machinery, equipment and supplies (in tril. yen).

Table 1. Input–output data analysed by the fuzzy GMDH

No	Year	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	TRD/CHD
1	1960	0.1	1.3	3.9	1.7	2.9	1.2	1.3	1.4	2.2	1.2	1.2	1.4	TRD
2	1961	0.2	1.6	4.7	2.1	3.7	1.7	1.7	1.6	2.6	1.4	1.5	1.6	CHD
3	1962	0.5	1.8	5.3	2.4	3.6	1.9	1.9	1.8	3.0	1.7	1.6	1.9	TRD
4	1963	1.0	2.1	5.8	2.8	4.1	2.0	2.0	2.0	3.5	1.9	2.0	2.4	CHD
5	1964	1.5	2.3	6.5	3.2	5.0	2.3	2.3	2.6	4.0	2.3	2.2	2.8	TRD
6	1965	3.2	2.6	7.4	3.4	5.2	2.2	2.3	3.0	4.4	2.8	2.2	3.4	CHD
7	1966	4.9	3.1	8.7	3.9	6.2	2.5	2.7	3.4	5.1	3.4	2.7	3.9	TRD
8	1967	8.7	3.6	10.3	4.6	7.6	3.3	3.6	4.3	6.0	4.0	3.0	4.6	CHD
9	1968	14.1	4.3	12.0	5.3	8.7	4.2	4.6	5.4	7.6	4.7	3.4	5.5	TRD
10	1969	17.1	5.0	14.0	6.3	10.9	5.4	6.0	6.2	8.6	5.7	4.0	6.5	CHD
11	1970	27.0	5.9	16.1	7.3	13.4	6.8	7.3	7.3	10.7	6.7	3.7	8.2	TRD
12	1971	31.1	6.6	18.7	8.1	13.0	7.0	7.5	8.2	12.0	7.3	5.3	9.1	CHD
13	1972	38.2	7.6	22.8	8.5	14.3	7.1	8.6	9.4	13.9	9.3	6.0	10.7	TRD
14	1973	42.8	9.2	28.2	10.5	19.5	9.4	10.5	11.4	16.3	12.4	6.6	13.2	CHD
15	1974	53.6	10.9	34.8	16.6	24.5	11.4	11.9	13.6	21.9	13.0	7.8	16.6	TRD
16	1975	49.7	12.0	39.8	18.0	21.8	10.6	10.8	14.9	24.6	15.1	9.6	18.9	CHD
17	1976	57.5	13.6	48.0	20.5	24.8	11.6	13.7	16.8	26.6	19.5	11.6	21.7	TRD
18	1977	65.9	14.9	57.1	21.4	26.3	12.7	15.1	19.1	27.6	21.3	12.8	18.6	CHD
19	1978	82.0	16.4	67.4	20.7	27.3	13.4	16.2	20.2	27.3	24.2	14.1	20.7	TRD
20	1979	100.6	17.6	77.1	24.7	31.7	15.6	18.5	21.7	26.3	25.7	15.5	23.3	CHD
21	1980	113.7	18.9	72.9	33.1	36.3	17.4	22.0	24.9	27.3	29.2	18.0	25.2	TRD
22	1981	131.5	20.0	94.3	34.0	35.5	19.4	25.7	28.2	29.1	30.3	18.6	26.1	CHD

x_7: Value of shipments of transport equipment and precision instruments (in tril. yen).

x_8: Net domestic product of wholesale and retail trade (in tril. yen).

x_9: Net domestic product of finance, insurance and real estate (in tril. yen.).

x_{10}: Net domestic product of electricity, gas, water, transport and communication (in tril. yen).

x_{11}: Net domestic product of services (in tril. yen).

Step 3: We separate data into TRD and CHD as described in Table 1. Thus, $N_t = 11$ and $N_c = 11$.

Step 4: We assume that a reference function $L(x) = \max(0, 1 - |x|)$. Thus, the membership function of the fuzzy parameter A_j is

$$\mu_{A_j}(a_j) = \begin{cases} 1 - |a_j - \alpha_j|/c_j, & \alpha_j - c_j \leq a_j \leq \alpha_j + c_j, \\ 0, & \text{otherwise}. \end{cases} \tag{25}$$

Fuzzy parameters in (20) are determined by solving the Min Problem (13) with $h = 0$. We obtain 55 partial descriptions $Y_1, \ldots, Y_{55}$, which is the number of combinations of the variables.

