# Double-criterion Choice of the Optimal Model in GMDH Algorithms

Ievgeniia Savchenko

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

savchenko@irtc.org.ua

***Abstract.*** *Investigation and choice of reasonable sequence of the external GMDH criteria for the optimal model selection is the purpose of the work. The inductive modelling is directed to the optimal model construction by the adjusted external criteria. The sequence search of model by orthogonal criteria for the removal of these failings is proposed. Two variants of optimal model search are considered with a different sequence of external criteria of accuracy and bias for the choice a reasonable sequence of criteria. Numerical examples show search the reasonable sequence of the external GMDH criteria.*

## Keywords

Inductive modelling, GMDH, criteria bias and regularity, double-criterion choice.

## 1 Introduction

The aim of this work is to develop the method of double-criterion selection of the optimal model in GMDH with sequence criteria applying. This method use in the real modelling tasks when optimal model of any system can't be find by selection on Combinatorial GMDH algorithm.

Inductive modelling solves the problems of structural and parametrical identification. These problems are solved together. Two or more data samples are applied in the GMDH for the decision of coefficients estimations problems and the choice of the best model from all generated models [1]. Dividing data sample into two or more parts provides on external character of criteria. One of them is used for coefficients estimation and on the second part values of criterion are calculated. Thus values of criteria are calculated on the new (external) information, which was not used for model construction. It is impossible to use only one criterion because the model is more complex, the model is more exact.

Choosing model by external criteria we can expect from model to be the most exact or the least bias. For this purpose the regularity or the bias criterion can be used. The regularity and bias criteria are orthogonal by nature. Combined criteria are used to simultaneously to take into account this features at model choice. Calculation of the criteria requires assigning the weight coefficients of criteria according to what quality of the model it is needed to get. But such a procedure requires large computing time as values of criteria must be calculated for all generated models. The combinred criteria are preferably because it simultaneously takes into account several section criteria.

Secection criteria are constructed so that they are reflected several obvious demands to model. For example, "the model must to give the well forecast" – regularity criterion, "coefficients of model must marginally depends on part of data sample" – bias criterion, "the model must reflects connections from variables" – balance criterion.

Generally the external selection criterion can be definded as the measure of numerical comparison of different complexity models. This measure allows to choice a set of several models from all generated models.

Separately taken criteria often are named individual, convolution of several criteria is the combined criteria and criteria consecution is appuing several criterion in series one to another.

Below the numerical example will be shown in which describes the search of reasonable consecution of the extarnal criteria: regularity and bias. At first the some set of models is choosen by one criterion. Then the optimal model is selected by the minimum of second criterion.

## 2. External GMDH criteria

Input data sample can be divided into three parts:

$A$ – learning data sample;

$B$ – testing data sample;

$C$ – examination data sample

$W = A \cup B$ – working data sample.

Among external GMDH criteria the following are the most widespread [1].

Regularity criterion:

$$AR(B) = AR(B|A) = \sum_{i \in B} (y_i - \hat{y}_i(A))^2 = \|y_B - \hat{y}_B(A)\|^2 = \\ = (y_B - X_B \hat{a}_A)^T (y_B - X_B \hat{a}_A) = \|y_B - X_B \hat{a}_A\|^2, \quad (1)$$

when recording $AR(B|A)$ means «the error on $B$ model, coefficients of which are finding on $A$» and

$$\hat{a}_A = \left(X_A^T X_A\right)^{-1} X_A^T y, \quad \hat{y}_B(A) = X_B \hat{a}_A .$$

The most widespread form of bias criterion is a minimum of bias decisions:

$$CB = CB(W|A,B) = \|\hat{y}_W(A) - \hat{y}_W(B)\|^2 = \\ \|X_W \hat{a}_A - X_W \hat{a}_B\|^2 = (\hat{a}_A - \hat{a}_B)^\tau X_W^\tau X_W (\hat{a}_A - \hat{a}_B). \quad (2)$$

