THE DEVELOPMENT OF SELF-ORGANIZATION TECHNIQUES IN MODELLING: A REVIEW OF THE GROUP METHOD OF DATA HANDLING (GMDH)

L. Anastasakis & N. Mort

Department of Automatic Control & Systems Engineering
The University of Sheffield
Mappin St, Sheffield, S1 3JD, United Kingdom

THE DEVELOPMENT OF SELF-ORGANIZATION TECHNIQUES IN MODELLING: A REVIEW OF THE GROUP METHOD OF DATA HANDLING (GMDH)

L. Anastasakis & N. Mort

Department of Automatic Control & Systems Engineering
The University of Sheffield
Mappin St., Sheffield, S1 3JD. United Kingdom

Abstract:

The necessity of modelling is well established since the structural identification of a process is essential in analysis, control and prediction. In the past, limited information on system behaviour has driven researchers to introduce modelling techniques with a broad range of assumptions on systems' characteristics. Statistical modelling methods, which are based on, these assumptions have generally failed to fully capture the dynamic characteristics of the process. The development of neural networks have partly improved the modelling procedure but their high degree of subjectiveness in the definition of some of their parameters as well as the demand of long data samples remain significant obstacles. On the other hand, real world systems like financial markets have a high degree of volatility and the utilisation of long data samples tends to remove and, in effect, filter the dynamic characteristics of the process. The Group Method of Data Handling (GMDH) belongs to the category of inductive self-organisation data driven approaches. It requires small data samples and is able to optimise models' structure objectively. In this report the stages of GMDH development and a broad spectrum of GMDH algorithms will be explored as well as a diversity of applications. A special study on the external criteria - a key feature in GMDH - will be also presented. Finally some key differences between neural networks and GMDH algorithms will be discussed.

Keywords: self-organising modelling, GMDH, data mining, prediction, external criteria

1. Introduction

Modelling real world systems is a difficult but essential task in control, identification and forecasting applications. Economic, ecological and engineering systems are generally complex with limited information on their underlying physical laws. Mathematical apparatus and assumptions about their features are necessary for their structural identification. The majority of those systems are consisted of subsystems with unclear and unknown interrelationships. In cybernetics the inherent behaviour of these elements is not as important as the exhibitory behaviour of the system is. Input and output variables of the system can be analysed, transformed and manipulated highlighting the system's exhibitory behaviour.

Complex systems have been characterised as black boxes where the only available information is the number and nature of input – output variables. Their investigation follows the principle of the black box concept according to which the internal mechanisms and relations amongst the elements of the system should be ignored while concentrating on the study of the relationships between input-output. Obviously, in the black box concept, the knowledge of the system can be extracted exclusively from the data. Either deductive or inductive sorting methods can be applied to identify and process this knowledge. Generally, deductive techniques should be applied on simple problems where the theory of the system is well established and physical laws can be applied to identify a physical model. Most engineering systems belong to this category while problems of economical or ecological nature due to the fuzziness of their objects require inductive sorting out methods. A first approach in modelling procedure with data was based on the well-known statistical methods and statistical regression analysis. These methods require a priori assumptions about the laws governing the data as well as their properties. These assumptions could reflect only special states of the object and therefore may produce inaccurate models.

The rapid development of artificial intelligence and neural networks in the last two decades due to the introduction of back propagation learning algorithms have urged a significant number of researchers to investigate its application in modelling tasks. However, neural networks despite the small number of assumptions in comparison to statistical methods require still a significant amount of a priori information about the model's structure. Experts should decide on the quality and quantity of input arguments, the number of hidden layers and neurons as well as the form of their activation function. Such an approach requires not only the knowledge about the theory of neural networks but also the rules for the translation of this knowledge into the language of neural networks [1]. The heuristic approach that follows the determination of network architecture corresponds to a subjective choice of the final model, which in the majority of the cases will not approximate the ideal. Sarle [2] comparing the behaviour of neural networks in data analysis to the statistical methods claims that it is not appropriate to be viewed as competitors since there are many overlaps between them.

A new approach, which attempts to overcome the subjectiveness of neural networks is an inductive approach, based on principle of self-organisation. An inductive approach is similar to neural networks but is unbounded in nature, where the independent variables of the system are shifted in a random way and activated so that the best match to the dependent variables is ultimately selected [3]. Following such a procedure there is a gradual increase of complexity and the optimum model is found with respect to the popular incompleteness theorem issued by Gödel in 1931. According to that, it is in principle impossible to obtain a unique model of an object on the basis of empirical data without using an external complement. The existence of a single optimum model is based on the principle of self-organisation which states that when a model's complexity gradually increases, certain criteria (i.e. selection criteria) or objective functions that hold the property of external complement pass through a minimum [4].

In inductive learning algorithms the experts have a limited role. Man communicates with the machine not in the difficult language of details but in a generalised language of integrate signals like selection criteria or objective functions [5]. On the other hand in deductive methods the dominant role belongs in the experts with computers being simple large calculators. The inductive approach have a similar concept to that of evolution introduced by Holland [6] where a number of solutions is created and an external criterion plays the role of finding the fittest.

