*Scientific Inquiry, vol. 9, No. 1, June, 2008, pp. 61 – 76* *IIGSS Academic Publisher*

# THE FUZZY GROUP METHOD OF DATA HANDLING WITH FUZZY INPUT VARIABLES

ZAYCHENKO Yu.

*Institute of Applied System Analysis, National Technical University of Ukraine "KPI"*
*03056, Kiev-56, Peremogy Avenue, 37, Ukraine*

*(Received January 8, 2007; In final form June 02, 2007)*

The problem of forecasting models constructing using experimental data in terms of fuzziness, when input variables are not known exactly and determined as intervals of uncertainty is considered in this paper. The fuzzy group method of data handling is proposed to solve this problem. The theory of this method was suggested and researched in [1-7]. As is well known, fuzzy GMDH allows to construct fuzzy models and has the following advantages:

1) The problem of optimal model finding is transformed to the problem of linear programming, which is always solvable;
2) There is interval regression model built as the result of method work out;
3) There is a possibility of adaptation of the obtained model;

The mathematical model of the problem mentioned above is built in this article and fuzzy GMDH with fuzzy inputs is elaborated in the paper. The corresponding program, which uses the suggested algorithm, was developed. And also the experimental researches and comparison of FGMDH with GMDH and neural nets in problems of stock prices forecasting was carried out and presented in this article.

*Keywords:* Group method of Data Handling, fuzzy, economic indexes, forecasting

## 1. MATH MODEL OF GROUP METHOD OF DATA HANDLING WITH FUZZY INPUT DATA

### 1.1 General view of FGMDH model with fuzzy input data

Let's consider a linear interval regression model:

$$Y = A_0Z_0 + A_1Z_1 + \ldots + A_nZ_n, \tag{1.0}$$

where $A_i$ are fuzzy numbers, which are described by threes of parameters $A_i = (\underline{A_i}, \breve{A}_i, \overline{A_i})$, where $\breve{A}_i$ – interval center, $\overline{A_i}$ – upper border of the interval, $\underline{A_i}$ - lower border of the interval, and $Z_i$ – also fuzzy numbers, which are determined by parameters $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ - lower border, $\breve{Z}_i$ - center, $\overline{Z_i}$ - upper border of fuzzy number. Then $Y$ – output fuzzy number, which parameters are defined as follows (in accordance with L-R numbers multiplying formulas):


*ISSN 1552-1222© 2008 IIGSS*
*http://www.iigss.net/Scientific-Inquiry/mission.html*

Center of interval: $\breve{y} = \sum \breve{A}_i * \breve{Z}_i$;

Deviation in the left part of the membership function:

$$\breve{y} - \underline{y} = \sum (\left|\breve{A}_i\right| * (\breve{Z}_i - \underline{Z}_i) + (\breve{A}_i - \underline{A}_i) * \left|\breve{Z}_i\right|);$$

And lower border of the interval:

$$\underline{y} = \sum (\breve{A}_i * \breve{Z}_i - \left|\breve{A}_i\right| * (\breve{Z}_i - \underline{Z}_i) - (\breve{A}_i - \underline{A}_i) * \left|\breve{Z}_i\right|);$$

Deviation in the right part of the membership function:

$$\overline{y} - \breve{y} = \sum (\left|\breve{A}_i\right| * (\overline{Z}_i - \breve{Z}_i) + \left|\breve{Z}_i\right| * (\overline{A}_i - \breve{A}_i)).$$

Thus the upper border of the interval is

$$\overline{y} = \sum (\left|\breve{A}_i\right| * (\overline{Z}_i - \breve{Z}_i) + \left|\breve{Z}_i\right| * (\overline{A}_i - \breve{A}_i) + \breve{A}_i * \breve{Z}_i).$$

For the interval model to be correct, the real value of input variable Y is needed to lay in the interval got by the method workflow. So, the general requirements to estimation linear interval model are to find such values of parameters $(\underline{A}_i, \breve{A}_i, \overline{A}_i)$ of fuzzy coefficients, which allow:

a) Observed values $y_k$ lay in estimation interval for $Y_k$;

b) Total width of estimation interval is minimal.

Input data for this task is $Z_k = [\ Z_{ki}]_{\mathrm{i}}$ - input training sample, and also $y_k$ – known output values, $k = \overline{1, M}$, $M$ – the number of observation points. There are two cases of fuzzy values membership function described in this work:

- Triangular membership functions
- Gaussian membership functions.

Quadratic partial descriptions were chosen:

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2.$$

## 1.2. FGMDH with fuzzy input data for triangular membership functions

### 1.2.1. The form of math model for triangular MF

Let's consider the linear interval regression model:

$$Y = A_0 Z_0 + A_1 Z_1 + \ldots + A_n Z_n, \tag{1.1}$$

where $A_i$ – fuzzy number of triangular shape, which is described by threes of parameters $A_i = (\underline{A}_i, a_i, \overline{A}_i)$, where $a_i$ – center of the interval, $\overline{A}_i$ – its upper border, $\underline{A}_i$ - its lower border. Current task contains the case of symmetrical membership function for parameters $A_i$, so they can be described via pair of parameters ($a_i, c_i$).

