

Recent Developments of Self-Organising Modeling in Prediction and Analysis of Stock Market

Ivakhnenko, A.G.

Glushkov Institute of Cybernetics, Ukraine, Kyiv 34, PO Box 298 -9,
e-mail: Gai@gmdh.kiev.ua <http://come.to/GMDH>

Müller, J.-A.

Fachbereich Informatik/Mathematik, Hochschule für Technik und Wirtschaft
D 01069 Dresden, F.-List-Platz 1, Germany, e-mail: Muellerj@informatik.htw-dresden.de

Review

***Abstract:** At present, GMDH algorithms give us the only way to get accurate identification and forecasts of different complex processes in the case of noised and short input sampling. In distinction to neural networks, the results are explicit mathematical models, obtained in a relative short time. For ill-defined objects with very big noises better results should be obtained by analogues complexing methods. Neural nets with active neurones should be applied to rise up accuracy of complex objects modelling algorithms.*

1. Introduction

Problems of complex objects modelling (functions approximation and extrapolation, identification, pattern recognition, forecasting of random processes and events) can be solved in general by deductive logical-mathematical or by inductive sorting-out methods. Deductive methods have advantages in the cases of rather simple modelling problems, when the theory of the object being modelled is known and therefore it is possible to develop a model from physically based principles employing the users knowledge of the process.

Decision making in such areas as process analysis in macroeconomy, financial forecasting, company solvency analysis and another requires tools, which are able to get accurate models on basis of processes forecasting. However, arise problems that connected with large amount of variables, very small number of observations and unknown dynamical between these variables. Such financial objects are complex ill-defined systems that can be characterised by:

- inadequate a priori information;
- great number of immeasurable variables;
- noisy and extremely short data samples;
- ill-defined objects with fuzzy characteristics.

Problems of complex objects modelling such as analysis and prediction of stock market and other, cannot be solved by deductive logical-mathematical methods with needed accuracy. In this case knowledge extraction from data, i.e. to derive a model from experimental measurements, has advantages in cases of rather complex objects, being only little a priori knowledge or no definite theory particularly for objects with fuzzy characteristics on hand. This is especially true for objects with fuzzy characteristics.

The task of knowledge extraction from data is to select mathematical description from data. But the required knowledge for designing of mathematical models or architecture of neural networks is not at the command of the users. In mathematical statistics it is need to have a priori information about the structure of the mathematical model. In neural networks the user estimates this structure by choosing the number of layers and the number and transfer functions of nodes of a neural network. This requires not only knowledge about the theory of neural networks, but also knowledge of the object nature and time. Besides this the knowledge from systems theory about the systems modelled is not applicable without transformation in neural network world. But the rules of translation are usually unknown.

GMDH type neural networks can overcome these problems - it can pick out knowledge about object directly from data sampling. The Group Method of Data Handling (GMDH) is the inductive sorting-out method, which has advantages in the cases of rather complex objects, having no definite theory, particularly for the objects with fuzzy characteristics. GMDH algorithms found the only optimal model using full sorting-out of model-candidates and operation of evaluation of them, by external criteria of accuracy or difference types [1,2].

2. Group Method of Data Handling (GMDH)

2.1. Brief description

The Group Method of Data Handling (GMDH) is self-organizing approach based on sorting-out of gradually complicated models and evaluation of them by external criterion on separate part of data sample. As input variables can be used any parameters, which can influence on the process. Computer is found structure of model and measures of selected parameters significance itself. That model is better that leads to the minimal value of external criterion. This inductive approach is different from commonly used deductive techniques or neural networks.

The GMDH was developed for complex systems modelling, prediction, identification and approximation of multivariate processes, decision support after "what-if" scenario, diagnostics, pattern recognition and clusterization of data sample. It was proved, that for inaccurate, noisy or small data can be found best optimal simplified model, accuracy of which is higher and structure is simpler than structure of usual full physical model.

There were defended more than 230 dissertations and published many papers and books devoted to GMDH theory and its applications. There are developed methods of mathematical induction for the solution of comparatively simple problems. GMDH can be considered as further propagation of inductive self-organising methods to the solution of more complex practical problems. It solves the problem of how to handle data samples of observations. The goal is to get mathematical model of the object (the problem of identification and pattern recognition) or to describe the processes, which will take place at object in the future (the problem of process forecasting).

GMDH solves, by sorting-out procedure, the multidimensional problem of model optimization:

$$\tilde{g} = \arg \min_{g \in G} CR(g), \quad CR(g) = f(P, S, \xi^2, T, V) \quad (1)$$

where: G - set of considered models; CR is an external criterion of model g quality from this set; P - number of variables set; S - model complexity; ξ^2 - noise dispersion; T - number of data sample transformation; V - type number of reference function. For definite reference function, each set of variables corresponds to definite model structure $P = S$. Problem transforms to much simpler one-dimensional

$$CR(g) = f(S),$$

when $\xi^2 = \text{const}$, $T = \text{const}$, and $V = \text{const}$.

Method is based on the sorting-out procedure, i.e. consequent testing of models, chosen from set of models-candidates in accordance with the given criterion. Most of GMDH algorithms use the polynomial reference functions. General connection between input and output variables can be expressed by Volterra functional series, discrete analogue of which is Kolmogorov-Gabor polynomial [1]:

$$y = a_0 + \sum_{i=1}^M a_i x_i + \sum_{i=1}^M \sum_{j=1}^M a_{ij} x_i x_j + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M a_{ijk} x_i x_j x_k,$$

where $X(x_1, x_2, \dots, x_M)$ - input variables vector;

$A(a_1, a_2, \dots, a_M)$ - vector of coefficients or weights.

Components of the input vector X can be independent variables, functional forms or finite difference terms. Other non-linear reference functions, such as difference, logistic, harmonic can also be used for model construction. The method allows to find simultaneously the structure of model and the dependence of modelled system output on the values of most significant inputs of the system.

The GMDH theory solve the problems of:

- long-term forecasting [3,18];
- short-term forecasting of processes and events [2];
- identification of physical regularities;
- approximation of multivariate processes;
- physical fields extrapolation [4];
- data samplings clusterization [5];
- pattern recognition in the case of continuous-valued or discrete variables;
- diagnostics and recognition by probabilistic sorting-out algorithms [6];
- vector process normative forecasting [7];
- modeless processes forecasting using analogues complexing [8];
- self-organization of twice-multilayered neuronet with active neurones [9,10].

In [12] were obtained the theoretical grounds of GMDH effectiveness as adequate method of robust forecasting models construction. Essence of it consists of automatically generation of models in given class by sequential selection of the best of them by criteria, which implicitly by sample dividing take into account the level of indeterminacy.

Since 1967 a big number of GMDH technique implementations for modelling of economic, ecological, environmental, medical, physical and military objects were done in several countries. Some outdated approaches are used in USA by Ward Systems Group, Inc. in "NeuroShell2", AbTech Corp. "ModelQuest", Barron Associates Co. "ASPN", and DeltaDesign Berlin Software "KnowledgeMiner" commercial software tools.

Self-organising modelling is based on statistical learning networks, which are networks of mathematical functions that capture complex, non-linear relationships in a compact and rapidly executable form. Such networks subdivide a problem into manageable pieces or nodes and then automatically apply advanced regression techniques to solve each of these simpler problems.

2.2. The "GMDH algorithms" and "algorithms of GMDH type"

It's necessary to make difference between the original "GMDH algorithms" and the "algorithms of GMDH type" [11]. The first ones - work using the minimum of an external criterion (Fig.1) and therefore realise objective choice of optimal model. This original GMDH technique is based on inductive approach: optimal models are founded by sorting-out of possible variants and evaluated by external criterion. It is calculated on separate part of data sample, which is not used for model creation. That model is better which leads to minimal value of criterion. To make objective choice, selection is done without thresholds or coefficients in criterion. We recommend to calculate criteria two times: first to find the best models at each layer of selection for structure identification and second time to find the optimal model. Selection procedure is stopped when minimal criterion value is reached.

Second is GMDH type algorithms - work on characteristic, expressed by words: "more complex is the model - more accurate it is". For it necessary to put definite threshold or to point out coefficients of weight for the members of the internal criterion formula, to find optimal model out in a some subjective way. But real problems usually are presented by short or noised data samples. Unfortunately, in almost all GMDH type software (ModelQuest, NeuroShell) and research works in USA and Japan this deductive approach is used, which is not effective for such kind of data.

The inductive approach does not eliminate the experts or take them away from the computer, but rather assigns them a special position. Experts indicate the selection criterion of a very general form and interpret the chosen models. They can influence the result of modelling by formulating new criteria. Computer becomes an objective referee for scientific controversies, if criteria ensemble is coordinated between experts, which take part in discussion.