2. The Group Method of Data Handling

Generally, the connection between input-output variables can be approximated by Volterra functional series, the discrete analogue of which is Kolmogorov-Gabor polynomial.

$$y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i-1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k + \ldots \tag{1}$$

where, x=(x_1, x_2, ……, x_m) the input variables vector and A=(α_0, α_1, α_2, ……., α_m) the vector of weights. The Kolmogorov-Gabor polynomial can approximate any stationary random sequence of observations and can be computed by either adaptive methods or a system of Gaussian normal equation [7]. However, in real world systems two problems forbid its establishment as an identification method. In majority of cases the vector of independent variables is long and incomplete while the set of observations is small. Furthermore, the computation time for solving all the necessary normal equations increases, as the input vector becomes wider.

Ivakhnenko [7], inspired by the form of Kolmogorov-Gabor polynomial, developed a new algorithm, which is called the "Group Method of Data Handling (GMDH)". Following a heuristic and perceptron type approach, he attempted to resemble the Kolmogorov-Gabor polynomial by using low order polynomials for every pair of the input variables. He proved that a second order polynomial (i.e. Ivakhnenko polynomial: $y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2$) can reconstruct the complete Kolmogorv-Gabor polynomial through an iterative perceptron type procedure. This approach offers better accuracy due to its perceptron type structure, which enables the classification of the information into "useful" and "harmful", requires a fewer number of observations and therefore reduces computation time. The GMDH method belongs to the category of heuristic self-organisation methods, where the black box concept, the concept of connectionism and induction are applied [8]. Ivakhnenko [9] claims that the self-organisation is necessary when it is impossible to trace all input-output

relationships through an entire system that is too complex. This ability made GMDH algorithms appropriate modelling procedure for real world systems.

During the modelling procedure, GMDH algorithm involves four heuristics that represent the main features of GMDH theory [9].

(i)   collect a set of observations that seems to be relevant to the object

(ii)   divide the observations into two groups. The first will be used to estimate the coefficients of model while the second will separate the information embedded in the data into either useful or harmful. Strictly speaking: 'no partition of the data, no GMDH' [10]

(iii)   create a set of elementary functions where complexity will increase through an iterative procedure producing different models

(iv)   according to Gödel's incompleteness theorem, apply an external criterion to choose the optimum model.

The GMDH has developed rapidly, particularly during the eighties where its theoretical background was formulated [10]-[11]-[12]. Ivakhnenko [10] points out that GMDH algorithms have been primarily developed on the basis of numerous computational experiments and in analogy to the justification of the Monte Carlo methods of statistical trials, multiple repetitions of an experimental result constitute its proof. Stepashko et al. [13] add that due to the absence of a satisfactory mathematical foundation for the model under other than statistical assumptions, the GMDH theory was developed as a division of regression analysis in its early steps. This lack of theory has been criticised by the researchers and a number of theoreticians have attempted to justify some of the aspects of GMDH theory like the convergence of multilayer algorithm [14]. Furthermore, Ivakhnenko & Kocherga [15] present a series of theorems for long-range two levels forecasting with GMDH, which can serve as a basis for a future general theory in that class of forecasting tasks.

Despite its limited theoretical background, the extensive number of GMDH algorithms and their ability to model ill-defined objects with satisfactory accuracy over other known statistical methods, have proved and strengthened its position as an appropriate non-linear method for structural identification and prediction tasks. The wide range of partial descriptions allows its application in different fields of modelling procedures as well as producing a respective number of GMDH algorithms.

3. GMDH Algorithms

The wide development of GMDH theory has led to a broad spectrum of algorithms with each one of them corresponding to some specific conditions of a particular application. GMDH algorithms may differ by the type of elementary function, the way of model's structure complexing, the external criteria or the type of modelling task itself. The choice of the algorithm depends on the level of noise in the data, their sufficiency as well as their type (e.g. continuous or discrete). The first GMDH algorithms had adopted three different types of elementary functions like, probabilistic graphs, Bayes' formulas and second order polynomials [9]. The rapid development of GMDH theory and the broad spectrum of its algorithms have resulted to a different classification approach where GMDH methods are grouped into two main categories, the parametric and non-parametric algorithms. The level of language representation of the system is used to distinguish between them. Parametric algorithms are recommended to describe systems characterised by either exact or low variance noisy data. On the other hand, in the case of ill-defined systems and high variance noisy data, the application of non-parametric algorithms is justified.

Parametric algorithms have been classified further according to the type of activation function (partial descriptions) or the type of model structure complexity. Combinatorial algorithms, also known as single-layer self-organising algorithms, perform an exhaustive search between all candidate models. Multilayer or iterative algorithms apply an iterative procedure, which increases model complexity while an external criterion identifies those models to be progressed in the next layer. In multilayer algorithms there is no exhaustive search of all candidate models but the computation time is reduced and the number of independent variables to be processed becomes larger. With respect to the type of activation function, GMDH algorithms are distinguished into polynomials, harmonic, multiplicative-additive and fuzzy. All these algorithms will be thoroughly discussed in the following sections. The adaptive nature of GMDH algorithms has allowed modifications of the basic GMDH so that they can better compensate for peculiarities in each and every application separately.

