

# Inductive Method of Optimal Model Selection by External Error Criterion with Additional Determination Using Bias

Ivakhnenko A.G., Savchenko E.A., Somina L.P.

*International Research and Educational Center of Information Technologies and Systems of NAS of Ukraine, pr. Akademika Glushkova, 40, Kyiv, 03680, Ukraine*

Savchenko@irtc.org.ua

**Abstract.** *The problem of optimal model selection often occurs in a real problems of modeling, because there can be a several accurate models. For this case the method of determination of model by additional bias criterion is proposed. At first the criterion of regularity is calculated, then if the optimal model is not possible to select, the additional bias criterion is calculated for the models which fall into interval of uncertainty. The examples of models determination are shown in problems of selection of aircrafts surface material, forecasting of glucose level for diabetes patients and prediction of solar activity by Wolf's numbers.*

## Keywords

Inductive modelling, GMDH, bias criterion, additional determination, forecasting.

## 1 Introduction

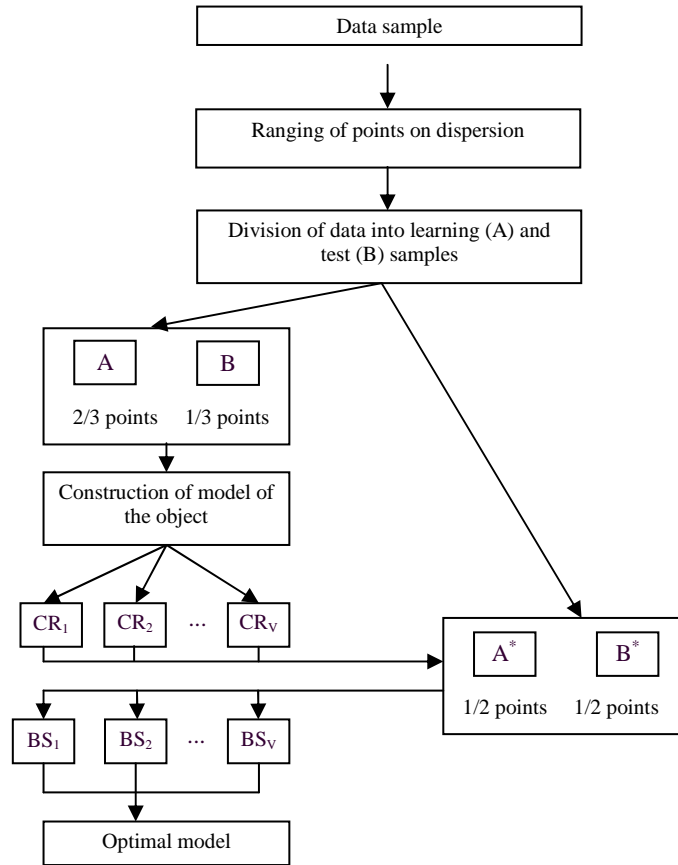
The aim of this work is to develop the method for selection of the optimal model from a set of the most accurate ones in complex systems of different physical nature. For this is used additional models determination by the bias criterion. That is dependence of a materials atomization coefficient from its physical properties, monitoring of diabetes and monitoring of Wolf's numbers.

The method of additional determination of model by bias criterion is considered in Section 2. A numerical example is shown in Section 3. The laws governing the interaction of ions with a surface by the Combinatorial GMDH algorithm with determination by bias are found in Section 4. The process of modelling on the basis of GMDH with determination of process for medical treatment of diabetes described. Also is shown the method for modelling in two-layered monitoring systems with trend definition by example of solar activity modelling.

In real modelling, when data contains noise, difficult problems can appear while searching for the optimal model of some complex system. That is, not only one model is selected by minimum of criterion, but some interval may contain several the most accurate models. In this case we propose to use the method of additional determination, which consists of consecutive models estimation by several criteria.

## 2 Method of models determination by the bias criterion

The idea of models determination or regularization in the GMDH algorithms was explored before in 1970s, when its efficiency was shown [1]. But then the method of additional determination was not developed. In this paper is described the determination of models by Combinatorial GMDH algorithm [2]. The developed algorithm of models determination is shown in Figure 1.



**Figure 1.** Algorithm of additional determination of optimal model.

At first step regularity criterion is computed by the equation:

$$CR_{A/B} = \frac{1}{N_B} \sum (y_{Bi} - \hat{y}_{(A/B)i}) \rightarrow \min \quad (1)$$

where  $y_{Bi}$  - real values of output on a test sample  $B$ ;  $\hat{y}_{(A/B)i}$  - the model values on a test sample  $B$ , coefficients of which are estimated on a learning sample  $A$ ;  $N_B$  - number of observations in a test sample  $B$ .