$\underline{A}_i = a_i - c_i$, $\overline{A}_i = a_i + c_i$, $c_i$ – interval width, $c_i \geq 0$,

$Z_i$ – also fuzzy numbers of triangular shape, which are defined by parameters $(\underline{Z}_i, \breve{Z}_i, \overline{Z}_i)$, $\underline{Z}_i$ - lower border, $\breve{Z}_i$ - center, $\overline{Z}_i$ - upper border of fuzzy number. Then $Y$ – fuzzy number, which parameters are defined as follows: Center of the interval:

$$\breve{y} = \sum a_i * \breve{Z}_i,$$

Deviation in the left part of the membership function:

$$\breve{y} - \underline{y} = \sum (a_i * (\breve{Z}_i - \underline{Z}_i) + c_i |\breve{Z}_i|),$$

Thus, the lower border of the interval:

$$\underline{y} = \sum (a_i * \underline{Z}_i - c_i |\breve{Z}_i|)$$

Deviation in the right part of the membership function:

$$\overline{y} - \breve{y} = \sum (a_i * (\overline{Z}_i - \breve{Z}_i) + c_i |\breve{Z}_i|) = \sum a_i \overline{Z}_i - a_i \breve{Z}_i + c_i |\breve{Z}_i|,$$

So, the upper border of the interval:

$$\overline{y} = \sum (a_i * \overline{Z}_i + c_i |\breve{Z}_i|)$$

For the interval model to be correct, the real value of input variable Y should lay in the interval got by the method workflow. It can be described in such a way:

$$\begin{cases} \sum (a_i * \underline{Z}_{ik} - c_i |\breve{Z}_{ik}|) \le y_k \\ \sum (a_i * \overline{Z}_{ki} + c_i |\breve{Z}_{ik}|) \ge y_k, k = \overline{1, M} \end{cases}$$

where $Z_k$ = [ $Z_k$]$_i$ is input training sample, $y_k$ –known output values, $k = \overline{1, M}$, $M$ – number of observation points. So, the general requirements to estimation linear interval model are to find such values of parameters $(a_i, c_i)$ of fuzzy coefficients, which enable:

a) Observed values $y_k$ lay in estimation interval for $Y_k$;
b) Total width of estimation interval is minimal.

These requirements can be redefined as a task of linear programming:

$$\min_{a_i, c_i} \sum_{k=1}^{M} (\sum (a_i * \overline{Z}_i + c_i |\breve{Z}_i|) - \sum (a_i * \underline{Z}_i - c_i |\breve{Z}_i|)) \quad (1.2)$$

under constraints:

$$\begin{cases} \sum (a_i * \underline{Z}_{ik} - c_i |\breve{Z}_{ik}|) \le y_k \\ \sum (a_i * \overline{Z}_{ki} + c_i |\breve{Z}_{ik}|) \ge y_k, k = \overline{1, M} \end{cases} \quad (1.3)$$

### 1.2.2. Formalized problem formulation in case of triangular membership functions

Let's consider partial description

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2 \tag{1.4}$$

Rewriting it in accordance with the model (1.1) needs such substitution $z_0 = 1$, $z_1 = x_i$, $z_2 = x_j$, $z_3 = x_i x_j$, $z_4 = x_i^2$, $z_5 = x_j^2$. Then math model (1.2)-(1.3) will take the form

$$\begin{aligned}
\min_{a_i, c_i} \; & (2Mc_0 + a_1\sum_{k=1}^{M}(\bar{x}_{ik} - \underline{x}_{ik}) + 2c_1\sum_{k=1}^{M}|\breve{x}_{ik}| + a_2\sum_{k=1}^{M}(\bar{x}_{jk} - \underline{x}_{jk}) + 2c_2\sum_{k=1}^{M}|\breve{x}_{jk}| + \\
& + a_3\sum_{k=1}^{M}(|\breve{x}_{ik}|(\bar{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}|(\bar{x}_{ik} - \underline{x}_{ik})) + 2c_3\sum_{k=1}^{M}|\breve{x}_{ik}\breve{x}_{jk}| + 2a_4\sum_{k=1}^{M}|\breve{x}_{ik}|(\bar{x}_{ik} - \underline{x}_{ik}) + \\
& + 2c_4\sum_{k=1}^{M}\breve{x}_{ik}^2 + 2a_5\sum_{k=1}^{M}|\breve{x}_{jk}|(\bar{x}_{jk} - \underline{x}_{jk}) + 2c_5\sum_{k=1}^{M}\breve{x}_{jk}^2)
\end{aligned}$$

with the following conditions:

$$\begin{aligned}
& a_0 + a_1\underline{x}_{ik} + a_2\underline{x}_{jk} + a_3(-|\breve{x}_{ik}|(\breve{x}_{jk} - \underline{x}_{jk}) - |\breve{x}_{jk}|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}\breve{x}_{jk}) + \\
& + a_4(-2|\breve{x}_{ik}|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}{}^2) + a_5(2|\breve{x}_{jk}|(\breve{x}_{jk} - \underline{x}_{jk}) + \breve{x}_{jk}{}^2) - c_0 - c_1|\breve{x}_{ik}| - \\
& - c_2|\breve{x}_{jk}| - c_3|\breve{x}_{ik}\breve{x}_{jk}| - c_4\breve{x}_{ik}^2 - c_5\breve{x}_{jk}^2 \le y_k
\end{aligned}$$

$$\begin{aligned}
& a_0 + a_1\bar{x}_{ik} + a_2\bar{x}_{jk} + a_3(|\breve{x}_{ik}|(\bar{x}_{jk} - \breve{x}_{jk}) + |\breve{x}_{jk}|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}\breve{x}_{jk}) + a_4(2|\breve{x}_{ik}|(\bar{x}_{ik} - \\
& - \breve{x}_{ik}) - \breve{x}_{ik}{}^2) + a_5(2|\breve{x}_{jk}|(\bar{x}_{jk} - \breve{x}_{jk}) - \breve{x}_{jk}{}^2) + c_0 + c_1|\breve{x}_{ik}| + c_2|\breve{x}_{jk}| + c_3|\breve{x}_{ik}\breve{x}_{jk}| + \\
& c_4\breve{x}_{ik}^2 + c_5\breve{x}_{jk}^2 \ge y_k
\end{aligned}$$

$$c_l \ge 0, \;\; l = \overline{0,5}.$$

As we can see, this is the linear programming problem, but there are still no limitations for non-negativity of variables $a_i$, so we need go to dual problem, introducing dual variables $\{\delta_k\}$ and $\{\delta_{k+M}\}$. Write down dual problem:

$$\max\left(\sum_{k=1}^{M} y_k \cdot \delta_{k+M} - \sum_{k=1}^{M} y_k \cdot \delta_k\right) \tag{1.5}$$

under constraints:

$$\sum_{k=1}^{M}\delta_{k+M} - \sum_{k=1}^{M}\delta_k = 0$$

$$\sum_{k=1}^{M}\bar{x}_{ik} \cdot \delta_{k+M} - \sum_{k=1}^{M}\underline{x}_{ik} \cdot \delta_k = \sum_{k=1}^{M}(\bar{x}_{ik} - \underline{x}_{ik})$$

$$\sum_{k=1}^{M}\bar{x}_{jk} \cdot \delta_{k+M} - \sum_{k=1}^{M}\underline{x}_{jk} \cdot \delta_k = \sum_{k=1}^{M}(\bar{x}_{jk} - \underline{x}_{jk})$$

$$\sum_{k=1}^{M}(|\breve{x}_{ik}|(\bar{x}_{jk}-\breve{x}_{jk})+|\breve{x}_{jk}|(\bar{x}_{ik}-\breve{x}_{ik})-\breve{x}_{ik}\breve{x}_{jk})\cdot\delta_{k+M}- \tag{1.6}$$

$$-\sum_{k=1}^{M}(-|\breve{x}_{ik}|(\breve{x}_{jk}-\underline{x}_{jk})-|\breve{x}_{jk}|(\breve{x}_{ik}-\underline{x}_{ik})+\breve{x}_{ik}\breve{x}_{jk})\cdot\delta_{k}=$$

$$=\sum_{k=1}^{M}(|\breve{x}_{ik}|(\bar{x}_{jk}-\underline{x}_{jk})+|\breve{x}_{jk}|(\bar{x}_{ik}-\underline{x}_{ik}))$$

$$\sum_{k=1}^{M}(2|\breve{x}_{ik}|(\bar{x}_{ik}-\breve{x}_{ik})-\breve{x}_{ik}{}^{2})\cdot\delta_{k+M}-\sum_{k=1}^{M}(-2|\breve{x}_{ik}|(\breve{x}_{ik}-\underline{x}_{ik})+\breve{x}_{ik}{}^{2})\cdot\delta_{k}=$$

$$=\sum_{k=1}^{M}|\breve{x}_{ik}|(\bar{x}_{ik}-\underline{x}_{ik})$$

$$\sum_{k=1}^{M}(2|\breve{x}_{jk}|(\bar{x}_{jk}-\breve{x}_{jk})-\breve{x}_{jk}{}^{2})\cdot\delta_{k+M}-\sum_{k=1}^{M}(-2|\breve{x}_{ik}|(\breve{x}_{jk}-\underline{x}_{jk})+\breve{x}_{jk}{}^{2})\cdot\delta_{k}=$$

$$=\sum_{k=1}^{M}|\breve{x}_{jk}|(\bar{x}_{jk}-\underline{x}_{jk})$$

$$\sum_{k=1}^{M}\delta_{k+M}+\sum_{k=1}^{M}\delta_{k}\le 2M$$

$$\sum_{k=1}^{M}|\breve{x}_{ik}|\cdot\delta_{k+M}+\sum_{k=1}^{M}|\breve{x}_{ik}|\cdot\delta_{k}\le 2\sum_{k=1}^{M}|\breve{x}_{ik}|$$

$$\sum_{k=1}^{M}|\breve{x}_{jk}|\cdot\delta_{k+M}+\sum_{k=1}^{M}|\breve{x}_{jk}|\cdot\delta_{k}\le 2\sum_{k=1}^{M}|\breve{x}_{jk}|$$

$$\sum_{k=1}^{M}|\breve{x}_{ik}\breve{x}_{jk}|\cdot\delta_{k+M}+\sum_{k=1}^{M}|\breve{x}_{ik}\breve{x}_{jk}|\cdot\delta_{k}\le 2\sum_{k=1}^{M}|\breve{x}_{ik}\breve{x}_{jk}| \tag{1.7}$$

$$\sum_{k=1}^{M}\breve{x}_{ik}^{2}\cdot\delta_{k+M}+\sum_{k=1}^{M}\breve{x}_{ik}^{2}\cdot\delta_{k}\le 2\sum_{k=1}^{M}\breve{x}_{ik}^{2},$$

$$\sum_{k=1}^{M}\breve{x}_{jk}^{2}\cdot\delta_{k+M}+\sum_{k=1}^{M}\breve{x}_{jk}^{2}\cdot\delta_{k}\le 2\sum_{k=1}^{M}\breve{x}_{jk}^{2},$$

$$\delta_k\ge 0,\ \delta_{k+M}\ge 0,\ k=\overline{1,M}\,.$$

The task (1.5)-(1.7) can be solved using simplex-method. Having optimal values of dual variables $\{\delta_k\}$, $\{\delta_{k+M}\}$, we easily obtain the optimal values of desired variables $c_i$, $a_i$, $i=\overline{0,5}$, and also a desired fuzzy model for given partial description.

## 2. FGMDH WITH FUZZY INPUT DATA FOR GAUSSIAN MEMBERSHIP FUNCTIONS

Let B be a fuzzy number with Gaussian MF, which looks like:

$$\mu_B(x)=e^{-\frac{1}{2}\cdot\frac{(x-a)^2}{c^2}}$$

This is defined by pair of numbers $\beta = (a, c)$, where $a$ - center, and $c$ - a value which characterizes the width of the interval.

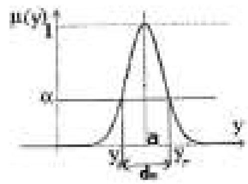

Fig. 1. Gaussian membership function. Level set α.