3.1 Combinatorial Algorithms (COMBI)

The classical combinatorial GMDH algorithm generates models of all possible input variable combinations and selects a final best model from the generated set of models according to a chosen selection criterion. Ivakhnenko et al. [16] describe combinatorial algorithm as a method of complete mathematical induction since not any possible model is missed to consider. Combinatorial algorithm sorts the models by gradually increasing the terms from 1 to n (i.e. the number of arguments) while an external criterion will indicate the optimum solution between models with the same complexity. The external criterion will have a minimum value in the plane of complexity vs. selection criterion, which corresponds to the optimum non-physical model.

The main disadvantage of combinatorial algorithms is related to computer capacity. Due to the full sorting out procedure the computation time is high and therefore the number of terms and power in the final model will be limited. Ivakhnenko et al. [16] propose two different ways to increase the number of terms in the final model up to 23-25 from the 18 that was initially proposed. The first approach indicates the calculation of the coefficients in the successfully obtained model with the help of a bordering method, which allows the processing of up to 25 variables. A second approach is known as the method of truncation of the sorting triangle according to which every model with more than a specified number of terms (17 or 18) is discarded from the sorting out procedure. Obviously, the second method violates the property of exhaustive searching for the optimum model but allows the processing of more complicated models. Problems with computer resolution were found in multilayer algorithms also. Ivakhnenko et al. [17] introduce a GMDH algorithm that achieves an increase in the input domain from 80 up to 120 by selecting randomly the number of pairs of input arguments. However, multilayer algorithms can produce models with up to 1000 terms in the final solution.

An additional shortcoming of combinatorial algorithms, which is common to all polynomial GMDH algorithms is the biased estimates of coefficients due to the method of least squares. Ivakhnenko et al. [18] argue that the method of instrumental variables could replace least squares producing less biased estimates. The advantages of single layer self-organising could be summarised in their simplicity and their ability to perform complete sorting of model structure. However, as the number of terms increases so does the number of required observations, which restrict them from ill-defined systems characterised by incomplete input vector and small number of available observations.

3.2 Multilayer Algorithms

The Multilayer GMDH algorithm was the first algorithm to be introduced by Ivakhnenko [9]. Its structure is very similar to that of multilayer feedforward neural networks but the number of layers as well as the number of nodes is objectively defined by an external criterion in accordance with the incompleteness theorem. Looking at figure 1 it is clear that in the first layer the number of nodes is equal to the number of inputs and for the subsequent layers it is equal to the number of pairs of variables for the characteristic vector (i.e. input variables). The number of hidden units can be either prespecified or change from layer to layer according to a threshold value of the external criterion. Multilayer algorithms do not perform an exhaustive search amongst all the candidate models but if the number of selected models in every layer is large enough, the optimum solution will never be lost [4].

The most popular type of activation function is the second order polynomial but a number of alternative partial descriptions have been also tested. Stepashko [19] provides a list of partial description as well as external criteria that have been used by researchers. Despite the popularity of polynomial partial descriptions, many researchers have applied different types of activation functions that enable them to transfer the concept of GMDH into different fields of research. Dimitrov [20] replaced the polynomial descriptions with probability distributions transferring the concept of GMDH into probabilistic models. In the probabilistic domain, the Kullbank approach is the analogue to the Kolmogorov-Gabor polynomial, stating that a nth-order probability distributions can be approximated by distributions of lower orders. The convergence of this algorithm is ensured by the theorems of convergence for the Kullbank approach. Additionally, Patereu et al. [21] presented the principle of self-organisation into pattern recognition designing a Bayessian type algorithm for classification where its main steps are analogues to GMDH.