Step 5: We set $r = 5$ and the partial descriptions Y_4, Y_5, Y_{28}, Y_{35} and Y_{45} are selected. The values of the index for the partial descriptions are obtained as

$$J_4 = 18.49, \quad J_5 = 17.68, \quad J_{28} = 16.43, \quad J_{35} = 15.68, \quad J_{45} = 16.70.$$

Thus, we have

$$\begin{aligned} &\theta_1 = \min J_k = 15.68, \\ &Y_{35} = A_0 + A_1 x_5 + A_2 x_6 + A_3 (x_5)^2 + A_4 (x_6)^2 + A_5 x_5 x_6, \\ &A_0 = (0, 5.64), \quad A_1 = (0, 0), \quad A_2 = (2.60, 0), \\ &A_3 = (0.18, 0), \quad A_4 = (0, 0), \quad A_5 = (0.02, 0). \end{aligned} \tag{26}$$

Step 6: Go to Step 4. Repeat Step 4 and 5. In the second layer, we obtain

$$\theta_2 = \min_{s=1,2,\ldots,10} J_s = 15.68 = \theta_1.$$

Thus, the algorithm of the fuzzy GMDH is stopped in the second layer. As a result of the algorithm, the estimated model Y_{35} in the first layer is obtained.

Figure 2 shows the observed values and the estimated fuzzy values from 1960 to 1981. The possibility of containment of observed data in the estimated values is shown in Table 2. Since 9 data among 11 CHD are contained by the estimated values, except No. 18 (1977) and No. 22 (1981), a good estimated model is obtained.

5.2. *Identification and prediction of water temperature in a dam reservoir*

Water temperature in dam reservoirs exerts influence on water supply, irrigation and so on. Accordingly, the prediction of water temperature is needed to avoid bad influences. But the fluctuation of water temperature is irregular and vague. Thus, we obtain a structural model by using the fuzzy GMDH.

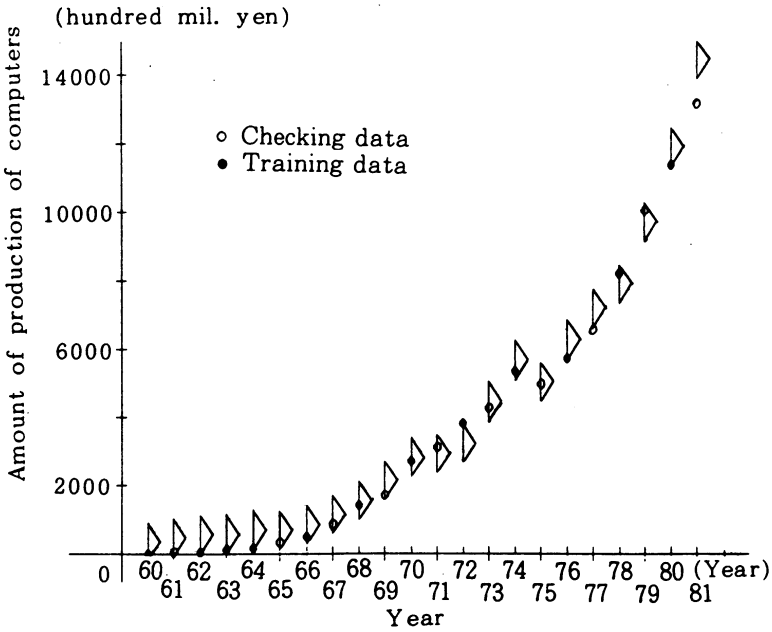

Fig. 2. The amount of production of computers.

Figure 3 shows the data of air temperatures and water temperature in the Daiichi Yahagi Dam in Aichi Prefecture from July 1 to August 31, 1980 [10]. We obtain an estimated model using data from July 1 to July 31, 1980. The estimated model is used to forecast water temperatures from August 1 to August 31, 1980.

Let us explain the algorithm of the fuzzy GMDH according to Section 4.