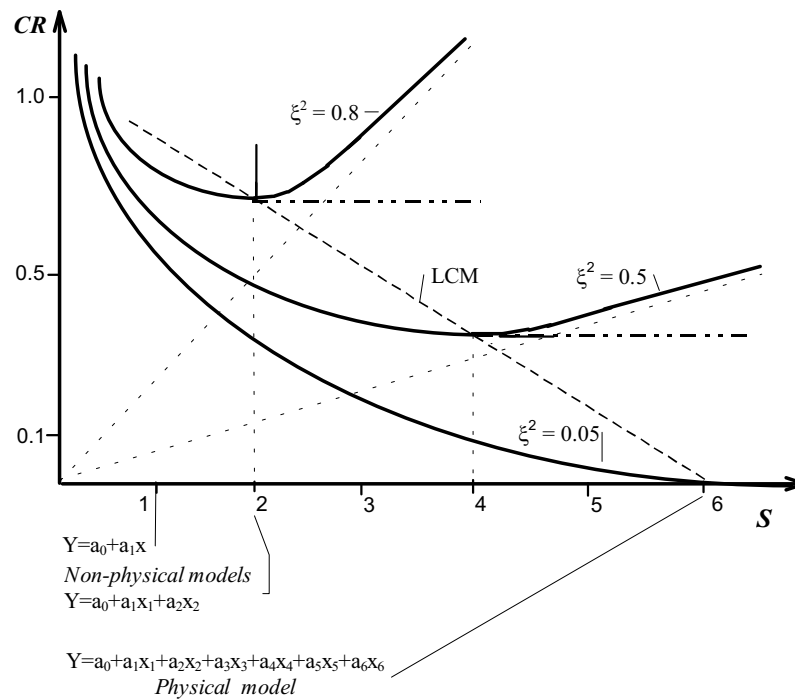


Fig.1. External accuracy criterion minima values plotted against complexity of model structure S for different noise variance ξ^2 .

LCM - locus of criterion minima line;

--- - model choice by criterion minimum.

The human element often involves errors and undesired decisions. Objective choice of optimal model by minimum of external criterion characteristic in actual GMDH algorithms often contradicts with the opinion of investigator. Objective algorithms give possibility to realise real artificial intelligence.

2.3. Special GMDH peculiarities

The main peculiarity of GMDH algorithms is that, when it uses continuous data with noise, it selects as optimal the simplified non-physical model. Only for accurate and discrete data the algorithms point out so-called physical model - the most simple optimal, from all unbiased models.

It is proved the convergence of multilayered GMDH algorithms [25] and it is proved that shortened non-physical model is better than full physical model (for noisy and continuous data for prediction and approximation solving, more simplified Shannon's non-physical models become more accurate [12]). It can be noted, that this conclusion has place in model selection on the basis of model entropy maximisation (Akaike approach), in average risk minimising (Vapnik approach) and in another modern approaches. The only way to get non-physical models is to use sorting-out GMDH algorithms. Regularity of optimal structure of forecasting models change in dependence on general indexes of data indeterminacy (noise level, data sample length, design of experiment, number of informational variables) was shown in [24,25,27].

The special peculiarities of GMDH are following:

- 1) External supplement: Following S.Beer work [13], only the external criteria, calculated on new independent information, can produce the minimum of sorting-out characteristic. Because of this data sampling is divided into parts for model construction and evaluation.
- 2) Freedom of choice: Following D.Gabor work [14], in multilayered GMDH algorithms are to be conveyed from one layer to the next layer not one but F best results of choice to provide "freedom of choice";
- 3) The rule of layers complication: Partial descriptions (forms of a mathematical description for iteration) should be simple, without quadratic members in them;
- 4) Additional model definition: In cases, when the choice of optimal physical model is difficult, because of noise level or oscillations of criterion minima characteristic, auxiliary discriminating criterion is used [15]. The choice of the main criterion and constrains of sorting-out procedure is the main heuristic of GMDH;
- 5) All algorithms have multilayered structure and parallel computing can be implemented for their realisation;
- 6) All questions that arise about type of algorithm, criterion, variables set etc. should be solved by minimal criterion value.

The main criteria used are: cross-validation $PRR(s)$, regularity $AR(s)$ and balance of variables criterion $BL(s)$. Estimation of their effectiveness (investigation of noise immunity, optimality and adequateness) and their comparison with another criteria was done in detail in [24,25,26,15]. The conditions, under which GMDH algorithm produces the minimum of characteristics are following:

- a) criterion of model choice is to be *external*, based on additional fresh information, which was not used for model construction;
- b) the data sample is not to be too long. Such data sample produce the same form of characteristic as the exact data sample without noises;
- c) when difference type balance criterion $BL(s)$ is used, small noise is necessary or the variables in the data sample should not be exactly measured [16].

Difference of the GMDH algorithms from another algorithms of structural identification, genetic and best regression selection algorithms consists of three main peculiarities:

- ♦ usage of external criteria, which are based on data sample dividing and are adequate to problem of forecasting models construction, by decreasing of requirements to volume of initial information;
- ♦ much more diversity of structure generators: usage like in regression algorithms of the ways of full or reduced sorting of structure variants and of original multilayered (iteration) procedures;
- ♦ better level of automatization: there are needed to enter initial data sample and type of external criterion only;

- ♦ automatic adaptation of optimal model complexity and external criteria to level of noises or statistical violations – effect of noiseimmunity cause robustness of the approach;
- ♦ implementation of principle of inconclusive decisions in process of gradual models complication.

2.4. Spectrum of GMDH algorithms

Solution of practical problems and GMDH theory design lead to development of broad spectrum of software algorithms. Each of them corresponds to some definite conditions of it application [17]. Algorithms mainly differ one from another by the models-candidates set generator arrangement for given basic function, by the way of models structure complexing and, at last, by the external criteria accepted. Algorithm choice depends on specifics of the problem, noise dispersion level, sufficiency of data sample, and if data sample is continuous-valued only.

Table 1. Spectrum of GMDH algorithms

<i>Variables</i>	GMDH algorithms	
	<i>Parametric</i>	<i>Non-parametric</i>
Continuous	<ul style="list-style-type: none"> - Combinatorial (COMBI) - Multilayered Iterational (MIA) - Objective System Analysis (OSA) - Harmonical - Two-level (ARIMAD) - Multiplicative-Additive 	<ul style="list-style-type: none"> - Objective Computer Clusterization (OCC); - "Pointing Finger" (PF) clusterization algorithm; - Analogues Complexing (AC)
Discrete and binary	<ul style="list-style-type: none"> - Harmonical Rediscretization 	<ul style="list-style-type: none"> - Algorithm on the base of Multilayered Theory of Statistical Decisions (MTSD)

Most often criteria of accuracy, differential or informative type are used. The work of GMDH algorithms has a straightforward analogy with the work of gardener during selection of a new hybrid plant [11].

The basic parametric GMDH algorithms listed in table 1 have been developed for continuous variables. Among the parametric algorithms [1,9] the most known are:

- the basic is *Combinatorial (COMBI)* algorithm. It is based on full or reduced sorting-out of gradually complicated models and evaluation of them by external criterion on separate part of data sample;
- *Multilayered Iteration (MIA)* algorithm use at each layer of sorting procedure the same partial description (iteration rule). It should be used when it is needed to handle a big number of variables;
- *Objective System Analysis (OSA)* algorithm. The key feature of it is that it examines not single equations, but systems of algebraic or difference equations, obtained by implicit templates

(without goal function). An advantage of the algorithm is that the information embedded in the data sample is utilised better and we get relationships between variables;

- *Two-level (ARIMAD)* algorithm for modelling of long-term cyclic processes (such as stock or weather). There are used systems of polynomial or difference equations for identification of models on two time scales and then choice of the best pair of models by external criterion value. For this can be used any parametric algorithm from described above [23].

Also less known parametric algorithms, which apply an exhaustive search to difference, harmonic or harmonic-exponential functions, and the Multiplicative-Additive algorithm, in which tested polynomial models are obtained by taking the logarithm of the product of input variables [18,19]. The parametric GMDH algorithms have proved to be highly efficient in cases where one is to model objects with non-fuzzy characteristics, such as engineering objects. In cases, where modelling involves objects with fuzzy characteristics, it is more efficient to use the non-parametric GMDH algorithms, in which polynomial models are replaced by a data sample divided into intervals or clusters. Such type algorithms completely solve the problem of coefficients estimates bias elimination.

Non-parametric algorithms are exemplified by:

- *Objective Computer Clusterization (OCC)* algorithm that operates with pairs of closely spaced sample points [5]. It finds physical clusterization that would as possible be the same on two subsamples;
- *"Pointing Finger" (PF)* algorithm for the search of physical clusterization. It is implemented by construction of two hierarchical clustering trees and estimation by the balance criterion [20];
- *Analogues Complexing (AC)* algorithm, which use the set of analogues instead of models and clusterizations [8]. It is recommended for the most fuzzy objects;
- algorithm, based on the *Multilayered Theory of Statistical Decisions* [6]. It is recommended for recognition of binary objects and for the variability of input data control to avoid the possible experts' errors in it.

Recent developments of the GMDH have led to neuronets with active neurons, which realise twice-multilayered structure: neurons are multilayered and they are connected into multilayered structure. This gives possibility to optimise the set of input variables at each layer, while the accuracy increases. The accuracy of forecasting, approximation or pattern recognition can be increased beyond the limits, which are reached by neuronet with single neurons, or by usual statistical methods [9,10,34]. In this approach, which corresponds to the actions of human nervous system, the connections between several neurons are not fixed but change depending on the neurons themselves. Such active neurons are able during the learning self-organising process to estimate which inputs are necessary to minimise the given objective function of neuron. This is possible on the condition that every neuron in its turn is multilayered unit, such as modelling GMDH algorithm. Neuronet with active neurons, which are described below, is considered as a

tool to increase AI problems accuracy and lead-time with the help of regression area extension for inaccurate, noisy data or small data samples.