Level set **α** of fuzzy set $A \subseteq X$ is the set $A_\alpha \subseteq X$ such, that

$$A_\alpha = \{x \in X: \ \mu^A(x) \ge \alpha, \quad \forall \alpha \in [0,1]\cdot$$

### 2.1. Math model in the case of Gaussian MF

Let's consider a linear interval regression model:

$$Y = A_0Z_0 + A_1Z_1 + \ldots + A_nZ_n,$$

where $A_i$ – fuzzy numbers of Gaussian shape, which are described via pair of parameters ($a_i, c_i$), $a_i$ - center, $c_i$ – value characterizing the width of the interval, $c_i \ge 0$,
$Z_i$ – Gaussian fuzzy numbers, which have membership functions of the following view:

$$\mu(x) = \begin{cases} e^{-\frac{1}{2}\cdot\frac{(x-a)^2}{c_1^2}}, x \le a \\ e^{-\frac{1}{2}\cdot\frac{(x-a)^2}{c_2^2}}, \ x > a \end{cases}$$

Then such fuzzy numbers are described by threes of parameters $(c_1, a, c_2)$, where a – center, $c_1$ - value characterizing the width of the interval in the left branch of membership function, $c_2$ - the same thing in the right part. So, the task is to find such fuzzy parameters $(a_i, c_i)$ of fuzzy coefficients $A_i$, that the following conditions would be true:

1) Observation $y_k$ should lay in the estimation interval for $Y_k$ with degree, which is not less than $\alpha$, $0 < \alpha < 1$.
2) The width of the estimation interval of level be minimal.

The width of estimation interval of level $\alpha$ equals (fig. 1.3):

$$d_\alpha = y_2 - y_1 \tag{1.8}$$

We can find $y_1$ and $y_2$ from the system:

$$\begin{cases} \alpha = \exp\left\{-\frac{1}{2}\cdot\frac{(y_2-a)^2}{{c_2}^2}\right\} \\ \alpha = \exp\left\{-\frac{1}{2}\cdot\frac{(y_1-a)^2}{{c_1}^2}\right\} \end{cases} \quad (1.9)$$

So,

$$\begin{cases} y_2 = c_2\sqrt{-2\ln\alpha} + a \\ y_1 = -c_1\sqrt{-2\ln\alpha} + a \end{cases} \quad \text{and} \quad d_\alpha = \sqrt{-2\ln\alpha}(c_2 - c_1)$$

Condition 1) can be defined as: $\mu(y_k) \geq \alpha$, which can be rewritten as follows:

$$\begin{cases} y_k \leq a_k + c_{2k}\sqrt{-2\ln\alpha} \\ y_k \geq a_k - c_{1k}\sqrt{-2\ln\alpha} \end{cases}$$

So, the task is:

$$\min\sum_{k=1}^{M} d_\alpha^k = \min\sum_{k=1}^{M}(c_{2k} - c_{1k})\sqrt{-2\ln\alpha} \quad (1.10)$$

$$\begin{cases} y_k \leq a_k + c_{2k}\sqrt{-2\ln\alpha} \\ y_k \geq a_k - c_{1k}\sqrt{-2\ln\alpha} \end{cases} \quad (1.11)$$

Variables $Z_i$ may be defined via parameters $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ - lower border, $\breve{Z}_i$ - center, $\overline{Z_i}$ - upper border of fuzzy number, where

$$\underline{Z_i} = \breve{Z}_i - c_{1i},\ \overline{Z_i} = \breve{Z}_i + c_{2i}.$$

Thus $Y$ – fuzzy number, which parameters are defined as follows:
Center of the interval:

$$\breve{y} = \sum a_i * \breve{Z}_i,$$

Deviation in the left part of membership function:

$$\breve{y} - \underline{y} = \sum(a_i * (\breve{Z}_i - \underline{Z}_i) + c_i|\breve{Z}_i|)$$

And the lower border of the interval

$$\underline{y} = \sum(a_i * \underline{Z}_i - c_i|\breve{Z}_i|)$$

Deviation in the right part of membership function:

$$\overline{y} - \breve{y} = \sum(a_i * (\overline{Z}_i - \breve{Z}_i) + c_i|\breve{Z}_i|) = \sum a_i\overline{Z}_i - a_i\breve{Z}_i + c_i|\breve{Z}_i|$$

So, the upper border of the interval:

$$\bar{y} = \sum (a_i * \bar{Z}_i + c_i |\breve{Z}_i|)$$

**2.2. Formalized problem formulation in the case of Gaussian MF**

Partial description

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2 \cdot$$

Let's rewrite it in accordance with model (1.1). It needs such definitions $z_0 = 1$, $z_1 = x_i$, $z_2 = x_j$, $z_3 = x_i x_j$, $z_4 = x_i^2$, $z_5 = x_j^2$. So the math model (1.10)-(1.11) will take the form:

$$\min_{a_i, c_i} \ (2Mc_0 + a_1 \sum_{k=1}^{M} (\bar{x}_{ik} - \underline{x}_{ik}) + 2c_1 \sum_{k=1}^{M} |\breve{x}_{ik}| + a_2 \sum_{k=1}^{M} (\bar{x}_{jk} - \underline{x}_{jk}) + 2c_2 \sum_{k=1}^{M} |\breve{x}_{jk}| +$$
$$+ a_3 \sum_{k=1}^{M} (|\breve{x}_{ik}| (\bar{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}| (\bar{x}_{ik} - \underline{x}_{ik})) + 2c_3 \sum_{k=1}^{M} |\breve{x}_{ik} \breve{x}_{jk}| + 2a_4 \sum_{k=1}^{M} |\breve{x}_{ik}| (\bar{x}_{ik} - \underline{x}_{ik}) +$$
$$+ 2c_4 \sum_{k=1}^{M} \breve{x}_{ik}^2 + 2a_5 \sum_{k=1}^{M} |\breve{x}_{jk}| (\bar{x}_{jk} - \underline{x}_{jk}) + 2c_5 \sum_{k=1}^{M} \breve{x}_{jk}^2 )$$

under conditions:

$$a_0 + a_1 (\breve{x}_{ik} - \sqrt{-2\ln\alpha} (\breve{x}_{ik} - \underline{x}_{ik})) + a_2 (\breve{x}_{jk} - \sqrt{-2\ln\alpha} (\breve{x}_{jk} - \underline{x}_{jk})) + a_3 (\breve{x}_{ik} \breve{x}_{jk} -$$
$$- (|\breve{x}_{ik}| (\breve{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}| (\breve{x}_{ik} - \underline{x}_{ik})) \sqrt{-2\ln\alpha}) + a_4 (\breve{x}_{ik}{}^2 - 2\sqrt{-2\ln\alpha} (|\breve{x}_{ik}| (\breve{x}_{ik} - \underline{x}_{ik})) +$$
$$+ a_5 (\breve{x}_{jk}{}^2 - 2\sqrt{-2\ln\alpha} (|\breve{x}_{jk}| (\breve{x}_{jk} - \underline{x}_{jk}))) - c_0 \sqrt{-2\ln\alpha} - c_1 |\breve{x}_{ik}| \sqrt{-2\ln\alpha} -$$
$$- c_2 |\breve{x}_{jk}| \sqrt{-2\ln\alpha} - c_3 |\breve{x}_{ik} \breve{x}_{jk}| \sqrt{-2\ln\alpha} - c_4 \breve{x}_{ik}^2 \sqrt{-2\ln\alpha} - c_5 \breve{x}_{jk}^2 \sqrt{-2\ln\alpha} \le y_k$$

$$a_0 + a_1 (\breve{x}_{ik} + \sqrt{-2\ln\alpha} (\bar{x}_{ik} - \breve{x}_{ik})) + a_2 (\breve{x}_{jk} + \sqrt{-2\ln\alpha} (\bar{x}_{jk} - \breve{x}_{jk})) + a_3 (\breve{x}_{ik} \breve{x}_{jk} +$$
$$+ (|\breve{x}_{ik}| (\bar{x}_{jk} - \breve{x}_{jk}) + |\breve{x}_{jk}| (\bar{x}_{ik} - \breve{x}_{ik})) \sqrt{-2\ln\alpha}) + a_4 (\breve{x}_{ik}{}^2 + 2\sqrt{-2\ln\alpha} (|\breve{x}_{ik}| (\bar{x}_{ik} - \breve{x}_{ik})) +$$
$$+ a_5 (\breve{x}_{jk}{}^2 + 2\sqrt{-2\ln\alpha} (|\breve{x}_{jk}| (\bar{x}_{jk} - \breve{x}_{jk})) + c_0 \sqrt{-2\ln\alpha} + c_1 |\breve{x}_{ik}| \sqrt{-2\ln\alpha} +$$
$$+ c_2 |\breve{x}_{jk}| \sqrt{-2\ln\alpha} + c_3 |\breve{x}_{ik} \breve{x}_{jk}| \sqrt{-2\ln\alpha} + c_4 \breve{x}_{ik}^2 \sqrt{-2\ln\alpha} + c_5 \breve{x}_{jk}^2 \sqrt{-2\ln\alpha} \ge y_k$$

$$c_l \ge 0, \ \ l = \overline{0,5} \cdot$$

As we can see, this is the same problem of linear programming as in case of triangular MF, but there are still no limitations for non-negativity of variables $a_i$, so we need to go to dual problem, introducing dual variables $\{\delta_k\}$ and $\{\delta_{k+M}\}$.

## 3. RESULTS OF FGMDH WITH FUZZY INPUT DATA WORKFLOW IN RTS INDEX FORECASTING

For estimation of efficiency of the suggested FGMDH method with fuzzy inputs the corresponding software kit was elaborated and numerous experiments of financial markets forecasting were carried out. Some of them are presented below.

### 3.1. Forecasting of RTS index.

#### 3.1.1. Experiment 1. RTS index forecasting (opening price)

In this experiment we used 5 fuzzy input variables, which represent stock prices of leading Russian energetic companies, which are included to the list of computations of RTS index:

LKOH – shares of "LUKOIL" joint-stock company,
EESR – shares of "РАО ЕЭС России"joint-stock company,
YUKO – shares of "ЮКОС" joint-stock company,
SNGSP – privileged shares of "Сургутнефтегаз" joint-stock company,
SNGS – common shares of "Сургутнефтегаз" joint-stock company.

Output variable is the RTS (opening price) index value of the same period (03.04.2006 – 18.05.2006). Sample size – 32 values. Training sample size – 18 values (optimal size of training sample for current experiment). The following results were obtained:

1. For triangular membership function
a) For normalized input data, normalizing was done using the following formula:

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, X_i \in [0;1].$$

Table 3.1. Experiment 1 results with triangular MF and normalized input data

| Date | Real value of Y | Forecasted value Y | | | Deviation | Square deviation |
|---|---|---|---|---|---|---|
| | | Lower border | Center | Upper border | | |
| 03.04.2006 | 0 | -0.26998 | -0.01999 | 0.2102 | 0.2102 | 0.04418404 |
| 04.04.2006 | 0.049767092 | -0.180273 | 0.0716071 | 0.341937 | 0.29217 | 0.085363255 |
| 05.04.2006 | 0.072967693 | -0.0554223 | 0.162758 | 0.377578 | 0.30461 | 0.092787439 |
| 06.04.2006 | 0.153087904 | -0.101492 | 0.103928 | 0.288868 | 0.13578 | 0.018436235 |
| 07.04.2006 | 0.176679189 | 0.0563192 | 0.243239 | 0.428249 | 0.25157 | 0.06328737 |
| 10.04.2006 | 0.268099174 | -0.0564308 | 0.195679 | 0.460669 | 0.19257 | 0.037083138 |
| 11.04.2006 | 0.366130729 | 0.266021 | 0.445581 | 0.596351 | 0.23022 | 0.053001373 |
| 12.04.2006 | 0.354259955 | 0.19167 | 0.43254 | 0.69384 | 0.33958 | 0.115314607 |
| 13.04.2006 | 0.382779865 | 0.15828 | 0.393 | 0.61536 | 0.23258 | 0.054093519 |
| 14.04.2006 | 0.337761082 | 0.230051 | 0.397011 | 0.603071 | 0.26531 | 0.070389353 |
| 17.04.2006 | 0.373553719 | 0.138454 | 0.410774 | 0.648944 | 0.27539 | 0.075839807 |
| 18.04.2006 | 0.447753569 | 0.232374 | 0.467904 | 0.734234 | 0.28648 | 0.082071037 |
| 19.04.2006 | 0.49959429 | 0.192934 | 0.429164 | 0.667664 | 0.16807 | 0.028247427 |
| 20.04.2006 | 0.542118708 | 0.318519 | 0.525329 | 0.736669 | 0.19455 | 0.037849816 |
| 21.04.2006 | 0.545484598 | 0.481095 | 0.640635 | 0.801735 | 0.25625 | 0.065664269 |
| 24.04.2006 | 0.552667168 | 0.450827 | 0.632127 | 0.802057 | 0.24939 | 0.062195289 |
| 25.04.2006 | 0.566190834 | 0.358941 | 0.624711 | 0.883731 | 0.31754 | 0.100831757 |
| 26.04.2006 | 0.579744553 | 0.319645 | 0.563455 | 0.819935 | 0.24019 | 0.057691451 |
| 27.04.2006 | 0.698302029 | 0.446212 | 0.659722 | 0.856082 | 0.15778 | 0.024894519 |
| 28.04.2006 | 0.569105935 | 0.385086 | 0.641006 | 0.867356 | 0.29825 | 0.088953101 |
| 02.05.2006 | 0.66945154 | 0.471102 | 0.742042 | 0.995382 | 0.32593 | 0.106230665 |
| 03.05.2006 | 0.84730278 | 0.660633 | 0.860383 | 1.06418 | 0.216877 | 0.047035729 |
| 04.05.2006 | 0.868309542 | 0.6836 | 0.89258 | 1.08695 | 0.21864 | 0.04780365 |
| 05.05.2006 | 0.919729527 | 0.7516 | 0.96899 | 1.17836 | 0.25863 | 0.066889722 |
| 06.05.2006 | 0.946025545 | 0.686906 | 0.935746 | 1.17631 | 0.230284 | 0.05303093 |
| 10.05.2006 | 1 | 0.8311 | 1.05219 | 1.29725 | 0.29725 | 0.088357563 |
| 11.05.2006 | 0.917986476 | 0.711086 | 0.977196 | 1.20893 | 0.290944 | 0.084648134 |
| 12.05.2006 | 0.869181067 | 0.675541 | 0.846741 | 1.04664 | 0.177459 | 0.031491673 |
| 15.05.2006 | 0.739444027 | 0.422684 | 0.699854 | 0.978754 | 0.23931 | 0.057269263 |
| 16.05.2006 | 0.457190083 | 0.25151 | 0.47063 | 0.66686 | 0.20967 | 0.043961474 |
| 17.05.2006 | 0.453042825 | 0.196883 | 0.478753 | 0.720013 | 0.26697 | 0.071273074 |
| 18.05.2006 | 0.307257701 | 0.113418 | 0.397508 | 0.690848 | 0.38359 | 0.147141518 |

Criterion value for current experiment were: MSE = 0.055557

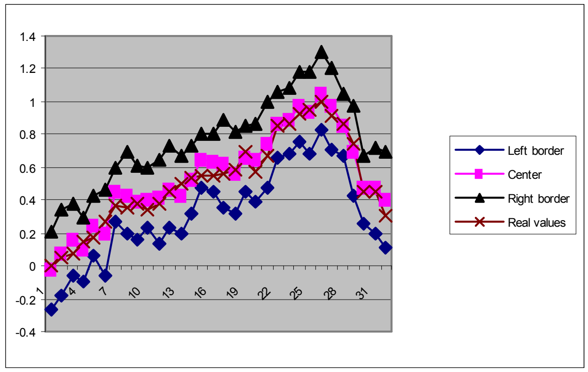


Fig. 2. Experiment 1 results for triangular membership function and normalized values of input variables

b) for non-normalized input data: Criterion values for this experiment were: MSE = 18.48657, MAPE = 0.8%

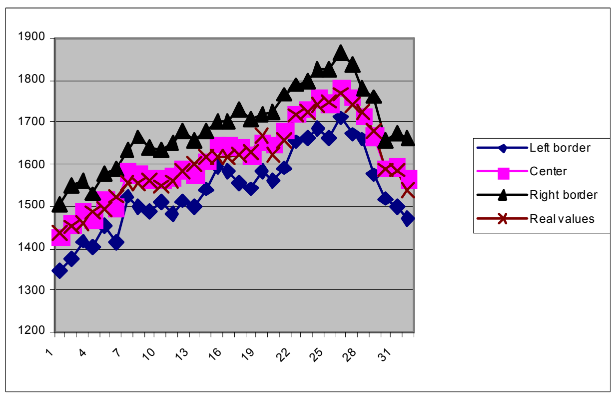


Fig. 3. Experiment 1 result for triangular MF and non-normalized values of input variables

2. For the case of Gaussian membership function (optimal level is α=0.8)
a) For normalized input data: Criterion values for this experiment were: MSE = 0.028013

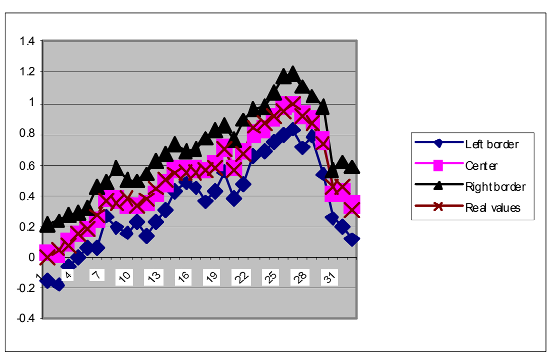


Fig. 4. Experiment 1 result for Gaussian MF and normalized input data

b) for non-normalized input data:

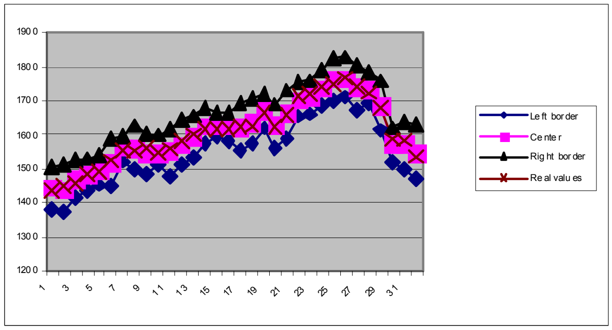


Fig. 5. Experiment 1 result with Gaussian MF and non-normalized values of input variables

Criterion values for current experiment were: MSE = 9.321461, MAPE = 0.4%.