$F(x_i, x_j)$

x_1
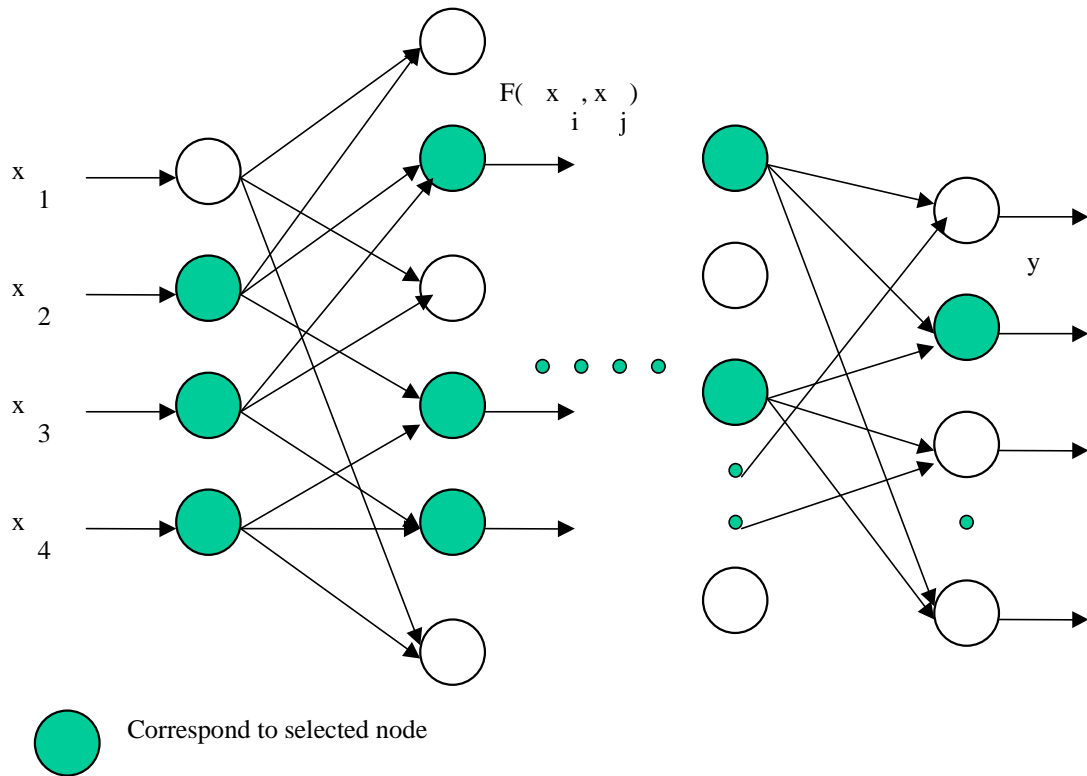x_2
x_3
x_4

y

⬤ Correspond to selected node

Figure 1. Multilayer GMDH Algorithm

A wide range of polynomial activation functions has been suggested to either increase accuracy or make the algorithm more compatible with the features of a specific application. Multilayer GMDH algorithms are superior to combinatorial algorithms with respect to the maximum number of terms in the final model. Ivakhnenko et al. [18] compared them and concluded that in multilayer algorithms there is an exponential growth to the volume of computation while increasing the number of variables in the data sample. They argued that it is better to apply them in cases where the number of input variables is small like underdetermined and ill-defined systems. However, they can only form single output equations, which restricts them to applications where the output is known a priori.

Triseyev [22] overcomes the above drawback, by designing a multilayer algorithm for multi-output systems. The only condition is that the output should be known a priori. Experts make the decision upon which variables of the system should be included in the output vector however, Ivakhnenko in [23] proposes some useful hints about that problem. Triseyev claims that it is better to underdetermine the system rather than overdetermine it because it is ensured that the system's laws would be exposed even though the accuracy may be lower. Polynomial partial descriptions of all the outputs are constructed and a rule R will select the system of equations, which approximates them best. The rule R is introduced to regulate the formation of the systems of equations since it is computationally expensive to perform complete sorting of all systems of equations. A distinctive feature of this algorithm is the introduction of an additional criterion in every layer (i.e. balance of variables criterion), which takes into account the forecasting background and decides on the candidates for the system of equations.

Yurachkovskiy [14], investigating the theoretic background of multilayer algorithms, proposes two properties that should be held by every multilayer algorithm. Firstly, the search in the set of structures must proceed in the direction of increasing complexity as well as the greatest degree of terms and secondly for any admissible structures a path leading to it from the previous layers must exceed. The "precision algorithm" which is a multilayer two-stage algorithm that satisfies these properties is also proposed by the author. In the first stage the search is limited to monomials with increasing powers. At the second stage the search is extended to all possible models having one or two terms within the partial descriptions. Ivakhnenko et al. [16] classify multilayer algorithms into those with calculation of remainders after each selection layer and those without calculation of remainders. In the latter class, the external criterion has a monotonic structure where in the plane complexity vs. criterion the area around the optimum models being flat. The application of "left corner rule" is proposed to overcome it

however, the subjectiveness of choice is increased. According to the left corner rule, at each layer the model is to be chosen not only among the models of given layer, but among the models of last two layers. The solution of the above two algorithms can coincide only for a rather large sample of initial data and in case of using the minimum bias criterion.

Multilayer GMDH algorithms have been applied in a broad spectrum of application and several modifications have been designed to increase their accuracy. In a later section all these modifications will be presented as well as the proposed solutions to their drawbacks, like multicollinearity, complexity and overfitting.

3.3 Harmonic Algorithms

Harmonic GMDH algorithms have been designed for oscillatory and periodic processes. Their functional form can resemble that of either combinatorial or multilayer algorithms where polynomial terms are replaced by harmonic components in the form (2).

$$y = a_0 + a_1 \sin \omega t + a_2 \cos \omega t \tag{2}$$

Stepashko et al. [24] introduce a polynomial GMDH algorithm to predict multidimensional cyclic processes via linear two level difference models, concluding that it is efficient for the construction of two level models of various natural cyclic processes. However, due to the periodic nature of such systems the utilisation of harmonic components should be explored. Vysotskiy et al. [25] describe two harmonic algorithms with prespecified non-multiple discrete frequencies. The balance of variables criterion replaces the regularity criterion since it can heuristically find a property of the process that holds exactly in both interpolation and extrapolation intervals. A modified form of the criterion appropriate to harmonic processes is also given.