The GMDH algorithms recently are applied in optimization to solve the problems of normative forecasting (after "what-if-then" scenario) and optimal control of multivariable ill-defined objects. Many ill-defined objects in macroeconomy, ecology, manufacturing etc. can be described accurately enough by static algebraic or by difference equations, which can be transformed into problems of linear programming by nomination of non-linear members by additional variables. GMDH algorithms are applied to evaluate deflections of output variables from their reference optimal values [7,21]. Examples of use of *Simplified Linear Programming (SLP)* algorithm should be used for expert computer advisor construction, normative forecasting and control optimization of averaged variables. An important example [10] gives the prediction of effects of experiments. The algorithm solves two problems: calculation of effects of a given experiment and calculation of parameters which are necessary to reach optimal results. It means, that the realisation of experiments can often be replaced by computer experiments.

As already noted, considered GMDH algorithms have been developed for continuous variables. In practice, however the sample will often include variables discretized into a small number of levels or even binary values. To extend these GMDH algorithms to discretized or binary variables, the Harmonic Rediscretization algorithm has been developed [22].

The existence of a broad gamut of GMDH algorithms is traceable to the fact, that it is impossible to define the characteristics of the rest or controlled objects exactly in advance. Therefore, it can be good practice to try several GMDH algorithms one after another and to decide which one suits a given type of objects best. All the questions, which arise during modelling process, are to be solved by the comparison of the criterion values: that variant is better, which leads to more deeper minimum of basic external criteria. In this way, the type of algorithm is chosen objectively, according to the value of the discriminating criterion.

Information about dispersion of noise level is very useful to decrease computer calculation time. For small dispersion level we can use the learning networks of GMDH type, based on the ordinary regression analysis using internal criteria. For considerable noise level the GMDH algorithms with external criteria are recommended. And for high level of noise dispersion non-parametric algorithms of clusterization or analogues complexing should be applied [8].

2.4.1. The Combinatorial GMDH algorithm (COMBI)

The flowchart of the algorithm is shown in Fig. 2. The input data sample is a matrix containing N levels (points) of observations over a set of M variables. The sample is divided into two parts. Approximately two-thirds of points make up the learning subsample N_A , and the remaining one-third of points (e.g. every third point) with same variance form the check subsample N_B . Before dividing, points are ranged by variation value. The learning sample is used to derive estimates for the coefficients of the polynomial, and the check subsample is used to choose the structure of the

optimal model, that is, one for which the external regularity criterion $AR(s)$ takes on a minimal value:

$$AR(s) = \frac{1}{N_B} \sum_{i=1}^{N_B} (y_i - y_i(B))^2 \rightarrow \min \quad (2)$$

or better to use the cross-validation criterion $PRR(s)$ (it takes into account all information in data sample and it can be computed without recalculating of system for each checking point):

$$PRR(s) = \frac{1}{N} \sum_1^N [y_i - y_i(B)]^2 \rightarrow \min, \quad N_A = N - 1; \quad N_B = 1.$$

To test a model for compliance with the differential balance criterion, the input data sample is divided into two equal parts. The criterion requires to choose a model that would, as far as possible, be the same on both subsamples. The balance criterion will yield the only optimal physical model solely if the input data are noisy.

To obtain a smooth exhaustive-search curve (Fig. 1), which would permit one to formulate the exhaustive-search termination rule, the exhaustive search is performed on models classed into groups of an equal complexity. For example, the first layer can use the information contained in every column of the sample; that is full search is applied to all possible models of the form:

$$y = a_0 + a_1 x_i, \quad i = 1, 2, \dots, M. \quad (3)$$

Non-linear members can be taken as new input variables in data sampling. The output variable is specified there in advance by the experimenter. At next layer are sorted all models of the form:

$$y = a_0 + a_1 x_i + a_2 x_j, \quad j = 1, 2, \dots, M \quad (4)$$

The models are evaluated for compliance with the criterion, and so on until the criterion value decrease. For limitation of calculation time recently it was proposed during full sorting of models to range variables according to criterion value after some time of calculation or after some layers of iteration. Then full sorting procedure continues for selected set of best variables till the minimal value of criterion will be found. This gives possibility to set much more input variables at input and to save effective variables between layers to found optimal model.

A salient feature of the GMDH algorithms is that, when they are presented continuous or noisy input data, they will yield as optimal some *simplified non-physical model*. If is only in the case of discrete or exact data that the exhaustive search for compliance with the precision criterion will yield what is called a *physical model*, the simplest of all unbiased models. With noisy and continuous input data, simplified (Shannon) models prove more precise [12,25] in approximation and for forecasting tasks.

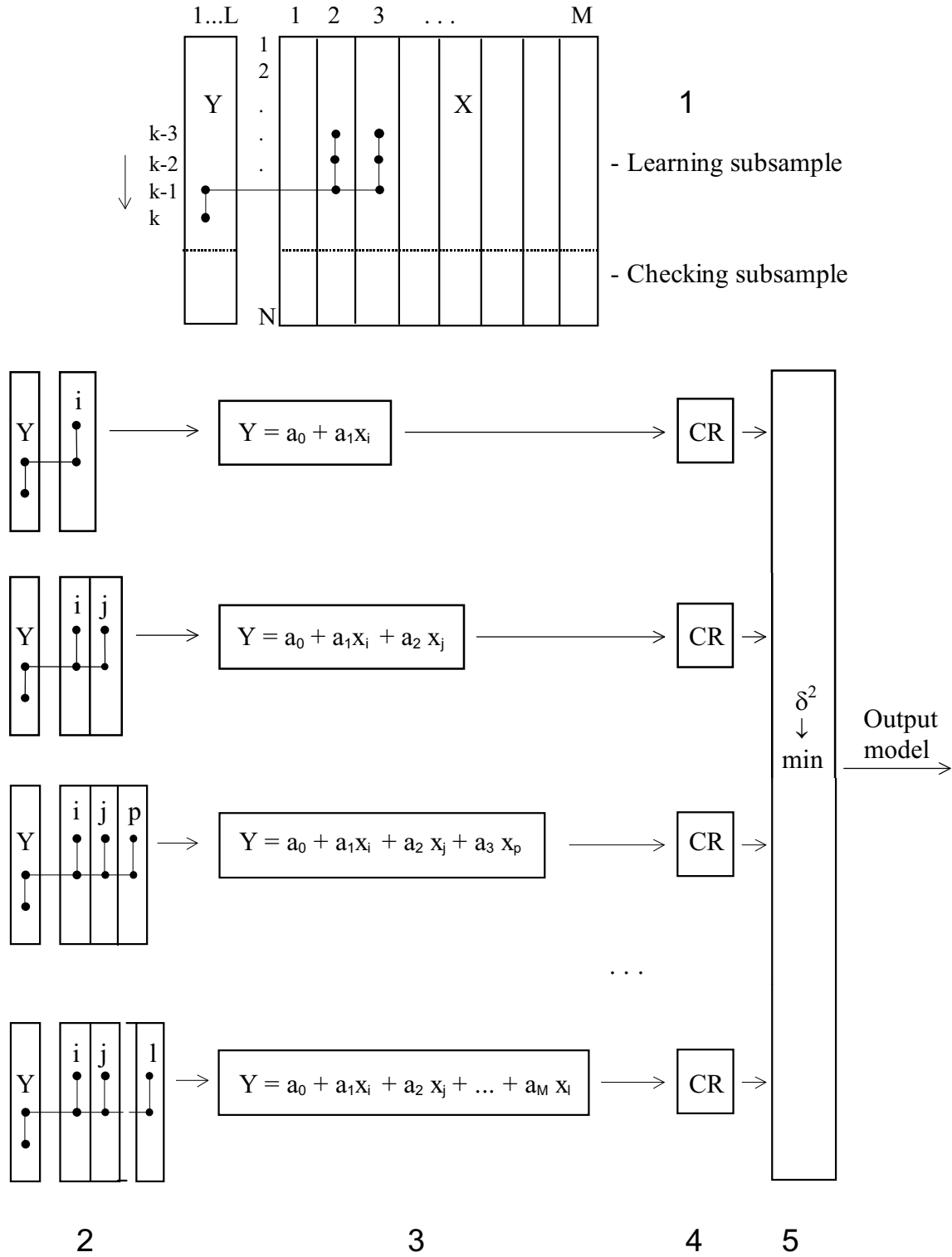
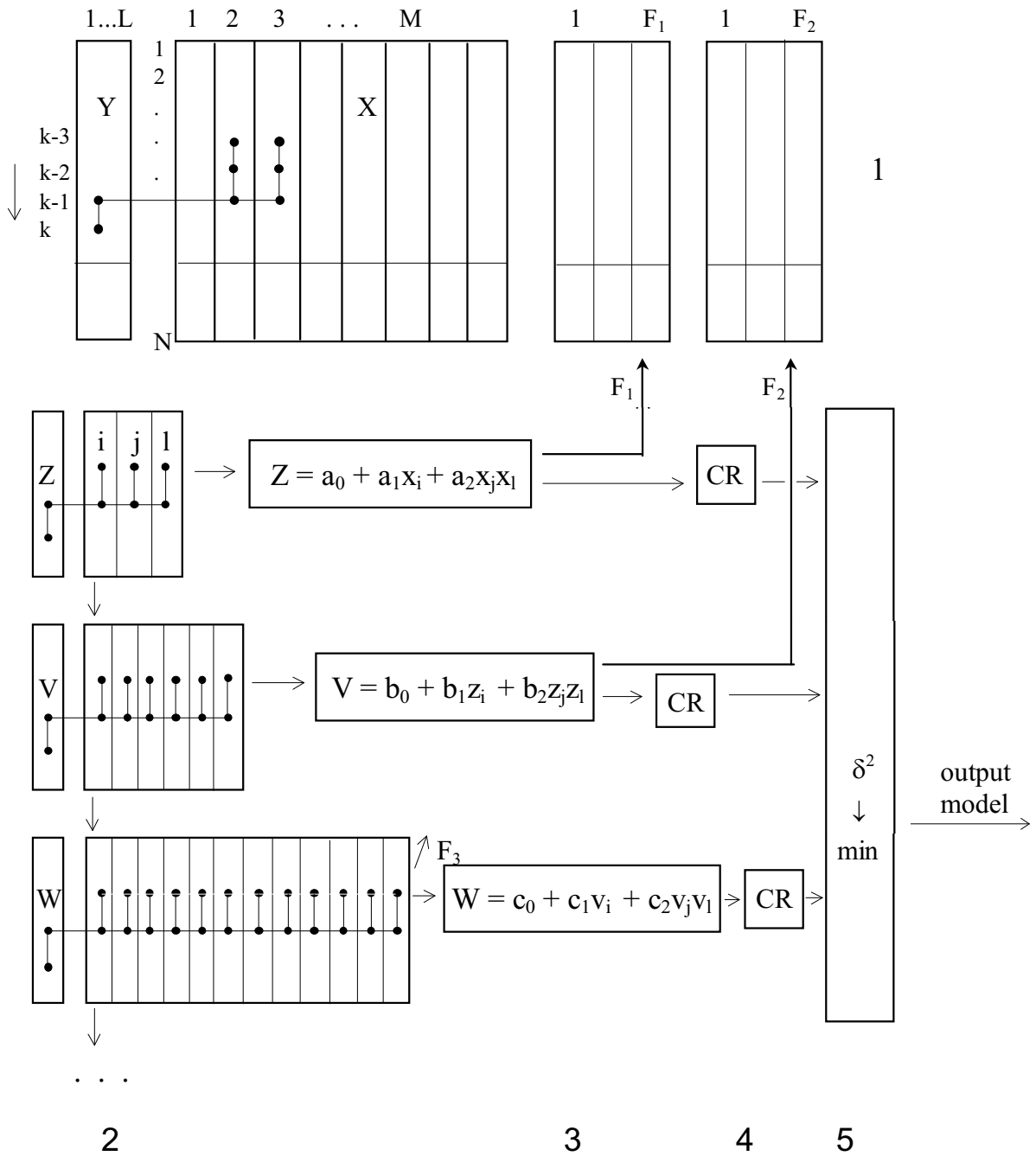