Then the criterion of bias is calculated for the models which belongs to the interval of uncertainty by the equation:

$$BS = |CR_{A^*/A^*+B^*} + CR_{B^*/A^*+B^*}| \rightarrow \min \quad (2)$$

where  $CR_{A^*/A^*+B^*}$  - total model error on all sample  $(A^* + B^*)$ , the coefficients of which are estimated on a sample  $A$ ;

$CR_{B^*/A^*+B^*}$  - total model error on all sample ( $A^* + B^*$ ), the coefficients of which are estimated on a sample  $B$ . On the Figure 2 the example of successive choice of model by two criteria is shown.

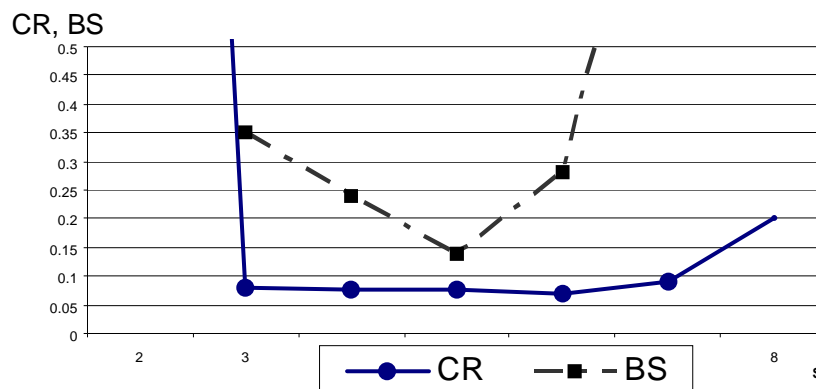


Figure 2. Successive application of two criteria: regularity and bias.

### 3 Numerical example of model determination

Let some model was set as:

$$y = -3x_3 + 7x_7 + \varepsilon \quad (3)$$

It is required to restore this dependence by data sample using GMDH algorithm with additional model determination. The output variable contains 10% noise in the data. Arguments  $X_1 - X_8$  were generated randomly and shown in the table 1. It is necessary to find the model with optimal structure.

Table 1. Data sample.

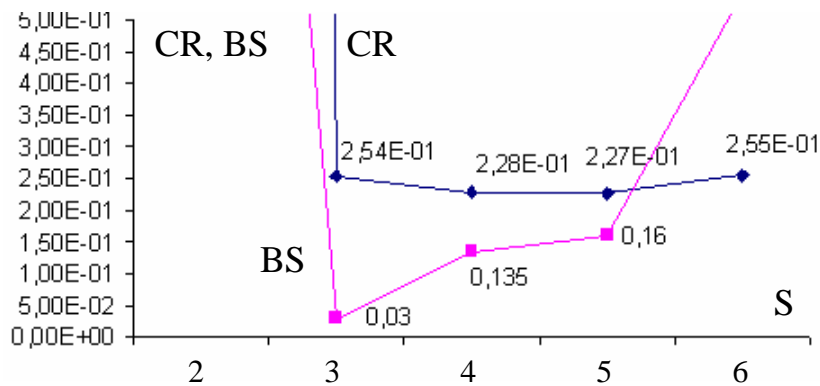
N <sub>o</sub>	X1	X2	X3	X4	X5	X6	X7	X8	Y
1	-3.300	-2.700	4.100	3.800	-0.750	-3.300	-4.900	-4.300	-46.000
2	1.700	0.029	-4.000	3.000	-0.570	1.100	4.200	-0.470	41.000
3	0.980	4.700	-1.200	-1.600	0.940	-2.500	4.700	-1.700	37.000
	...		...		...				
30	1.400	-4.800	4.900	1.900	4.900	4.500	0.610	-0.410	-10.000

The results of search of models by Combinatorial GMDH algorithm with the external criterion (CR) of regularity are shown in the Table 2.

Table 2. The models found by Combinatorial GMDH algorithm.

S	CR	BS	Models
3	0,254	7,1	$y = 0,055 - 2,948x_3 - 6,980 x_7$
4	0,228	7,48	$y = 0,068 - 2,960x_3 + 6,982 x_7 - 0,022 x_8$
5	0,227	8,38	$Y = 0,05 - 2,95 x_3 - 0,032x_5 + 6,987 x_7 - 0,022x_8$
6	0,255	12,4	$y = 0,035 - 2,94 x_3 - 0,452x_4 - 0,285x_5 + 6,97 x_7 - 0,026x_8$

Table 2 shows four models ( $S=3,4,5,6$ ), which get into the interval of uncertainty and have similar regularity criterion CR values. For these models bias criterion BS was calculated by equation (2). The results of computation are shown in Figure 3.