In the above harmonic algorithms, the number of frequencies is prespecified which generates a significant high possibility of excluding the optimum harmonic. An increase in the number of harmonics should be avoided due to high computational load. Ivakhnenko et al. [16] propose a modified harmonic algorithm which allows the expression of frequencies and amplitudes of a process model as a sum of harmonics with non-multiple frequencies. A similar modified harmonic algorithm uses sum of harmonics where the frequencies are not assigned in advance but are analytically determined in order to become close to the true harmonics of the oscillatory process. Harmonic components have been also applied in multilayer algorithms without calculations of remainders where the spectrum of frequencies is expanded by using non-linear functions.

Sarychev [26] proposed a different form of harmonic component (3), which create a sequence of harmonics for different values of t and specified ω, φ.

$$x = a \sin(\omega t + \varphi) \tag{3}$$

The main feature of this algorithm is that the values of ω, φ are not prespecified but the computer objectively determines them during the modelling procedure. The values of ω are coming from the local minima of the intensity function I(ω) which is created by the spectral analysis of time series while those of φ are calculated according to the period of each harmonic.

A priori information about elements of the system and their behaviour may indicate the form of functional descriptions. Ivakhnenko et al. [27] observe that some processes exceed both an exponential and harmonic behaviour and therefore propose an appropriate GMDH algorithm, which can perform either an exhaustive or a limited search.

A list of problems with respect to the estimation of the coefficients of the harmonic components is presented in [28]. Generally, least squares method cannot be applied in these algorithms as the unknown parameters ω are contained non-linearly in the model's final equation. Shelekhova [28] proposed a new parameter estimation scheme where least squares have been replaced by orthogonal regression. Additionally, a formula is given calculating the maximum number of data under which least square estimates will not be considered as inconsistent. A complete review of GMDH harmonic algorithms can be found in [16] where the analogy between polynomial and harmonic algorithms is also highlighted.

3.4 Fuzzy Algorithms

In the conventional GMDH method the deviation between the observed values of the output and its estimates is supposed to be Gaussian-distributed and hence linear regression analysis is used to estimate the parameters. However, such an assumption is often violated and the method of least square estimates is not appropriate. Furthermore, the majority of real world systems seem to follow Zadeh's principle of incompatibility and fuzzy theory is a more appropriate approach in modelling procedure. Hayashi et al. in [29] attempt to represent input-output relationships with possibility models where the parameters are fuzzy numbers. The structure of the GMDH algorithm remains the same but fuzzy parameters are used in partial descriptions while possibilistic linear regression is applied to find them. The above algorithm is one of the first applications of fuzzy methods in GMDH theory and this is where Japanese researchers have focused their attention in the recent years. Fuzzy GMDH algorithms have a structure similar to that of conventional algorithms but models are replaced by fuzzy rules. Matushita et al. [30] present a GMDH algorithm that follows the principle of a hierarchical fuzzy modelling method. It is able to identify comprehensible rules searching simultaneously more potential structures of fuzzy models while avoiding the exponential increment of fuzzy rules for multi-input systems.

Yokode et al. [31] point out that fuzzy rules can be obtained by either expert knowledge or numerical data using neural networks. However, when applying neural networks, an optimisation problem appears. This problem can be eliminated by applying fuzzy if-then rules with a certainty factor and introduce a multilayer GMDH algorithm where polynomial descriptions are replaced by these rules with a fuzzy reasoning method. Its main advantage is that multi-input systems using this multilayer structure of GMDH with the pair of fuzzy sets in each fuzzy rule avoid an explosion in the number of rules.

A number of Japanese researchers during the last decade motivated by the similarities between neural networks and GMDH designed a series of multilayer algorithms that were introduced as GMDH neural networks. The main characteristic of these algorithms is the adoption of internal criteria in the selection process, which do not require the partition of the data sample. Ivakhnenko [32] criticised this work arguing that one of the main features of GMDH theory is the application of principle of external complement and therefore any algorithm that does not satisfy this principle should not be characterised as a GMDH algorithm. However, Ichihashi et al. [33] and Nagasaka et al. [34] proved that Akaike's Information Criterion (AIC), as well as the differential minimum bias criterion (DMC), are capable of selecting an optimum model and therefore the number of hidden layers of GMDH NN. In this review we will accept the argument made by Ivakhnenko [32] and call all these algorithms as GMDH-type NN. Details of these algorithms will be presented later but at this point we shall focus on those which appear to contain a fuzzy or neurofuzzy approach.

Nagasaka et al. [35] develop a neurofuzzy GMDH algorithm that seems to perform identification tasks better than the conventional GMDH. The name 'neurofuzzy' derives from the utilisation of Gaussian Radial Basis Functions as partial descriptions. According to Brown & Harris [36] an RBF can reinterpreted as both a simplified fuzzy reasoning and a three layer NN, therefore the GMDH type NN can be called as NF-GMDH type NN. However, the main shortcoming of the algorithms described in [31] and [35] is the limited number of fuzzy model structures among a wide range of potential structures. Ohtani et al. [37]-[38] propose a method that eliminates 'harmful' weights of a NF-GMDH network. This technique will be able to identify the optimal architecture of a GMDH type NN. The proposed method is a two-stage technique where, in the first stage, a destructive technique (i.e. structural learning with forgetting) is applied that constantly eliminates weights with less influence on the output. In the second stage, for every input of each neuron the Minkowski's norm is calculated which identifies the links (i.e. pair of inputs) that should be deleted.