Fig. 2. Combinatorial GMDH algorithm.

- 1 - data sampling;
- 2 - layers of partial descriptions complexing;
- 3 - form of partial descriptions;
- 4 - choice of optimal models;
- 5 - additional model definition by discriminating criterion.



$$\text{Output model: } Y_{k+1} = d_0 + d_1 x_{1k} + d_2 x_{2k} + \dots + d_m x_{Mk} x_{M-1k}$$

Fig.3 Multilayered Iterational algorithm:

- 1 - data sampling;
- 2 - layers of partial descriptions complexing;
- 3 - form of partial descriptions;
- 4 - choice of optimal models;
- 5 - additional model definition by discriminating criterion;
- F1 and F2 - number of variables for data sampling extension.

Calculations are faster when following techniques are used [24,25]:

- a) in all formulae informational array $W^T W$ is used instead of data sampling array $W=(XY)$;
- b) model's parameters are estimated by recursion method of "framing" which allows to use arrays calculated on previous steps;
- c) faster generation of variables ensemble is done using Garsaid binary counter, where current ensemble is differ from previous in one digit only.

2.4.2. The Multilayered Iterative GMDH algorithm (MIA)

As with the Combinatorial algorithm, the output variable must be specified in advance by the person in charge of modelling, which corresponds to the use of so-called explicit templates (Fig.4). In each layer, new output variables values, calculated by the F best models in each point are used to successively extend the data sample (Fig. 3).

In Multilayered Iterative algorithm the iteration rule remains unchanged from one layer to next. As is shown in Fig. 3, the first layer tests models that can be derived from the information contained in any two columns of the sample. The second layer uses information from four columns; the third, from any eight columns, etc. The exhaustive-search termination rule is the same as for the Combinatorial algorithm: in each layer the optimal models are selected by the minimum of the criterion [16,25].

2.4.3. The Objective System Analysis algorithm (OSA)

In discrete mathematics, the term template refers to a graph indicating which of the delayed arguments are used in setting up conditional and normal Gauss equations. A gradual increase in the structural complexity of candidate models corresponds to an increase in the complexity of templates whose explicit (a) and implicit (b) forms are shown in Fig. 4.

When one uses implicit templates, one has, beginning from the second layer of the exhaustive search, to solve a system of equations and to evaluate the model, using a system criterion.

The system criterion is a convolution of the criteria calculated by the equations that make up the system

$$CR_{\text{sys}t} = \frac{1}{S} \sqrt{CR_1^2 + CR_2^2 + \dots + CR_s^2} \rightarrow \min, \quad (5)$$

where s is the number of equations in the system. The flowchart of the OSA algorithm is shown in Fig. 5. The key feature of the algorithm is that it uses implicit templates, and an optimal model is therefore found as a system of algebraic or difference equations. An advantage of this algorithm is that the number of regressors is increased and in consequence, the information embedded in the data sample) is utilised better. A disadvantage is that it calls for a large amount of calculations in order to solve the system of equations and a greater number of candidate models have to be searched. The amount of search can be reduced, using a constraint in the form of an auxiliary precision criterion.

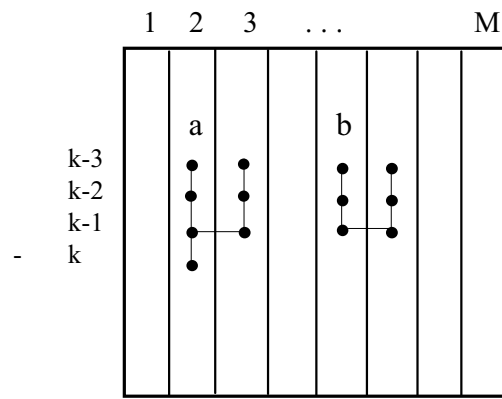


Fig.4. Derivation of conditional equations on a data sample

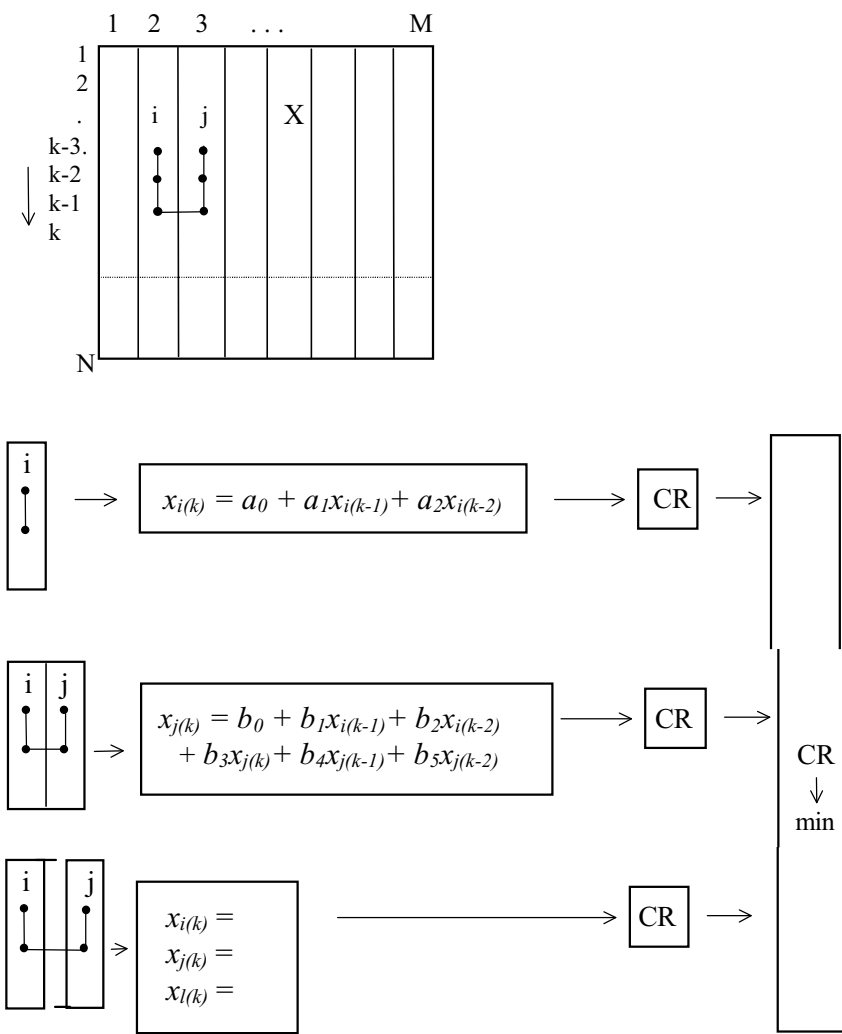


Fig. 5. Objective System Analysis (OSA) algorithm

In setting up the system of equations, one then discards the poorly forecasting equation (using equation only) for which the *variation accuracy criterion* for the forecast is less than unity (narrowing operation):

$$\delta_i^2 = \frac{\sum_1^N (y_i - \hat{y}_i)^2}{\sum_1^N (y_i - \bar{y})^2} \rightarrow \min, \quad (6)$$

where: y_i - is the variable values in the table;

\hat{y}_i - is the value calculated according to the model and

\bar{y} is the mean value.