In work [2] the minimum of bias errors is proposed:

$$BS = |AR_{W/A} - AR_{W/B}| = \left| \left\| y - X_W \hat{\theta}_A \right\|^2 - \left\| y - X_W \hat{\theta}_B \right\|^2 \right|, \quad (3)$$

when $AR_{W/A}$ and $AR_{W/B}$ is a total errors on the incorporated sample $W = A \cup B$ of model of the same structures with the parameters which estimated on samples $A$ и $B$ accordingly:

$$AR_{W/A} = \left\| y - X_W \hat{\theta}_A \right\|^2; \; AR_{W/B} = \left\| y - X_W \hat{\theta}_B \right\|^2. \quad (4)$$

## 3. Double-criterion of the optimal model choice. Combined criteria

Either the combined criteria or a sequence of criteria usually are applied for the solution of a problem of double-criterion model choice. Often such problem is result of linear convolution of criteria particularly of the combined of criteria:

$$k^2 = \alpha c_1^2 + (1 - \alpha) c_2^2, \quad (5)$$

when $a_j$ - weight coefficients; $c_j$ - value of criteria.

However combined criterion calculation has some complexities:

1. Needs the normalization (weightings) of criteria;
2. Large computing time of value critetia calculation.

These complexities are eliminated by consecutive application of external criteria. At the beginning some number of models $F$ are found by minimum of first criterion. Then the optimal model is selected by the second criterion form the set of $F$ best models by the first criterion.

The regularity criterion usually is less effective in conditions of noise because it depends on the variant of data sample division. The examples show that noisestability essentially increase at stage-by-stage selection when some set of models are selected by first criterion and the best model are found by second criterion.

Various combinations of external criteria were developed in the works [1, 2]. Criteria of a regularity, bias and accuracy on examination data sample were used there. The sequence of two criteria is proposed in the work [3]. Here the 100 of most accurate models are found by minimum of bias criterion on combinatory GMDH algorithm. The coefficients of each of these equations are recalculated on all samples by the method of the least squares. The ten models are defined by minimum of regularity criterion. The optimal model is chosen by minimum of balance criterion.

## 3. A numerical example of double-criterion model choice

The aim of this example is selection of reasonable sequence of the external GMDH criteria for the optimal model choice. The input data sample is generated accidentally and contains 6 variables, 40 lines. Data sample are presented in the Table 1.

Data sample are divided in three parts. Fist and second parts are the learning and testing samples and they contain 30 lines together (working sample). This sample is used for coefficients estimation and value of criteria calculation. Third part is examination sample. This sample (10 lines) is used for testing obtained models.

**Tab.1**. Input data sample

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $y$ |
|---|---|---|---|---|---|---|
| -0.85 | -3.53 | 4.47 | 1.52 | -3.03 | 3.45 | -7.27 |
| 1.35 | -2.52 | 4.76 | 1.00 | -2.53 | 1.74 | -4.50 |
| -3.09 | -2.18 | 2.11 | -2.68 | 1.86 | -3.50 | 4.07 |
| 2.30 | 0.81 | -0.09 | 4.12 | 1.49 | -0.71 | -3.15 |
| -0.90 | 2.60 | -1.72 | 1.36 | -2.28 | -3.13 | -1.28 |
| -2.73 | 1.63 | -1.17 | -3.14 | -2.01 | -0.18 | 3.52 |
| -0.40 | -3.70 | 2.46 | 3.29 | 2.26 | -1.80 | -4.98 |
| -0.59 | -2.74 | 3.92 | -3.55 | 3.39 | 3.48 | 5.03 |
| 2.80 | 3.72 | 2.89 | 1.01 | -4.32 | -4.96 | -2.24 |
| -0.29 | -1.45 | -0.13 | -4.14 | 4.20 | -2.77 | 8.39 |
| … | … | … | … | … | … | … |
| 2.70 | 1.62 | -0.30 | 3.29 | 1.34 | 1.40 | -2.27 |
| -1.14 | 0.95 | 4.36 | 4.15 | -1.20 | 2.28 | -5.01 |
| -1.74 | -2.32 | -2.01 | 1.70 | -2.68 | 3.13 | -5.32 |

The output variable contains 30% noise in the data:

$$y = 0{,}3 + 0{,}7x_2 - 1{,}3x_4 + 0{,}8x_5 \tag{6}$$

It is required to consider two variants of model search for the of reasonable sequence of the external criteria:

1. At beginning regularity criterion is used and then the bias criterion is applied (*AR – BS*).
2. At first bias criterion is used and then the regularity criterion is applied (*BS - AR*).