Park et al. [39] indicate that fuzzy polynomial neural networks can be considered as a means to overcome the weakness of mathematical modelling methods. Fuzzy rules are applied in the first layer of the NN while polynomial partial descriptions are explicitly applied in the subsequent layers. Comparing fuzzy polynomial neural networks to other fuzzy modelling methods, it is concluded that they are able to produce better results in most of the cases considered. A software package known as 'Design' is produced by Zaychenko et al. [40], which implements a fuzzy GMDH algorithm. The proposed algorithm uses partial linear descriptions. However, non-linear terms can be included if they are a priori transformed into linear approximations. The coefficients of partial descriptions are not calculated by least squares but are the centres of triangle fuzzy numbers and their width.

3.5 Active Neurons

Similarities between the structure of neural networks and GMDH algorithms have urged researchers to explore areas where these two concepts could combine. Inductive methods cannot substitute the necessary analysis of causes of events by means of theoretical systems analysis and hence a pragmatic solution to the model building problems should be explored in a union of deductive and inductive methodology [41]. Ivakhnenko et al. [42] propose such a combined method, which extend the theory of self-organisation from isolated models to active neural networks.

The proposed algorithm is known as 'neural networks with active neurons" where in place of passive neurons GMDH algorithms are applied. Neural networks with active neurons have a two fold multilayered structure since neurons themselves are multilayered and can be united into common matrix in a multilayred way [8]. These networks have the ability to simulate in a more suitable way the structure of human brains, proving that neural networks can be characterised as an accurate description of human brain. Generally, statistical learning networks (like GMDH algorithms) follow two different strategies for the synthesis of the network depending on whether the structure of the network is fixed or allowed to evolve during the synthesis [43]. In the latter case no a priori information about the target is needed and the network architecture is chosen objectively avoiding the overfitting problem. It is also proved that least squares estimates can find the network architecture better than recursive algorithms do.

Both multilayered and combinatorial GMDH algorithms can be used as active neurons. Ivakhnenko [44] proposes a modified combinatorial algorithm as active neurons, which is shown to increase accuracy and reduce computation time. The input dimension is extended by adding pair of inputs in several forms such as square roots as well as the inclusion of outputs from the previous layer into the new layer. Additionally, a variables protection system is developed according to which experts will propose a number of variables that should be definitely included in the final model and these will never be eliminated during the sorting out procedure. Finally, a discrimination criterion will calculate the quality of every model attempting to narrow the space of candidate solutions. The narrowing of argument set is realised by the exclusion of arguments, which are not included to the set of second layer equation arguments. In that way only a number of the best variables will continue in the sorting out procedure until the minimum of the criterion is achieved.

A neural network with active neurons has the ability to select the most relevant input arguments thus increasing its accuracy in prediction tasks and overcoming the inscrutability of neural networks with logistic or other passive neurons [45]. Additionally, the accuracy is increased since the output of active neurons generates new variables, which can be used as a new factor in the next layer of neurons and the set of input factors can be optimised at each layer [46]. Neural networks with active neurons have been applied in prediction of economic and ecological systems where their superiority is clear [47]-[43]-[44].

3.6 Non Parametric Algorithms

Clusterization is a weak mathematical method of modelling, which in accordance to the law of adequacy of an object established by Beer [48], it is the most appropriate technique for modelling ill-defined, complex real world systems. The development of non-parametric GMDH algorithms have been motivated by the weakness of parametric algorithms as well as the advantage of a weak mathematical method to overcome effectively some difficulties of underdetermined tasks of experimental systems analysis (i.e. Nalimov's approach) [49]. Weak methods are able to produce descriptions of objects whose properties are compatible to the properties of those with fuzzy characteristics, overcome difficulties caused by great size of initial data and allow the utilisation of internal criteria in model exhaustion and clustering. This enhances the reliability of coefficients and does not require the division of data samples into two groups [50].

Cluster analysis is a three dimensional problem where the width of cluster, the length and their numbers must be found. Generally, this problem can be solved by either a deductive or an inductive sorting procedure [51]. Deductive methods are subjective since they are totally based on detailed instructions from the person performing the clustering as well as his/her idea about the object behaviour. On the other hand inductive methods of clustering, based on the principle of self-organisation, are objective and hence more appropriate to such modelling tasks. Ivakhnenko et al. [52]-

[53] argue that the analogy between self-organisation of models and self-organisation of clusters enables the transfer of the concept of a sorting model to that of sorting clusters. Computers can successively replace human beings in clusterization and using an external criterion objectively divide a given set of objects representable by points in a multidimensional space of attributes, into a given number of compact groups or clusters. Ivakhnenko et al. [54] claim that clusterization should be accurate by successfully dividing the observational data and consistent so that hold for all possible sets of data obtained on the same observed object. Humans may not be able to do this but computers following the principle of self-organisation can do it easily.