This criterion is recommended in the literature in order to evaluate the success of an approximation or of a forecast [15]. With $\delta^2 < 0.5$, the result of modelling is taken to be good; with $0.5 < \delta^2 < 0.8$ it is taken to be satisfactory; with $\delta^2 > 1.0$, modelling is considered to have failed, and the model yields misinformation.)

2.5. Extended definition of the only optimal model by the theory of discriminating criteria

It has been demonstrated theoretically and experimentally that the exhaustive-search curves shown in Fig. 1 are gradual and unimodal for the expected value of the criterion [25]. The number of candidate models tested in each exhaustive-search layer cannot be infinitely large. In other words, in constructing exhaustive search curves, the expected value of the criterion is in effect replaced by its mean (or least) value. Because of this, the curves take on a slightly wavy shape, and a small error may creep into the optimal model structure choice.

The theory of discriminating criteria has been developed by Fedorov and Yurachkovsky [24] with special reference to experimental design. It has however proved its relevance to the self-organisation of models and active-neuron neural networks. The theory proceeds from the following premises: (1) there exists a "true" model represented in the data sample; (2) the assumed few object descriptions fit the model to a different degree; (3) the model that comes closest to the true model can be selected from its compliance with an *auxiliary discriminating criterion*.

With such an approach, every GMDH algorithm consecutively uses two criteria. At first, an exhaustive search is applied to all candidate models for compliance with the main criterion, and a small number of models whose structure is close to optimal are selected. Then only one optimal model is selected that complies with a special discriminating criterion. The theory of optimal discriminating criteria is still in the developmental stage, but successful discriminating criteria are already known.

In cases involving the selection of a structure for optimal polynomial models, the approximation or forecast variation criterion serves well. In the selection of optimal clusterization, good results are obtained with the symmetry criterion for the clusters distance matrix calculated relative to the secondary diagonal [21], etc.

3. Data analysis: neural networks versus self-organising modelling

The table 2 gives a comparison of both methodologies: neural networks and self-organising modelling in connection with their application to data analysis.

Table 2. Neural networks versus self-organising modelling.

	Neural networks	Statistical learning networks
Data analysis	universal approximator	structure identifier
Analytical model	indirect by approximation	direct
Architecture	unbounded network structure; experimental selection of adequate architecture demands time and experience	bounded network structure [1]; adaptively synthesised structure
A-priori-Information	without transformation in the world of neural networks not usable	can be used directly to select the reference functions and criteria
Self-organisation	deductive, given number of layers and number of nodes (subjective choice)	inductive, number of layers and of nodes estimated by minimum of external criterion (objective choice)
Parameter estimation	in a recursive way; demands long samples	estimation on training set by means of maximum likelihood techniques, selection on testing set (extremely short)
Feature	result depends from initial solution, time-consuming technique, necessary knowledge about the theory of neural networks	existence of a model of optimal complexity, not time-consuming technique, necessary knowledge about the task (criteria) and class of system (linear, non-linear)

Results obtained by statistical learning networks and especially GMDH type algorithms are comparable with results obtained by neural networks [30]. In distinction to neural networks, the results of GMDH algorithms are explicit mathematical models obtained in a relative short time on the base of extremely short samples. The well-known problems of an optimal (subjective) choice of the neural network architecture are solved in the GMDH algorithms by means of an adaptive synthesis (objective choice) of the architecture. There are to estimate networks of the right size with a structure evolved during the estimation process to provide a parsimonious model for the particular desired function. Such algorithms combining the best features of neural nets and statistical techniques in a powerful way discover the entire model structure - in the form of a network of polynomial functions, difference equations and other. Models are selected automatically based on their ability to solve the task (approximation, identification, prediction, and classification).

4. Nets of active neurons

4.1. Self-organisation of twice-multilayered neural network

A neural network is designed to handle a particular task. This may involve relation identification (approximation), pattern and situation recognition, or a forecast of random processes and repetitive events from information contained in a sample of observations over a test or control object.

The present stage of computer technology allows a new approach in neural networks, which increases the accuracy of classical modelling algorithms. Such complex system can solve complex problems. We can use the GMDH algorithms as the complex neurons, where the self-organisation processes are well studied.

Only by this inductive self-organising method for small, inaccurate or noisy data samples optimal non-physical model, accuracy of which is higher and structure is simpler than structure of usual full physical model can be found. GMDH algorithms are the examples of complex active neurons, because they choose the effective inputs and corresponding coefficients of them by themselves, in process of self-organisation. The problem of neuronet links structure self-organisation is solved in a rather simple way.

Each neuron is an elementary system that handles the same task. The objective sought in combining many neurons into a network is to enhance the accuracy in achieving the assigned task through a better use of input data. As already noted, the function of active neurons can be performed by various recognition systems, notably by Rosenblatt's two-layer perceptrons - such neural network achieves the task of pattern recognition. In the self-organisation of a neural network, the exhaustive search is first applied to determine the number of neuron layers and the sets of input and output variables for each neuron. The minimum of the discriminating criterion suggests the variables for which it is advantageous to build a neural network and how many neuron layers should be used. Thus, the theory of neural network self-organisation is similar in many respects to that of each active neuron.

Active neurons are able, during the self-organizing process, to estimate which inputs are necessary to minimise the given objective function of the neuron. In the neuronet with such neurons, we shall have twofold multilayered structure: neurons themselves are multilayered, and they will be united into a multilayered network. They can provide generation of new features of special type (the outputs of neurons from previous layer) and the choice of effective set of factors at each layer of neurons. The output variables of previous layers are very effective secondary inputs for the neurons of next layer. First layer of active neurons acts similar to Kalman filter: output set of variables repeated the input set but with filtration of noises. Number of active neurons in each layer is equal to number of variables given in initial data sampling.

Neuronet structure is given in Fig. 6. Solely including the output calculated variables from each previous layer of neurons effects sample extension. The samples show the form of the discrete template used to teach the first neurons of a layer by the Combinatorial GMDH algorithm. In particular, when four input variables are used and two time delays are allowed for ($\tau=2$), the first template corresponds (to the following complete difference equation:

$$x_{1(k)} = a_0 + a_1x_{1(k-1)} + a_2x_{1(k-2)} + \dots + a_8x_{4(k-1)} + a_9x_{4(k-2)}$$

The algorithm will suggest which of the proposed arguments should be taken into consideration and will help to estimate the connectivity coefficients.

To begin with, we construct the first layer of neurons in the network. Then we will be able to determine how accurate the forecast will be for all variables. For this purpose, we use a discrete template that allows a delay of one or two days for all variables. Then we add a second, a third, etc. layer to the neural network, as shown in Fig. 6, and go on doing so as long as this improves the forecast or decrease external criterion value.

For each neuron, we have applied the extended definition procedure to one model (out of the five closest to the optimal one). For the optimal models, we have calculated the forecast variation criterion. It may be inferred, that there is no need to construct a neural network in order to form a forecast for those variables, for which variation criterion value takes on the least value in the first layer. It is advisable to use a neural network to form a forecast for the variables, for which the variation criterion takes on the least value in the last layers of neurons.

The equations for the neurons of the network define the connections that must be implemented in the neural network; in this way they help achieve the task of structural self-organisation of the neural network. For brevity, the data sample in the above example is extended in only one way: tile output variables of the first layer are passed on as additional variables to the second, third, etc. layer of neurons. It is possible to compare different schemes of data sample extension by external criterion value.

The task for self-organisation of such networks of active neurons by selection is to estimate the number of layers of active neurons and the set of possible potential inputs and outputs of every neuron. The sorting characteristic - "number of neuronet layers - variables, given in data sample" - defines the optimum number of layers for each variable separately. Neuronets with active neurons should be applied to raise the accuracy of short-term and long-term forecasts.

Not only GMDH algorithms, but also many modelling or pattern recognition algorithms can be used as active neurons. Its accuracy can be increased in two ways:

- each output of algorithm (active neuron) generate new variable which can be used as a new factor in next layers of neuronet;
- the set of factors can be optimised at each layer. The factors (including new generated) can be ranked after their efficiency and several of the most efficient factors can be used as inputs for next layers of neurons. In usual once-multilayered ANN the set of input variables can be chosen once only.