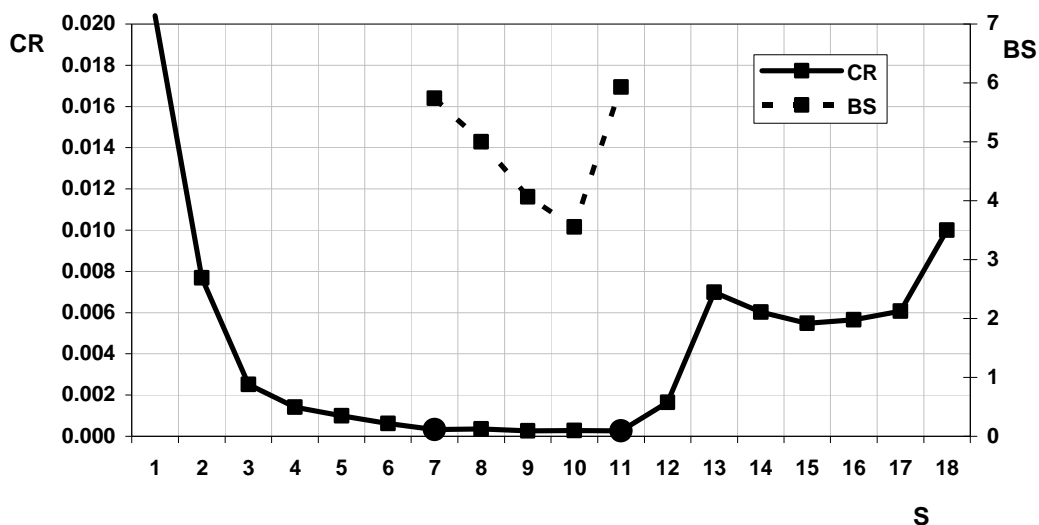
**Figure 3.** Dependence of accurate (CR) and bias (BS) criterion on model complexity S.

Thus the minimum of bias criterion indicates the optimal model structure when  $S = 3$ . The method described above allows us to find the dependence, given before by the minimum of bias criterion.

#### 4 Detection of dependencies of ion interaction with a surface by the Combinatorial GMDH algorithm with additional determination

The purpose of this example is to find the dependence of atomization coefficient from the physical properties of materials (molecular weight, sublimation temperature, heat capacity and energy of connection) by GMDH algorithm with additional determination.

Using the Combinatorial GMDH algorithm to determine the optimal model, the dependence of external regularity criterion from model complexity was founded. It has an interval of uncertainty which contain five models with almost equal values of regularity criterion. For each of these models the values of the bias criterion were calculated (Figure 4).



**Figure 4.** Dependence of regularity CR and bias BS criteria from models complexity S.

The following optimal model was received:

$$y = 5.856 x_6 + 0.4302 x_2 - 4.4347 x_2 x_4 - 1.309 x_2 x_5 + 7.891 x_2 x_6 - 7.189 x_3 x_4 + 4.781 x_3 x_5 + 0.7226 x_4^2 - 0.58 x_4 x_5 + 4.992 x_4 x_6.$$

Parameters of model: CR = 0.0002728; BS = 3.5506.

## 5 Modelling of a process of diabetes treatment by GMDH algorithm with model determination

The problem consists of prediction of blood glucose level (BGL) for patient with the diabetes II. BGL is measured four times per day at 6, 12, 17 and 22 hours ( $x_{6k}$ ,  $x_{12k}$ ,  $x_{17k}$ ,  $x_{22k}$ ) and is written down as the doses of insulin, which patients inject four times a day ( $i_{6k}$ ,  $i_{12k}$ ,  $i_{17k}$ ,  $i_{22k}$ ). The output variable is value of BGL at 6 o'clock. In Table 3 we show the example of data sample for diabetes treatment.

**Table 3.** Data sample of monitoring diabetes treatment.

N <sup>o</sup>	Date	$x_{6k}$	$x_{12k}$	$x_{17k}$	$x_{22k}$	$i_{6k}$	$i_{12k}$	$i_{17k}$	$i_{22k}$
1	12.08.1993	9.70	17.80	22.70	13.20	17	10	13	12
2	13.08.1993	14.70	16.30	12.30	22.50	17	10	13	12
3	14.08.1993	18.50	14.00	14.70	17.20	18	10	13	12
4	15.08.1993	8.90	8.70	17.80	23.00	18	10	13	12
5	16.08.1993	15.40	12.40	9.70	7.90	20	10	14	12
6	17.08.1993	10.50	10.60	5.50	7.00	20	12	14	12
7	18.08.1993	13.50	24.60	4.40	13.60	20	12	14	12
8	19.08.1993	14.90	16.40	5.80	10.70	20	10	14	12
9	20.08.1993	22.50	11.40	6.10	13.70	20	12	14	12
10	21.08.1993	10.00	18.50	18.10	19.10	20	12	12	12

The following forecasting model was received by Combinatorial algorithm GMDH with determination by bias:

$$y = \hat{x}_{6k+1} = -0,32 + 0,16 x_{6k} - 0,05 x_{6k-1} + 0,2 x_{22k} + 0,5 i_{12k} + 0,5 i_{17k}.$$

Parameters of model: CR = 0,667; BS = 3,50.