The objective computerised clusterisation (OCC) algorithm has the ability of performing objective clusterization as well as the orientation to search for those clusterizations that are unique and optimal for each level of noise variance [53]-[55]. The computerised clusterization procedure is affected by the completeness of the source data representation. If it is complete, a polynomial GMDH will indicate the optimal composition of input vector and then the clusterization procedure will take place in the combined input-output space [53]. In case of incomplete data representation where the dimension of the space of the goal function is not known, either experts will indicate it, or the Korhunen-Loeve orthogonalised expansion will be used [55]. Another important aspect in clusterization is the procedure of finding the effective ensembles of attributes and Ivakhnenko [55] proposes three different techniques. The first approach involves sorting of all possible clustering of different ensembles of attributes but is clearly time consuming. The second approach involves the utilisation of a combi-GMDH to find the most effective arguments, while a third approach introduce the utilisation of the correlation algorithm of 'Wroclaw taxonomy", which is strongly recommended in cases where the length of data is less than half of the number of attributes. According to Wroclaw taxonomy algorithm, the features are ordered according to their efficiency and then exclude them one by one in the process of optimisation of the event forecast algorithm to find the optimal feature set [62].

The selection of optimum clustering is based on the minimum of consistency criterion. However in [56] it is claimed that the application of a regularisation criterion into a third independent set of variables can avoid false clusterings which may be indicated by the consistency criterion. Ivakhnenko et al. [57] propose the utilisation of a balance of variables criterion where the samples are partitioned into two levels (e.g. N and N/2 – 1) instead of two subsamples. For each level a different clusterization takes place and the one, which minimise the balance of variables criterion is chosen. In contrast, similar to parametric GMDH algorithms, an auxiliary criterion like the symmetry regularity criterion will ensure the regularisation of solution [50]. In the same article it is pointed out that despite the way that clustering is found it is important to test if the selected clustering corresponds to the properties of the real. This can be tested by the multirow theory of statistical decision algorithms, which will improve the selected clustering so that it becomes closer to the real. Ivakhnenko et al. [58] highlight that the basic OCC fails to find an optimal clusterization when the constructed dipoles (i.e. a graph that connects two closest points) are large and a modified OCC (modOCC) is proposed. The modOCC is based on the calculation of the balance criterion of the clusterizations obtained by constructing two trees. One the initial data sample and the other on a sample the elements of which are calculated by a weighted summing formula. The weighting summing formula takes into account both the initial matrix of data and the matrix of each row.

Non parametric algorithms have been applied in a variety of financial applications [59] and clustering has been a more efficient method for long range prediction [51]. The reasons for this are the problems of estimating coefficients for the difference models which are not well defined since the number of variables is usually larger than the number of observations. Additionally, the weak mathematical language for the description of an object is required for solving the problem of maximal increase of the limiting achievable lead-time for forecasting random processes and events. Muller [49] points out the advantage of cluster analysis is, in general, to provide more exact predictions than parametric algorithms do, which will not be as detailed as the latter. However, it would be useful to apply a combination of those two where GMDH type sorting out algorithms will be used to validate the width and number of clusters.

3.7 Method of Analogue Complexing

In parametric GMDH algorithms where regression analysis techniques are used in the parameter estimation, it is essential to include all relevant regressors in the input vector. The exclusion of essential regressors will appear as noise and will effect the model performance. In the majority of real

world systems is impossible to know in advance the complete set of regressors so parametric GMDH algorithms usually fail. Non parametric GMDH algorithms are appropriate modelling methods to system with fuzzy characteristics and limited information since the equal fuzziness of the outputs is reached automatically if the output itself is used for forecast [8].

The method of analogues, which satisfy the above arguments have been successfully applied in meteorology in the past. A long history of the output under investigation is searched in order to find a state (i.e. pattern) which will be similar to the current known as reference analogue. Then the continuation of that analogue in the history can be considered as the next state of the output (i.e. reference analogue) as well. Ivakhnenko et al. [60] introduce the concept of induction into the method of analogues where the length and number of analogues in the analogue group is objectively chosen by means of an objective computer clusterization (OCC) algorithm. A significant disadvantage of that procedure is the requirement of a long prehistory as well as the identification of one very accurate analogue. The combination of a group of analogues may increase its accuracy and eliminates the above drawback. The method of analogue complexing involves four basic steps. Parametric GMDH will be applied to reduce input dimension, the analogues should be transformed so that stationarity is fulfilled, the most similar analogues will be selected according to a proximity criterion and finally analogues will be combined by using a GMDH algorithm to determine the parameters of the complexed analogues [8].

The Euclidean distance for short-range predictions and the canonical correlation coefficient for long-range predictions can be the proximity criterion that will determine the degree of similarity between the candidates of analogue patterns in prehistory [61]. The combination of analogues can be done by either applying a combinatorial GMDH using a consistency criterion or following a different method based on the extrapolation of forecasting space by spline functions [61]-[62]. The basic interval of data is subdivided in subintervals and different functions are found in every subinterval. Then, these functions are joint together at the interval endpoints in such a way that a certain degree of smoothness of the resulting functions is guaranteed. The latter method involves an exponential factor which value cannot be optimised but is subjectively chosen and it is proposed to be taken smaller by one the dimensionality of the analogue space. Regarding the number of complexed analogues there is no restriction but it has been found that more than three do not significantly improve the accuracy [61]. Ivakhnenko et al. [63] point out that with respect to the selection criterion used by Combi algorithm to find the combination of these analogues two different approaches could be followed. In the exactness approach a cross-validation criterion can be applied while in a robust approach the criterion of balance of discretization is preferred.