Step 1 and 2: Table 3 shows input–output relations. y is water temperature at 10 m depth, input variables $x_{j,t}$, $j = 1, 2$, are air temperature and water temperatures at 5 m depth with the time lag at day t, $t = 1, 2, \ldots, 10$, respectively. Input variables in Table 3 are obtained from the correlation coefficients between the input variables with time lag t and output variables. As a result, input variables of the air temperature with $t = 1, 2, 3$ and 10, and the water temperature at 5 m depth with $t = 1, 2, 3, 4$ and 5 are selected.

Step 3: We separate data into $N_t = 20$ and $N_c = 11$ as described in Table 3. N_t is about double N_c, since the estimated model is more fitting for TRD then CHD.

Step 4: Assume that $L(x) = \max(0, 1 - |x|)$. Thus, the membership function of a fuzzy parameter A_j is the same as (25). Fuzzy parameters in (20) are determined by solving Min Problem (13) with $h = 0$. We obtain ${}_9C_2 = 36$ partial descriptions.

Step 5: We set $r = 5$.

Step 6: Go to Step 4. Repeat Step 4 and 5. In the second layer, we obtain

$$\theta_2 = \min_{s=1,\ldots,10} J_s = 0.55 = \theta_1. \tag{27}$$

As a result, we obtain the best estimated model:

$$\begin{aligned} Y = A_0 + A_1 x_{2,1} x_{2,3} + A_2 x_{2,2} x_{2,5} + A_3 (x_{2,2})^2 + A_4 x_{2,1} (x_{2,2})^2 x_{2,3} \\ + A_5 x_{2,1} x_{2,2} x_{2,3} x_{2,5} + A_6 (x_{2,2})^2 (x_{2,5})^2 \end{aligned} \tag{28}$$

Table 2. Observed values and Estimated fuzzy values

Training data						Checking data					
		Estimate fuzzy values						Estimated fuzzy values			
Year	Observed values	Lower	Center	Upper	Possibility	Year	Observed values	Lower	Center	Upper	Possibility
1960	0.1	0	5.7	11.3	0.38	1961	0.2	0	5.7	11.3	0.18
1962	0.5	0	5.7	11.3	0.08	1963	1.0	0.3	5.9	11.6	0.12
1964	1.5	1.5	7.1	12.8	0	1965	3.2	1.3	7.0	12.6	0.33
1966	4.9	2.7	8.4	14.0	0.38	1967	8.7	5.8	11.5	17.1	0.50
1968	14.1	10.0	15.6	21.3	0.74	1969	17.1	15.8	21.4	27.1	0.22
1970	27.0	22.7	28.4	34.0	0.75	1971	31.1	23.8	29.4	35.0	0.70
1972	38.2	26.9	32.6	38.2	0	1973	42.8	39.4	45.0	50.7	0.60
1974	53.6	51.5	57.1	62.8	0.38	1975	49.7	45.1	50.7	56.3	0.83
1976	57.5	57.5	63.1	68.8	0	1977	65.9	66.4	72.1	77.7	0
1978	82.0	73.5	79.1	84.8	0.50	1979	100.6	92.0	97.6	103.7	0.48
1980	113.7	113.7	119.4	125.0	0	1981	131.5	139.4	145.0	150.6	0