### 3.1. The first variant of model search *(AR – BS)*

Lines of data sample range by a dispersion. Value of regularity criterion calculated by Combinatorial GMDH algorithm. Dependence of regularity criterion on model complexity is shown on Fig 1.

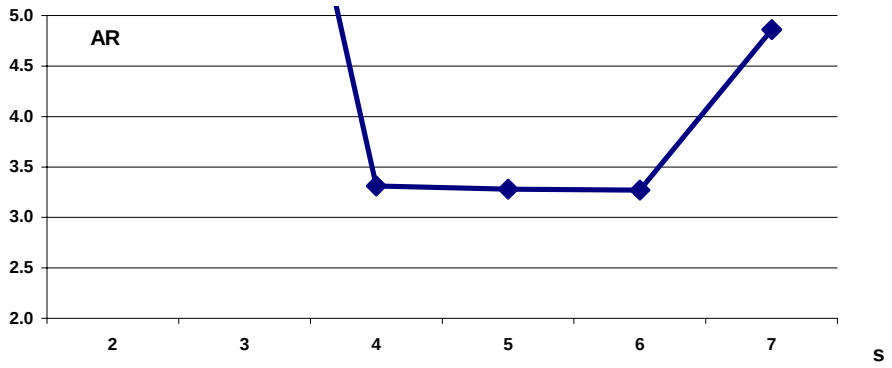


**Fig. 1**

Fig 1 shows that 2 models (S=4,5,6) get into the interval of uncertainty and have almost equal regularity criterion values. For these models bias criterion BS was calculated by (3). Fig.2 shows the optimal model choice by minimum of bias criterion.

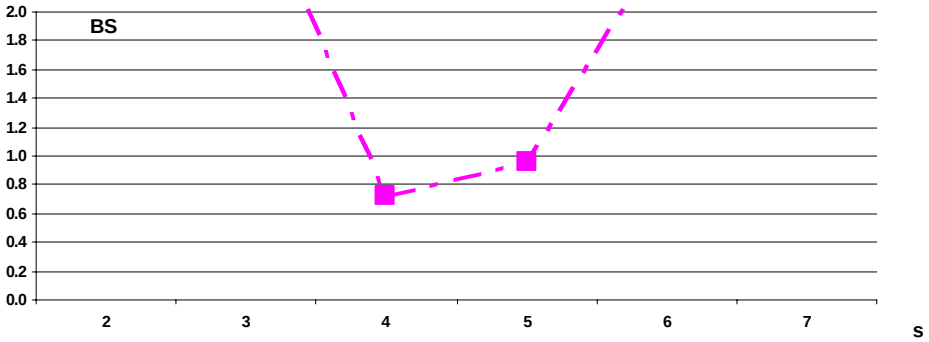


**Fig. 2**

The results of computation of both criteria regularity (CR) and bias (BS) are shown in Table 2.

**Tab.2.** The set of models chosen by GMDH

| *s* | *AR* | *BS* | *AR(C)* | *Models* | *Number of models* |
|---|---|---|---|---|---|
| 4 | 3,310 | 0,722 | 2,33 | $y = 0,2500 + 0,6501x_2 - 1,1590x_4 + 0,6317x_5$ | I |
| 5 | 3,280 | 0,764 | 2,36 | $y = 0,2339 + 0,6514x_2 - 0,0925x_3 - 1,1899x_4 + 0,6427x_5$ | II |
| 6 | 3,270 | 2,517 | 2,36 | $y = 0,2365 + 0,6518x_2 - 0,0922x_3 - 1,1886x_4 + 0,64327x_5 + 0,0048x_6$ | III |

Optimal model is complexity $s = 4$. Value of criteria are $AR = 3{,}31$; $BS = 0{,}722$. This model also have minimum of error on the examination data sample $AR(C) = 2{,}33$.