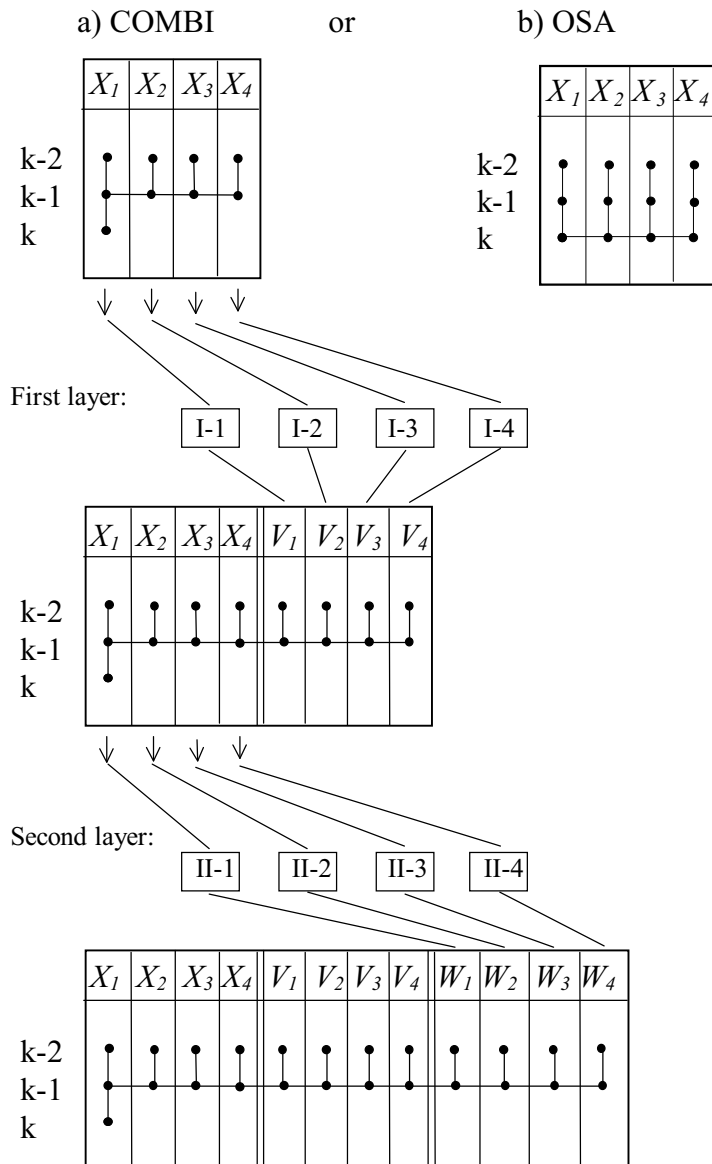


Fig 6. Schematic arrangement of the first two rows of a neural network.

4.2. The search termination rule

In self-organisation, the layers of neurons are extended as long as this improves the accuracy of the solution yielded by the neural network. This will be demonstrated later with reference to a relevant example.

4.3. Group allowance for arguments

We will call as the *exhaustive-search characteristic* of a neural network the graph that relates the main precision criterion for a specified variable to the layer number. This characteristic is similar to that of the GMDH algorithms. To obtain a smooth and unimodal curve, the exhaustive-search characteristic is calculated for many tools in the sample, and the results are averaged. Theoretically, the exhaustive-search characteristic has been investigated for the expected value of

the criterion [24]. In practice, the exhaustive-search curve has to be constructed not for the expected value and even not for the mean value of the criterion. Rather, it is constructed for the best results of the exhaustive-search applied to a group for which the criterion takes on the least value. This exhaustive-search termination rule holds only when many approximation or forecast results are average.

4.4. The selection of a discrete template

What type of template to use depends on the task at hand (Fig. 4). In an approximation task, the template does not contain delayed arguments; in a forecast task, two or three delays have to be allowed for. In the former case, one obtains single-moment equations; in the latter, difference equations.

4.5. Extended definition of one optimal model for each neuron in a network

Self-organisation of each neuron taken separately uses the differential balance criterion or the regularity precision criterion. As already noted, the exhaustive-search curve approaches its minimum in a gradual manner, and the criteria of models close to the optimal one differ only slightly from one another in value. This explains why one has to use an extended definition algorithm. This algorithm, instead of one, selects several of the best models. From them chosen only one that complies with another variation discriminating criterion.

4.6. Readout of modelling results

Each layer in a neural network contains neurons, whose outputs correspond each to a particular specified variable: the output of the first neuron to the first variable, the output of the second neuron to the second variable, etc. Each column consists of neurons whose outputs correspond to one of the variables. From each column in turn, one neuron with a minimal variation criterion is selected. More specifically, one neuron having the best result is selected from the first column of neurons for which the output is the first variable; similarly, one neuron is selected from the second column of neurons for which the output is the second variable, etc. This selection procedure uniquely defines the number of layers for each variable and, thus, the structure of the neural network.

4.7. The exhaustive search of methods for data -sample extension and narrowing

The principal method of data-sample extension is by including the output variables from the previous layer that have complied with the criterion best of all. It will also be a good plan to test against the criterion the advisability of sample extension by simple non-linear transformations of input variables. In the example that follows, three variables are involved. They are x_1 , x_2 , and x_3 .

(a) The extension using the covariance of the variables

$$v_1 = x_1x_2, \quad v_2 = x_1x_3, \quad v_3 = x_2x_3.$$

(b) The extension using the reciprocals of the variables

$$v_1 = \frac{1}{x_1}, \quad v_2 = \frac{1}{x_2}, \quad v_3 = \frac{1}{x_3}.$$

The reciprocals should above all be proposed for the variables that take a minus sign in the equation; that is, they reduce the value of the output.

4.8. Sample extension by consecutive elimination of the most efficient variables

The diversity of the variables that come in for the exhaustive search (performed by each neuron) can further be increased by eliminating the most efficient variables, thus producing partial subsets. This can be best illustrated by an example.

Let the input of a neural network accept a data sample containing just $M=25$ variables. Suppose further that we have used the OSA algorithm and found in the first neuron an optimal system of forecasting difference equations in the variables $x_2, x_{12}, x_{13}, x_{18}, x_{22}$. These variables are least "fuzzy" and lend themselves to forecasting by this system of equation. We eliminate from the sample the variables thus found and apply the OSA algorithm to a second neuron. This yields a second optimal system of equations in the variables x_3, x_9, x_{14}, x_{32} . As a result, the minimum of the criterion increases (because the second set contains other than the best variables) and shifts to the left (Fig. 1). Now we eliminate from the sample the nine variables thus found, and apply the OSA algorithm to a third neuron. This yields an optimal system of equations in only three variables x_5, x_6, x_{11} . The minimum of the criterion goes up still more and again shifts to the left etc.

This shift in the minimum of the system criterion bears out the adequacy law, which states that for more fuzzy systems the optimal description (model) must be likewise more fuzzy and simple; that is, it must have a smaller number of equations [24]. Computer experiments confirm the above form of exhaustive-search curve. In the above example, the number of variables used for decision-making is increased from 5 (in the first neuron) to $5 + 4 + 3 + 2 + 1 = 15$ (in five neurons). Ten features are discarded as inefficient. So, we shall have $5 \times 15 = 75$ neurons in each layer.

4.9. Simultaneous and successive algorithms for neural networks

In a computer program, neurons can be implemented simultaneously or successively, using memory devices.

4.10. Neuronets self-organisation and algorithms for optimization of control systems

The principal roadblock in the use of linear and non-linear programming algorithms for complex system optimization is that it is often impossible to specify either the goal function or the applicable constraints with sufficient accuracy. Meanwhile, even minute inaccuracies in their specification may have a strong impact on the outcome of optimization. Active-neuron networks can be readily combined with linear and non-linear programming algorithms.

One of the output functions is taken as the objective function, the equations of the other output variables can serve as equality-type constraints. This removes the subjective factor from the specification of the goal function and constraints. The human operator defines criteria for their choice, and not the objective function and constraints themselves [21].

6. Examples of applications.

Besides the applications of commercial GMDH software there were a lot of implementations made in very different fields. Many of them are described in the ukrainian journal "Avtomatica" (translated in "Soviet Automatic Control", "Soviet Journal of Automation and Information Sciences" and then in "Journal of Automation and Information Sciences" in full size). The basic GMDH technique applications include the studies on: economical systems (analysis and forecasting of macroeconomy parameters, decision support and optimization), ecological systems analysis and prediction (forecasting of oil fields and river flow, harvest analysis and ionosphere state definition), environment systems analysis, medical diagnostics, demographic forecasting, weather modelling, econometric modelling and marketing, manufacturing, planning of physical experiments, materials estimation, multisensor signal processing, microprocessor-based hardware, eddy currents, x-ray, acoustic and seismic analysis and widely in military systems (radar, infrared, ultrasonic and acoustics emission, missile guidance).