As we can see from the results of experiment 1, forecasting using triangular and Gaussian membership functions gives good results. Results of experiments with Gaussian MF are better than results of experiments with triangular MF.

For normalized data and non-normalized data, we have

| | Triangular MF | Gaussian MF |
|---|---|---|
| MSE | 0.055557 | 0.028013 |

and

| | Triangular MF | Gaussian MF |
|---|---|---|
| MSE | 18.48657 | 9.321461 |
| MAPE | 0.8% | 0.4% |

### 3.1.2. Experiment 2. Forecasting of RTS index (closing price)

This experiment uses the same input variables as the experiment 1 does. Output variable is the value of RTS index (closing price) for the same period (03.04.2006 – 18.05.2006). Sample size – 32 values. Training sample size – 18 values (optimal size of training sample for current experiment). The following results were obtained:

1. For triangular membership function
 a) For normalized input data: Criterion value: MSE = 0.057379.

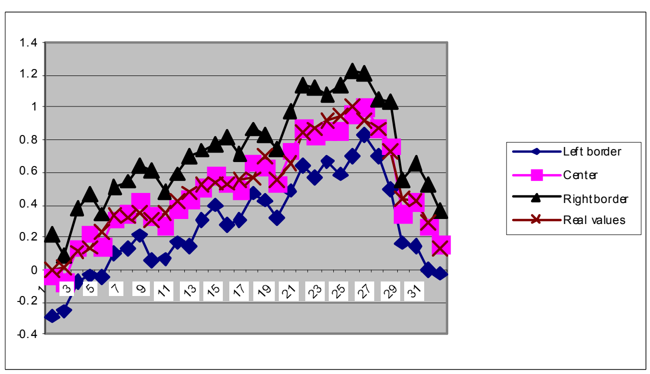


Fig. 6. Experiment 2 result for triangular MF and normalized values of input variables

b) For non-normalized input data: Criterion values:  MSE = 18.04394, MAPE =0.78%.

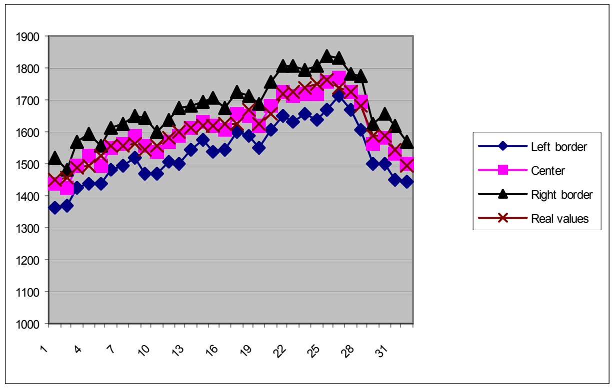


Fig. 7. Experiment 2 result for triangular MF and non-normalized values of input variables

1. For Gaussian membership function (optimal level α=0.85)
a) For normalized input data: Criterion value for current experiment was: MSE = 0.029582.

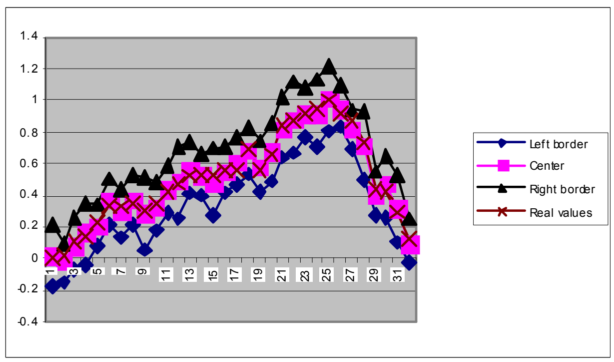


Fig. 8. Experiment 2 result for Gaussian MF and normalized values of input variables

b) For non-normalized input data: Criterion values for this experiment: MSE = 9.302766; MAPE =0.37%.

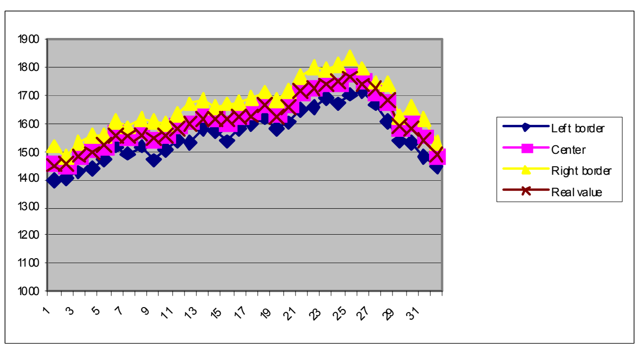


Fig. 9. Experiment 2 result for Gaussian MF and non-normalized values of input variables

As we can see from the results of experiment 2, forecasting using triangular and Gaussian membership functions gives good results. Results of experiments with Gaussian MF are better than results of experiments with triangular MF.

For normalized data and non-normalized data, we have

| | Triangular MF | Gaussian MF |
|---|---|---|
| MSE | 0.057379 | 0.029582 |

and

| | Triangular MF | Gaussian MF |
|---|---|---|
| MSE | 18.04394 | 9.302766 |
| MAPE | 0.78% | 0.37% |

### 3.2. Stock price forecasting

#### 3.2.1 Experiment 4. Stock price forecasting

The following experiment uses stock prices of 4 leading energetic companies of Russia:

EESR – shares of "РАО ЕЭС России" joint-stock company,
YUKO – shares of "ЮКОС" joint-stock company,
SNGSP – privileged shares of "Сургутнефтегаз" joint-stock company,
SNGS – ordinary shares of "Сургутнефтегаз" joint-stock company.