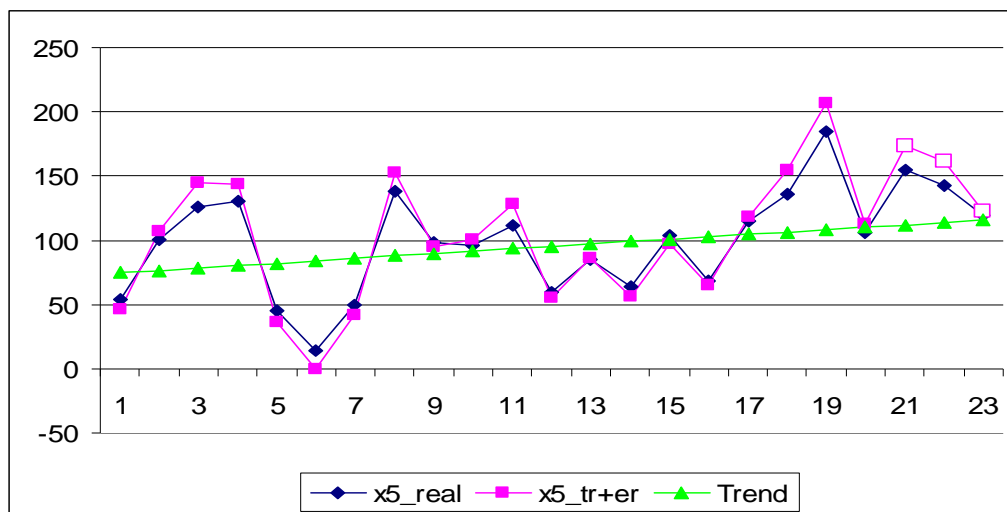
## 6 Method of modelling in two-layered monitoring systems with extraction of trend by the example of solar activity modelling

The problem of harmonic processes modelling, which has a characteristic cyclic or seasonal trend is considered. There exists some experience in modelling of natural cyclic processes with application of the GMDH algorithms in class of two-layered models, particularly in creation of solar activity models [5]. The 11-years cycle is observed in the Wolf's numbers, which characterise the solar activity. The input data contain 23 such cycles of solar activity.

In order to increase the accuracy of forecasting we recommend:

1. extraction of trend with optimal complexity;
2. estimation of remainder value, i.e. the difference between real values of output variables and values estimated by trend. In the case when the difference is small the forecast can be made by trend or to compute forecast as the sum of forecasts of difference and trend.

Forecast of solar activity by described method is shown in Figure 6.



**Figure 6.** Forecast of solar activity by GMDH algorithm.

Forecast is calculated for the fifth year of each cycle. In figure 6 the three curves are shown: real values of solar activity, the values calculated by trend and the values received by GMDH algorithm with preliminary extraction of trend. The last three points are calculated by model, received on the values of 20 previous cycles.

## 7 Conclusions

The described method of additional model determination by the bias criterion allows us to find the optimal model. This was shown in the numeral examples of construction of model of atomization coefficient dependencies from properties of different materials and forecasting model of blood glucose level for patients with diabetes. To increase the model accuracy we suggest at first to find trend of process, and then by GMDH algorithm to calculate the prediction of remainder value. This method is shown by the example of solar activity modelling.

## References

- [1] Ihara D. About unity of model choice by criterion of balance of variables // Automation.- №1.- 1976.- C. 86-87 (in Ukrainian).
- [2] Ivakhnenko A.G., Ivakhnenko G.A., Savchenko E.A. GMDH Algorithm for Optimal Model Choice by the External Error Criterion with the Extension of Definition by Model Bias and Its Applications to the Committees and Neural Networks // Pattern Recognition and Image Analysis, Vol. 12, № 4, 2002, pp.347-353.
- [3] Ivakhnenko A.G., Savchenko E.A., Ivakhnenko G.A., Nadiradze A.B., Rogov A.O. Detection of dependencies of ion interaction with a surface by the Combinatorial GMDH algorithm // Problems of control and informatics, № 2, 2003, c. 80-89 (in Russia).
- [4] Savchenko E.A. Express-prediction of blood glucose level taking into account analogue and temporal descriptions // Control Systems and Computers, №2, 2003, c. 107-112 (in Russia).
- [5] Stepashko V.S., Kostenko Yu.V. GMDH algorithm for two-layered modelling of multidimensional cyclic processes // Automation. -1087. -№4. - C. 51-60 (in Russia).