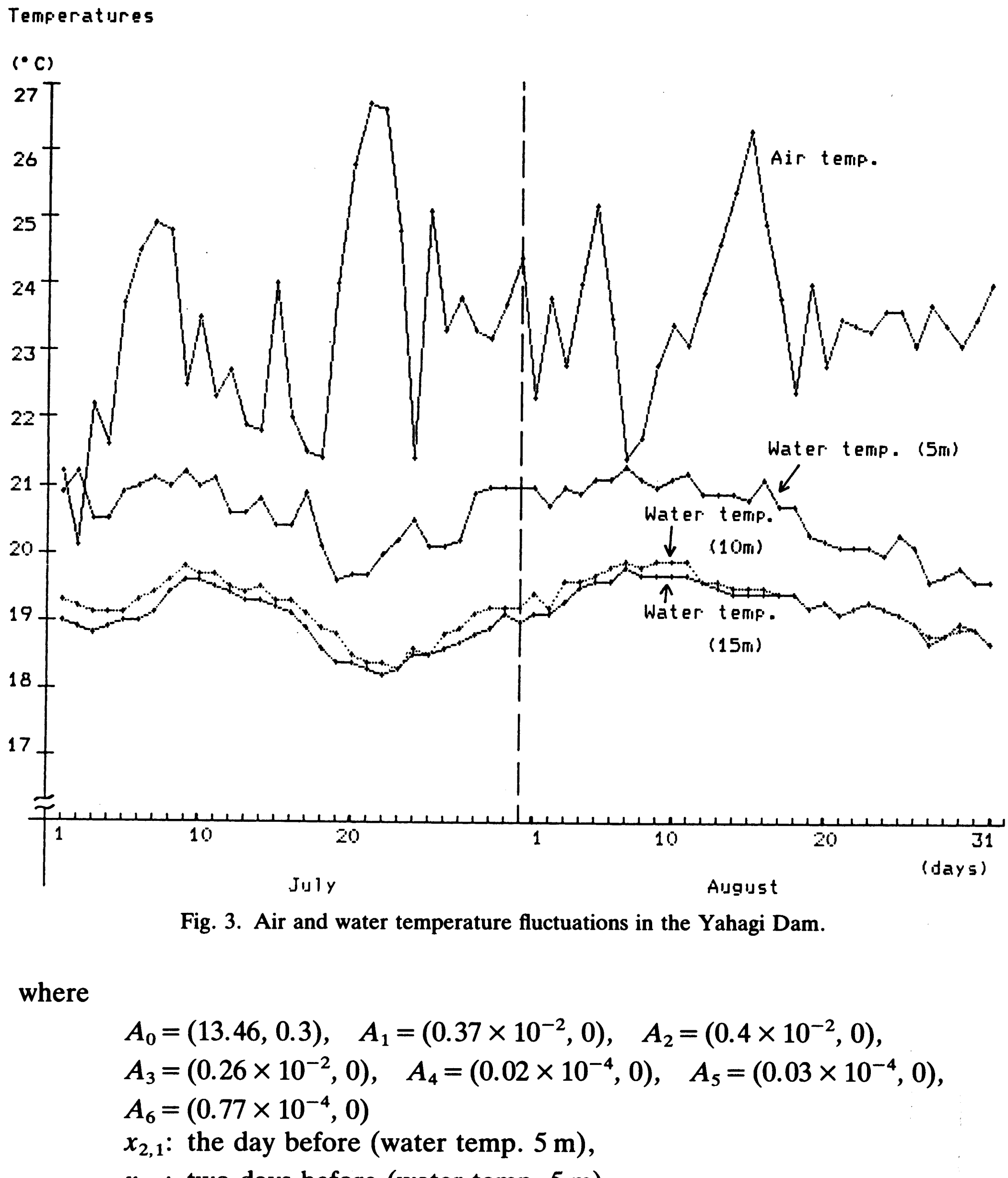

Fig. 3. Air and water temperature fluctuations in the Yahagi Dam.

where

$A_0 = (13.46, 0.3)$, $A_1 = (0.37 \times 10^{-2}, 0)$, $A_2 = (0.4 \times 10^{-2}, 0)$,
$A_3 = (0.26 \times 10^{-2}, 0)$, $A_4 = (0.02 \times 10^{-4}, 0)$, $A_5 = (0.03 \times 10^{-4}, 0)$,
$A_6 = (0.77 \times 10^{-4}, 0)$

$x_{2,1}$: the day before (water temp. 5 m),
$x_{2,2}$: two days before (water temp. 5 m),
$x_{2,3}$: three days before (water temp. 5 m),
$x_{2,5}$: five days before (water temp. 5 m).

Figure 4 shows fluctuations of observed water temperatures and estimated fuzzy values at 10 m depth. The estimated model Y is used to forecast water temperatures at 10 m depth from August 1 to August 31. Figure 5 shows the prediction of water temperatures at 10 m depth. The estimated values contain 22 data among 31 CHD. Since the average width of the estimated fuzzy values is small, 0.3 (°C), a good estimated model is obtained.