6.1. Prediction of characteristics of stock market

Currency, international stock trading and derivatives contracts play an increasing role for many investors. Commonly used a portfolio consisting of a number of contracts. Assets returns must be predicted and controlled by a prediction/control module. Control of risk via prediction/control module of individual investments returns inside the portfolio provides the most likely process.

It is known that in most economic applications i.e. financial risk control, neural networks give success only of 70-80%. By means of the new approach of GMDH twice-multilayered neural networks it will be improved by 5-10%. Prediction accuracy for short and very noised data also increases in short and long-time predictions by 10-50% in comparison to statistical methods and neural networks, especially for stochastic processes [30,31]. On the base of predictive control it increases the results of a repetitive control.

As an example prediction of the activity on the stock exchange in New York was considered in [10]. In the following on the base of observations in the period of February 22 up to June 14, 1995 in seven periods 7 variables of the stock market (DAX, Dow Jones, F.A.Z., Dollar and other) are predicted. In the information base delays of all variables up to 35 are included. Also there were used not only linear reference functions to describe the variables, but also non-linear. It was to model and to predict 7 time series not independently as time series models but rather as highly interactions network (input - output - model). Table 3 shows the accuracy of predictions for all variables (mean MAD [%]).

Using the results of model generation (at first level of neuronet) it is possible to improve the accuracy of models in a second model generation, where are used the model outputs from previous for models generation of optimal complexity. This procedure can be continued up to decreasing accuracy of models.

Table 3. Observation and prediction periods.

	Observation period		Long-term prediction period			Model	Prediction
	Period up to	Days	Begin	End	Days	Max delay	Mean MAD [%]
a	March, 17	18	March,20	March, 31	10	5	0.985 %
b	March, 31	28	April,3	April, 18	10	10	2.055 %
c	April, 18	38	April, 19	May, 3	10	15	0.809 %
d	April, 28	46	May, 2	May, 15	10	20	1.642 %
e	May, 15	56	May, 16	May, 30	10	26	1.217 %
f	May, 30	66	May, 31	June ,14	10	30	1.206 %
g	June, 14	76	June, 16	June, 29	10	35	0.760 %

Table 4 shows the resulting model error (MAPE [%]) and prediction error (MAD [%]) of Dollar, Dax, F.A.Z., Dow Jones and the mean values for all 7 variables obtained on 3 levels. The table 4 shows that the repeated application of self-organization gives more accurate approximation, which results in better predictions in the second level. The models obtained in the 3 level are overfitted, therefore the prediction error increases.

Table 4: Multilevel application (model f).

Level	MAPE [%]			MAD [%]		
	1	2	3	1	2	3
Dollar	0.68	0.51	0.11	2.32	2.17	11.67
Dax	0.35	0.24	0.10	2.20	1.24	5.21
F.A.Z.	0.22	0.23	0.03	1.54	1.27	2.32
Dow Jones	0.27	0.16	0.06	2.15	0.84	4.84
Mean	0.267	0.184	0.051	1.43	0.98	3.67

The efforts in using the GMDH type neural networks are much less than in neural networks, where the architecture must be chosen by trial and error. Only an adaptive synthesis of the network structure allows an automatic model generation and therefore applications in the fields

where lots of decisions and forecasts (monitoring of complex systems with many controlled variables) repeating over short time periods are needed.

7. Objective selection of the best model

It is the aim of self-organising modelling to get in an objective way models of optimal complexity. But there are several freedoms in choice of class of systems to be model (linear/non-linear), time lag and in selection of appropriate parameters (number of best models, complexity etc.). To reduce such a subjectivity it is recommended to generate several alternative models (linear, non-linear, with several complexity and time lags) and in a second layer to select the best model outputs or to generate there combination. Table 5 shows obtained results.

Table 5. Selection of best model results (model g): prediction error MAD [%].

Model	Linear			Non-linear			Second layer
	1	2	3	1	2	3	
Dollar	2.88	2.10	0.89	1.25	1.41	1.40	1.55
F.A.Z.	1.22	1.45	1.01	0.82	1.12	1.57	0.88
Dax	1.36	2.41	1.51	1.69	2.43	4.54	1.94
Dow Jones	1.14	1.26	1.44	3.75	3.25	3.79	2.93
<i>Mean</i>	<i>1.14</i>	<i>1.29</i>	<i>0.90</i>	<i>1.21</i>	<i>1.35</i>	<i>1.81</i>	<i>1.20</i>

8. Non-parametric inductive selection methods

8.1. Modeling of fuzzy systems

The physical model is the best tool for function approximation and random process forecasting of deterministic objects where inputs and outputs are measured accurately with absence of noises. In the case of insufficient a priori information, not very accurate measurements, noisy and short data sample, better results can be reached by the use of non-physical models. But in the case of so-called ill-defined objects, dispersion of noise is too big, even for the use of non-physical models. In this case application of clustering of data sample is to be recommended, which can be considered as discrete form of physical model of ill-defined objects.

Almost all objects of recognition and control in economics, ecology, biology and medicine are undeterministic or fuzzy. Deterministic (robust) part and additional black boxes acting on each output of object can represent them. The only information about these boxes is that they have limited values of output variables, which are similar to the corresponding states of object.

According to Ashby [33] diversity of control system is to be not smaller, than diversity of the object itself. The *Law of Adequateness*, given by S.Beer, establishes that for optimal control the objects are to be compensated by corresponding black boxes of the control system [13]. For

optimal pattern recognition and clustering only partial compensation is necessary. More of what we are interested in is to minimise the degree of compensation by all means to get more accurate results.

The methods of cluster analysis and selection of analogous patterns discussed below are denoted as non-parametric because there is no need to estimate parameters. The method of cluster analysis was described in [20] in more detail.

8.2. Method of analogues complexing

The equal fuzziness of the model and object is reached automatically if the object itself is used for forecasting. This is done by searching analogues from the given data sample which are equivalent to the physical model. Forecasts are not calculated in the classical sense but selected from the table of observation data.

The main assumptions are the following:

- the system to be modelled is described by a multidimensional process;
- observations of data sample are enough (long time series);
- the multidimensional process is sufficiently representative, i.e. the essential system variables are included in the observations;
- it is possible that a part of past behaviour will be repeated.

If we succeed in finding for the last part of behaviour trajectory (starting pattern), one or more analogous parts in the past (analogous pattern) the prediction can be achieved by applying the known continuation of these analogous patterns [8].

Using a sliding window which generates the set of possible patterns $\{P_{i,k+1}\}$, where $P_{i,k+1} = (\underline{x}_i, \underline{x}_{i+1}, \underline{x}_{i+2}, \dots, \underline{x}_{i+k})$ and $k+1$ is the width of sliding window and also of the patterns, the output pattern is $P_k^A = P_{N-k,k+1}$.

The algorithm of selection of the analogous pattern has the following task:

For the given output pattern P_k^A it is necessary to select the most similar patterns $P_{i,k+1}, i \in J$ and to evaluate the forecast with the help of these patterns.

Method of analogue complexing is recommended in the case when the input observations of data sample is long enough. Analogues substitute the physical model. It means that optimal analogue can be found by selection sorting-out procedure, using internal accuracy type criterion. To divide data sample into two parts is not necessary. There should be provided several optimization of algorithms parameters, to rise up the accuracy of processes short-term forecasting. The selection task is a four-dimensional problem with the following dimensions:

- set of variables used;
- number of analogues selected;
- width of the patterns (number of lines, used in each);
- values of weight coefficients with which patterns are complexed.

As method of optimization the comparison of variants by internal criterion of accuracy is used. The criterion is calculated on the whole length of sample. This is the way of short-time forecasting problem solution on one step ahead. More difficult is the problem of long-term step-by-step random processes forecast. To select similar patterns from all possible patterns in the time series, the following steps are developed:

A. Reducing variable set size

The choice of an optimal set of variables can be realised by preselection. It is necessary to identify a subset of effective variables, which were defined as the nucleus [17].

One method of automatic generation of the nucleus is the automatic classification of variables by means of the algorithm of objective cluster analysis, described in [20,22]. Another method gives the GMDH algorithm for linear model construction. The models selected in the last layer indicate an ensemble of variables for which we have to seek the most consistent pattern analysis.

B. Transformation of analogues

Most processes in large-scale systems are evolutionary. In this case stationarity as one important condition of successful use of the method of analogues is not fulfilled. As time-series may be non-stationary patterns with similar shapes may have different mean values, standard deviations and trends. In the literature, it is recommended to evaluate the difference between the process and its trend, which is an unknown function of time. Another possibility gives the selection of differences where the criterion of stationarity is used as selection criterion. The results of the method of analogues depend on the selected trend function.

It is advisable to determine transformed patterns $P_{i,k+1}^* = (\underline{x}_i^*, \underline{x}_{i+1}^*, \dots, \underline{x}_{i+k}^*)$, where $\underline{x}_{jt}^* = w_{j0} + w_{j1}x_{jt}$.

The weights w_0, w_1 for each pattern $P_{i,k+1}$, $k > 1$ can be estimated by means of the least squares method, which gives not only the unknown weights but also the total sum of squares s_i^2 , which can be used in the following (step 3) as a measure of similarity.