### 3.2. The second variant of model search *(BS – AR)*

Let's consider other sequence of criteria: at first bias criterion then regularity criterion. Dependence of bias and regularity criterion on model complexity is shown on the Fig.3.

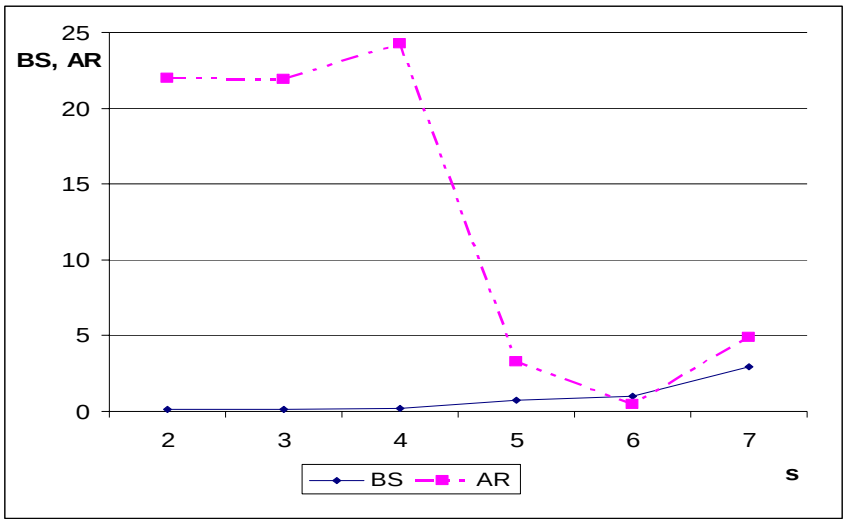


**Fig.3**

Fig. 3 has shown that the models having very great values of regularity criterion are selected by minimum of bias criterion, i.e. will be find very inaccurate models. Besides, true model have not appeared among selected by bias criterion models, i.e. such sequence of model can lead to the serious errors.

## Conclusion

Examples have shown that the most reasonable sequence of the external GMDH criteria for the optimal model selection is the sequence when at beginning some set of models is selected by criterion regularity, then the one model is selected by bias criterion. Using of sequence *BS – AR* faces with loss of optimum model as models are selected more "rough". It will lead to selection of simpler model in conditions of noise.

## References


[1] Ivakhnenko A.G., Stepashko V.S. Pomekhoustoychivost' modelirovaniya (Noise – immunity of modeling): Kiev.- Naukova Dumka. - 1985. – 216 p. (in Russian).

[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 - 2002.-V.12, N4. - P.347-353.

[3] Ivakhnenko A.G., Stepashko V.S., Khomovnenko M.G., Galyamin E.P. Selforganization of the dynamics models of agricultural cultures growth for crop rotation irrigated control // Automation. – 1977. - № 5. – P. 32 – 44 (in Russian).

[4] Akishin B.A., Yunusov N. Sequential applying of selforganization and statistical testing of hypothesis for adequate mathematical models choice // Automation.- 1978.- №3. – P. 64 – 67 (in Russian).

[5] Ivakhnenko A.G., Savchenko E.A., Ivakhnenko G.A. GMDH algorithm for optimal model selection by error criterion with additional determination by bias criterion and its application in committee and neural networks // Proceedings of the International Conference of Inductive Modelling 2002. – Lviv, 2002.- P. 47- 56 (in Ukranian).

[6] Ivakhnenko A.G., Savchenko E.A., Ivakhnenko G.A., Nadiradze A.B., Rogov A.O. Inductive method of Choosing a Model with the Least Error and Bias for Solving Interpolation Tasks of Artificial Intelligence // Pattern Recognition and Image Analysis.-2003.-V.13, N3.-P. 452-458.