Table 3. Input data of air temperature and water temperature at 5 m depth and output data of water temperature at 10 m depth (°C)

	Output data	Input data/Air temp.				Input data/Water temp. at 5 m depth				
July date	Water temp. at 10 m depth	The day before $x_{1,1}$	Two days before $x_{1,2}$	Ten days before $x_{1,10}$	Three days before $x_{1,3}$	The day before $x_{2,1}$	Two days before $x_{2,2}$	Three days before $x_{2,3}$	Four days before $x_{2,4}$	Five days before $x_{2,5}$
1	19.3	21.7	20.4	20.2	23.3	21.0	21.6	20.2	20.7	20.8
2	19.2	21.2	21.7	21.8	20.4	20.9	21.0	21.6	20.2	20.7
3	19.1	20.1	21.2	22.3	21.7	21.2	20.9	21.0	21.6	20.2*
4	19.1	22.2	20.1	24.1	21.2	20.5	21.2	20.9	21.0	21.6
5	19.1	21.6	22.2	23.6	20.1	20.5	20.5	21.2	20.9	21.0
6	19.3	23.7	21.6	23.0	22.2	20.9	20.5	20.5	21.2	20.9*
7	19.4	24.5	23.7	23.3	21.6	21.0	20.9	20.5	20.5	21.2
8	19.6	24.9	24.5	23.3	23.7	21.1	21.0	20.9	20.5	20.5
9	19.8	24.8	24.9	20.4	24.5	21.0	21.1	21.0	20.9	20.5*
10	19.7	22.5	24.8	21.7	24.9	21.2	21.0	21.1	21.0	20.9
11	19.7	23.5	22.5	21.2	24.8	21.0	21.2	21.0	21.1	21.0
12	19.5	22.3	23.5	20.1	22.5	21.1	21.0	21.2	21.0	21.1*
13	19.4	22.7	22.3	22.2	23.5	20.6	21.1	21.0	21.2	21.0
14	19.5	21.9	22.7	21.6	22.3	20.6	20.6	21.1	21.0	21.2
15	19.3	21.8	21.9	23.7	22.7	20.8	20.6	20.6	21.1	21.0*
16	19.3	24.0	21.8	24.5	21.9	20.4	20.8	20.6	20.6	21.1
17	19.1	22.0	24.0	24.9	21.8	20.4	20.4	20.8	20.6	20.6
18	18.9	21.5	22.0	24.8	24.0	20.9	20.4	20.4	20.8	20.6*
19	18.8	21.4	21.5	22.5	22.0	20.1	20.9	20.4	20.4	20.8
20	18.5	24.0	21.4	23.5	21.5	19.6	20.1	20.9	20.4	20.4
21	18.4	25.8	24.0	22.3	21.4	19.7	19.6	20.1	20.9	20.4*
22	18.4	26.7	25.8	22.7	24.0	19.7	19.7	19.6	20.1	20.9
23	18.3	26.6	26.7	21.9	25.8	20.0	19.7	19.7	19.6	20.1
24	18.6	24.8	26.6	21.8	26.7	20.2	20.0	19.7	19.7	19.6*
25	18.5	21.4	24.8	24.0	26.6	20.5	20.2	20.0	19.7	19.7
26	18.8	25.1	21.4	22.0	24.8	20.1	20.5	20.2	20.0	19.7
27	18.9	23.3	25.1	21.5	21.4	20.1	20.1	20.5	20.2	20.0*
28	19.1	23.8	23.3	21.4	25.1	20.2	20.1	20.1	20.5	20.2
29	19.2	23.3	23.8	24.0	23.3	20.9	20.2	20.1	20.1	20.5
30	19.2	23.2	23.3	25.8	23.8	21.0	20.9	20.2	20.1	20.1*
31	19.2	23.7	23.2	26.7	23.3	21.0	21.0	20.9	20.2	20.1*

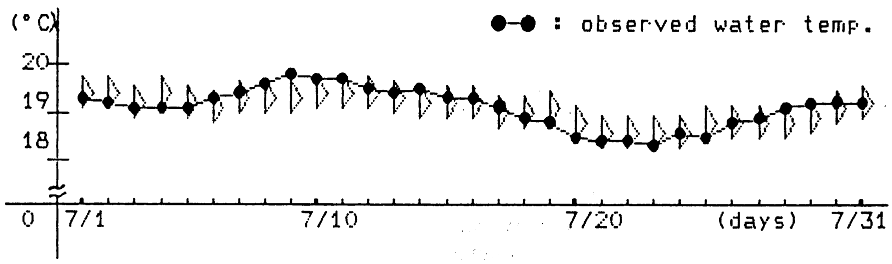

Fig. 4. Observed water temperature and estimated fuzzy numbers at 10 m depth.