C. Selection of the most similar analogues

The closest analogue is called the first analogue A_1 , the next one in distance A_2 is called the second analogue and so on until the last analogue A_F . Distances can be measured by means of the Euclidean distance of points of the output pattern and the analogue or by other measures of distance. In our case it is not necessary to find a proximity measure, but the total sum of the squares s_i^2 gives us information about the proximity between P_k^A and $P_{i,k+1}^*$.

D. Combining forecasts

Every selected analogous has its continuation which gives a forecast. In such a way we obtain F forecasts, which are to combine. In the literature there are several principles for combination of forecasts.

The unknown predictions \underline{x}_{N+k} , of the M systems variables can be assumed as a linear combination of the continuations of selected analogous patterns, i.e.:

$$\underline{x}_{N+j} = g_0 + \sum_{j \in J} g_k \underline{x}_{j+k+i}^*$$

The unknown parameters $g_0, g_j, j \in J$, will be estimated by means of parametric selection procedures e.g. using self-organising methods. The only problem is the small number of observations, i.e. the number of selected patterns.

8.3. Prediction of characteristics of stock market by analogues complexing

On the base of observations in the period of February 22 to May 30, 1995 (66 days) the analogue complexing algorithm was used. Table 7 shows prediction error (MAD [%]) of four variables (Dollar, Dax, F.A.Z., Dow Jones) and the mean prediction error over all 7 variables. The width of the patterns varies from 6 to 15 days.

Table 7. Prediction error (MAD [%]) of analogues complexing

Width	6	7	8	9	10	11	12	13	14	15
Dollar	2.61	3.28	2.62	2.79	2.91	2.91	1.86	1.48	2.16	8.96
F.A.Z.	1.418	2.609	1.597	1.485	1.187	1.391	1.435	1.869	0.723	1.118
Dax	1.427	1.702	1.7	2.307	2.962	2.761	2.761	2.612	2.372	1.122
Dow J	1.36	1.708	1.622	6.979	5.393	4.647	4.966	3.849	3.363	2.793
Mean	1.174	1.575	1.581	2.356	2.458	2.08	1.944	1.789	1.632	1.877

The forecasts are obtained by means of linear combination of the continuations of 5 selected analogous pattern, where the unknown weights g_j are estimated by means of parametric selection procedures.

References

1. Madala,H.R. and Ivakhnenko,A.G. Inductive Learning Algorithms for Complex Systems Modeling. CRC Press Inc., Boca Raton, 1994, p.384.
2. Müller,J.-A. and Ivakhnenko,A.G. Selbstorganisation von Vorhersagemodellen. Berlin, VEB Verlag Technik, 1984.
3. Ivakhnenko,A.G., and Osipenko,V.V. Algorithms of Transformation of Probability Characteristics into Deterministic Forecast. Sov. J. of Automation and Information Sciences, 1982, vol.15, no.2, pp.7-15.
4. Ivakhnenko,A.G., Peka,P.Yu., and Vostrov,N.N. Kombinirovannyj Metod Modelirovaniya Vodnykh i Neftianykh Polej (Combined Method of Water and Oil Fields Modeling). Kiev: Naukova Dumka, 1984.
5. Ivakhnenko,A.G., and Müller,J.A. Problems of Computer Clustering of the Data Sampling of Objects under Study. Sov. J. of Automation and Information Sciences, 1991, vol.24, no.1, pp.58-67.
6. Ivakhnenko,A.G., Petukhova,S.A., et al. Objective Choice of Optimal Clustering of Data Sampling under Non-robust Random Disturbances Compensation. Sov. J. of Automation and Information Sciences, 1993, vol.26, no.4, pp. 58-65.
7. Ivakhnenko,A.G. and Ivakhnenko,G.A. Simplified Linear Programming Algorithm as Basic Tool for Open-Loop Control. System Analysis Modeling Simulation (SAMS), 1996, vol.22, pp.177-184.
8. Ivakhnenko,A.G. An Inductive Sorting Method for the Forecasting of Multidimensional Random Processes and Events with the Help of Analogues Forecast Complexing. Pattern Recognition and Image Analysis, 1991, vol.1, no.1, pp.99-108.
9. Ivakhnenko,A.G., Ivakhnenko,G.A. and Müller,J.A. Self-Organisation of Neuronets with Active Neurons. Pattern Recognition and Image Analysis, 1994, vol.4, no.2, pp.177-188.
- 10.Ivakhnenko,G.A. Self-Organisation of Neuronet with Active Neurons for Effects of Nuclear Test Explosions Forecastings. System Analysis Modeling Simulation (SAMS), 1995, vol.20, pp.107-116.
- 11.Farlow,S.J.,(ed.) Self-organising Methods in Modeling (Statistics: Textbooks and Monographs, vol.54), Marcel Dekker Inc., New York and Basel, 1984.
- 12.Aksenova,T.I. and Yurachkovsky,Yu.P. A Characterisation at Unbiased Structure and Conditions of Their J-Optimality, Sov. J. of Automation and Information Sciences, vol.21, no.4, 1988, pp.36-42.
- 13.Beer,S. Cybernetics and Management, English Univ.Press, London, 1959, p.280.
- 14.Gabor D. Perspectives of Planning. Organisation of Economic Cooperation and Development. Emp. College of Sci. and Technology, London, 1971.
- 15.Belogurov,V.P. A criterion of model suitability for forecasting quantitative processes. Soviet J. of Automation and Information Sciences, 1990, vol.23, no.3, p.21-25.
- 16.Sawaragi,Y., Soeda,T. et al. Statistical Prediction of Air Pollution Levels Using Non-Physical Models, Automatica (IFAC), vol.15, no.4, 1979, p.441-452.
- 17.Ivakhnenko,A.G., and Müller,J.A. Parametric and Non-parametric Selection Procedures in Experimental Systems Analysis. Systems Analysis, Modeling and Simulation (SAMS), 1992, vol.9, pp.157-175.

- 18.Ivakhnenko A.G., Krotov G.I. and Cheberkus V.I. Harmonic and exponential-harmonic GMDH algorithms for long-term prediction of oscillating processes. Part I. Sov. J. of Automation and Information Sciences, v.14, no.1, 1981, p.3-17.
- 19.Ivakhnenko A.G., Krotov G.I. Multiplicative and Additive Non-linear GMDH Algorithm with factor degree optimization. Sov. J. of Automation and Information Sciences, v.17, no.3, 1984,p.13-18.
- 20.Ivakhnenko,A.G., Ivakhnenko,G.A., and Müller,J.A. Self-Organisation of Optimum Physical Clustering of the Data Sample for Weakened Description and Forecasting of Fuzzy Objects. Pattern Recognition and Image Analysis, 1993, vol.3, no.4, pp.415-421.
- 21.Triseev,Yu.P. Approaches to the Solution of Mathematical Programming Problems on the Basis of Heuristic Self-Organisation. Soviet J. of Automation and Information Sciences, 1987, vol.20, no.3, pp.30-37.
- 22.Zholnarsky,A.A. Agglomerative Cluster Analysis Procedures for Multidimensional Objects: A Test for Convergence. Pattern Recognition and Image Analysis, 1992, vol.25, no.4, pp.389-390.
- 23.Stepashko V.S. and Kostenko Ju.V. GMDH Algorithm for Two-Level Modeling of Multivariate Cyclic Processes, Sov. J. of Automation and Information Sciences, 1987, vol.20, no.4.
- 24.Ivakhnenko, A.G. and Stepashko,V.S., Pomekhoustojchivost' Modelirovaniya (Noise Immunity of Modeling). Kiev: Naukova Dumka, 1985.
- 25.Ivakhnenko,A.G. and Yurachkovsky,Yu.P. Modelirovanie Slozhnykh System po Eksperimentalnym Dannym (Modeling of Complex Systems after Experimental Data). Moscow: Radio i Svyaz, 1986, p.118.
- 26.Stepashko V.S. Asymptotic Properties of External Criteria for Model Selection, Sov. J. of Automation and Information Sciences, 1988, vol.21, no.6, pp.84-92.
- 27.Stepashko V.S. Structural Identification of Predictive Models under Conditions of a Planned Experiment, Sov. J. of Automation and Information Sciences, 1992, vol.25, no.1, pp.24-32.
- 28.Stepashko V.S. GMDH Algorithms as Basis of Modeling Process Automation after Experimental Data, Sov. J. of Automation and Information Sciences, vol.21, no.4, 1988, pp.43- 53.
- 29.Ivakhnenko, A.G., Müller, J.-A.: Present state and new problems of further GMDH development. SAMS, 20 (1995), no.1-2, 3-16.
- 30.Müller, J.-A., Lemke,F.: Self-Organising modelling and decision support in economics. In „Proceedings of the IMACS Symposium on Systems Analysis and Simulation“. Gordon and Breach Publ. 1995, 135-138.
- 31.Lemke, F.: SelfOrganize! - software tool for modelling and prediction of complex systems. SAMS, 20 (1995), no.1-2, 17-28.
- 32.Müller, J.-A.: Analysis and prediction of ecological systems. SAMS, 21 (1996).
- 33.Ashby An introduction to cybernetics. J. Wiley, New York 1958.
- 34.Ivakhnenko, A.G., Müller, J.-A.: Self-organisation of nets of active neurons. SAMS, 20 (1995) no.1-2, 93-106.