Stock price of other company – "LUKOIL" joint-stock for the same period (03.04.2006 – 18.05.2006) was also forecasted. Sample size – 32 values. Training sample size – 17 values (optimal size of training sample for this experiment). The following results were obtained:

1. For triangular membership function

a) For normalized input data: Criterion value: MSE = 0.056481

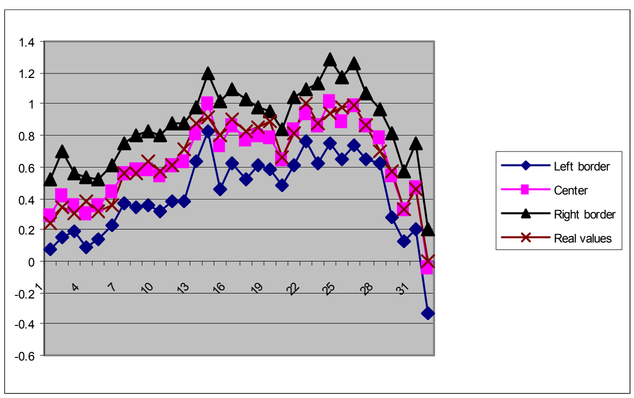


Fig. 10. Experiment 4 results for triangular MF and normalized values of input variables

b) for non-normalized input data: Criterion values: MSE = 0.914998; MAPE = 0.73%.

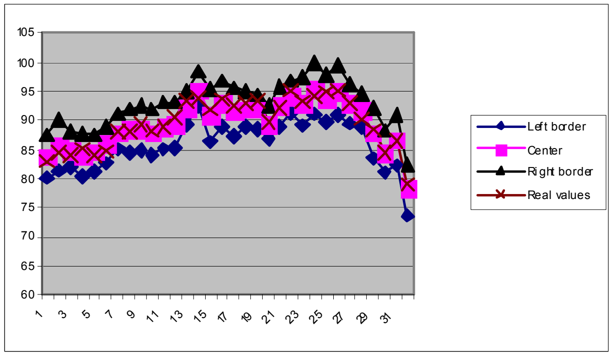


Fig. 11. Experiment 4 results for triangular MF and non-normalized values of input variables

2. For Gaussian membership function (optimal level of α=0.9)

a) For normalized input data: Criterion value for this experiment: MSE = 0.030464

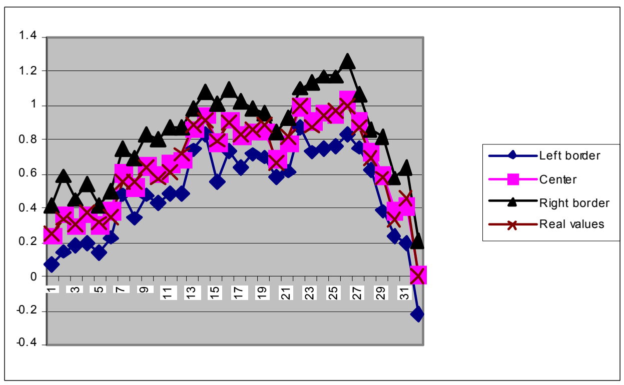


Fig. 12. Experiment 4 results for Gaussian MF and normalized values of input variables

b) For non-normalized input data: Criterion values for this experiment were: MSE = 0.493511, MAPE = 0.33% .

As we can see from the results of experiment 4, forecasting using triangular and Gaussian membership functions gives good results. Results of experiments with Gaussian MF are better than results of experiments with triangular MF.

For normalized data and non-normalized data, we have

| | Triangular MF | Gaussian MF |
|---|---|---|
| MSE | 0.056481 | 0.030464 |

and

| | Triangular MF | Gaussian MF |
|---|---|---|
| MSE | 0.914998 | 0.493511 |
| MAPE | 0.73% | 0.33% |

## 4. CONCLUSIONS

In this article new method of inductive modeling FGMDH with fuzzy inputs was suggested. This method represents the development of fuzzy GMDH when information is fuzzy and given in the form of uncertainty intervals. The mathematical model was constructed and corresponding algorithm was elaborated. The experimental results of application of the suggested method in the forecasting of market index and stock prices are presented and discussed. The main advantages of the suggested method are following:

- It operates with fuzzy and uncertain input information and constructs the fuzzy model;
- The constructed model has minimal possible total width and in this sense it is optimal;
- For finding optimal model we solve corresponding linear programming problem which is always solvable for this task;
- We should not a priori set the form of a model the algorithm finds it itself using the ideas of evolution.

## References

Zaychenko Yu. "The Fuzzy Group Method of Data Handling and Its Application for Economical Processes Forecasting" - *Scientific Inquiry,* - Vol. 7, No.1, June, 2006 - p.83-96.

Zaychenko Yu. "Fuzzy method of inductive modeling in problems of macroeconomic indexes forecasting." *System researches and informational technologies,* #3 of 2003, p. 25-45.

Zaychenko Yu. P., and Zayetz I.O. "The synthesis and adaptation of fuzzy forecasting model based of self-organization method. *Science News of NTUU "KPI",* #2 of 2001.

Zaychenko Yu. P., and Zayetz I.O. "Research of different types of partial descriptions in problems of synthesis of fuzzy forecasting models", *Science works of Donetsk NTU,* vol. 47, p. 341-349.

Zaychenko Yu. P., Zayetz I.O., O.V. Kamotsky, O.V. Pavlyuk. Research of different kinds of membership functions of fuzzy forecasting models parameters in fuzzy group method of data handling. *USiM,* 2003, #2, p.56-67.

Zaychenko Yu. P. and Zayetz I.O. Comparative analysis of GMDH algorithms using different method of single-step adaptation of coefficients. *The NTUU Herald.*

Zaychenko Yu. P. Comparative analysis of forecasting models built using distinct and fuzzy GMDH with different algorithms of fuzzy forecasting models generation. *Materials of international seminar of inductive modeling IWIM 2005.*