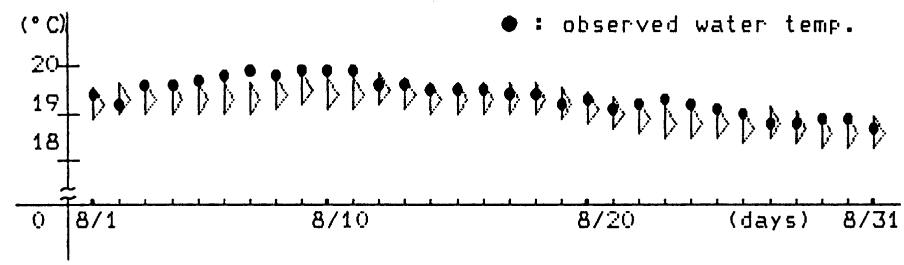

Fig. 5. Observed water temperature and predicted temperature at 10 m depth by the fuzzy GMDH.

6. Conclusion remarks

Conventional regression analysis is constructed by using a probability model whose deviations between the observed data and the estimated values are supposed to be Gaussian distributions. The conventional GMDH is based on regression analysis. On the other hand, possibilistic linear regression is constructed by a possibility model whose parameters are fuzzy numbers. In this paper, the fuzzy GMDH is formulated from the view-point of possibility model. In the algorithm of the fuzzy GMDH, the structure of the model can be automatically found by combining possibilistic linear systems. In order to explain the applicability of the algorithm of the fuzzy GMDH, two practical applications are analyzed by the fuzzy GMDH. As a result, good estimated models are obtained. Thus, the extension of possibilistic linear regression to the fuzzy GMDH can be said to be more efficient in the application aspect than in the theoretical aspect. This is the first paper dealing with the fuzzy GMDH, which will be a powerful analysis method for fuzzy phenomena in the near future.

Acknowledgement

We are indebted to Prof. F. Takagi, Department of Civil Engineering, Nagoya University and Prof. T. Ohono, Department of Civil Engineering, Toyota College of Technology, for their supplying data and discussion.

References

[1] H. Tanaka, S. Uejima and K. Asai, Linear regression analysis with fuzzy model, *IEEE Trans. Systems Man Cybernet.* **12**(6) (1982) 903–907.

[2] H. Tanaka, T. Shimomura, J. Watada and K. Asai, Fuzzy linear regression analysis of the number of staff in local government, *FIP*-84 (Kauai, Hawaii, July 1984).

[3] B. Heshmaty and A. Kandel, Fuzzy linear regression and its applications to forecasting in uncertain environment, *Fuzzy Sets and Systems* **15**(2) (1985) 159–191.

[4] A.G. Ivakhnenko, Polynomial theory complex systems, *IEEE Trans. Systems Man Comput.* **1**(14) (1971) 364–378.

[5] A.G. Ivakhnenko, The group method of data handling; a rival of the method of stochastic approximation, *Soviet Automat. Control* **13**(3) (1968) 43–55.

[6] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications* (Academic Press, New York, 1980).

[7] H. Tanaka, Fuzzy data analysis by possibilistic linear models, *Fuzzy Sets and Systems* **24**(3) (1987) 363–375.

[8] S. Nahmias, Fuzzy variables, *Fuzzy Sets and Systems* **1**(2) (1978) 97–111.

[9] Bureau of Statistics Office of the Prime Minister, *Japan Statistical Yearbook* 11–32 (1960–1981).

[10] I. Hayashi, H. Tanaka, T. Ohono and F. Takagi, Analysis and prediction of water temperature in a reservoir by the fuzzy GMDH, *SICE* **23**(12) (1987) 1304–1311 (in Japanese).