A key sensitive issue in the method of analogue complexing is the requirement of stationarity. Using a stationary series coming from the difference between the trend and actual series can solve this problem. The trend function can be found in an objective way by applying a parametric GMDH algorithm [49]. However, due to the high dependence of analogue complexing on the selected trend function an alternative four-step procedure is described in [64]. Due to the evolutionary nature of many processes like financial systems, it is possible for similar patterns to have different mean, standard deviation and trends therefore the choice of most similar patterns in past is a difficult task. Lemke and Muller [64] overcome that difficulty by measuring the 'similarity' between the reference pattern and a transformed pattern which describes the difference between the similar analogues characteristics. The transformed pattern will be a linear combination of the original pattern where least square method will be used for parameter identification. As measurement of similarities between two patterns, the total sum of squares obtained by least squares estimation of unknown parameters is considered.

According to [64] the combination of analogues can be difficult in case of small patterns length (i.e. patterns consisted by less than ten observations) therefore it is proposed to use the weighted mean of the continuations of the selected analogues. Two alternative methods are proposed for the combining procedure in [65]. The first is the 'rigid complexing' where a specific form for estimating the values of the weights is given. The second is the 'relaxed complexing' where the parameters of the rigid complexing are slightly varied and according to the forecasting variance criterion RR (i.e.

$$RR = \frac{1}{N-1} \frac{\sum_{i=1}^{N-1}(y - y_{pr})^2}{\sum_{i=1}^{N-1}\left(y - \bar{y}\right)^2}$$, where N number of data, y_{pr} the predicted output and $\bar{y}$ the mean output)

the optimum values are obtained. The rigid approach will produce values close to the optimum and the

relaxed method will slightly improve its prediction accuracy. However, the authors argue that the forecasting accuracy can be improved further by constructing a neural network with active neurons. In that case the active neurons will follow the method of analogue complexing.

4. GMDH vs. Neural Networks

In the early stage of the development of GMDH theory the similarity between neural networks and multilayer GMDH algorithms had been highlighted. Ivakhnenko in one of the introductory articles claims that since the differences between perceptron and GMDH are neither significant nor fundamental it is appropriate to call GMDH systems as "systems of perceptron type" [9]. Generally, either deductive logical mathematical techniques or inductive sorting out methods could solve modelling complex objects [66]. According to Muller et al. [1], statistical methods, neural networks and statistical neural networks are deductive methods that cannot identify complex objects because of their requirement for a large amount of a priori information. On the other hand GMDH algorithms can be considered as regression based method that combines the best of both statistic and neural networks while embedding the very important additional property of induction [67]. Based on that property GMDH algorithm are capable of overcoming the drawbacks of NN while statistical neural networks can partly solve them.

Muller et al. [1] and Madala [3] have thoroughly investigated the differences between deductive and inductive methods. The major differences can be summarised in the ability of GMDH to objectively select the optimum model, avoids overfitting problems and selects the most relevant input variables. In table 1 the features of both NN and self-organising modelling in a variety of categories can be found.

	Neural Networks	Self-Organising Modelling
Data analysis	Universal approximator	Structure identificator
Analytical model	Indirect approximation	Direct
Architecture	Preselected unbounded network structure, Experimental selection of adequate architecture demands time and experience	Bounded network structure Structure evolved during the estimation process
Network synthesis	Globally optimised fixed network structure	Adaptive synthesised structure
Threshold	Threshold transfer functions	Threshold objective functions
Self-organisation	Deductive, given number of layers and number of nodes	Inductive, number of layers and of nodes estimated by minimum of external criterion
Parameter estimation	In a recursive way demands long samples	Estimation in batch by means of maximum likelihood techniques using all the observational data, extremely short samples
Optimisation	Global search of a highly multimodal surface, result depends on initial solutions, slow and tedious, requiring the user to set various algorithmic parameters by trail and error, time consuming techniques	Group method of data handling, not-time consuming technique adaptively synthesised networks are more parsimonious, parts of the network which are inappropriate are automatically not included
On/off line	Observation is available transiently in a real-time environment	Data are usually stores and repeatedly accessible
Regularisation	Without, only internal information	Estimation on training set, selection on testing set
A priori information knowledge	Without transformation in the world of neural networks not usable	Can be used directly to select the reference functions and criteria
Knowledge	Needs knowledge about the theory of neural networks	Necessary knowledge about the task and the class of system (linear, non-linear)
Convergence	Global convergence is difficult to	Existence of a model of optimal

	guarantee	complexity
Computing	Suitable for implementation in hardware using massively parallel computation, for ordinary computers insufficient	Efficient for ordinary computers but also for massively parallel computation
Feature	General-purpose, flexible, non-linear (especially linear) static or dynamic nonparametric models	General-purpose, flexible linear or non-linear, static or dynamic parametric models

Table 1. Comparison of neural networks and self-organising modelling

The similarities between the structures of NN and GMDH have urged Pham and Liu in [68] to produce a GMDH neural network where N-adalines have replaced the Ivakhnenko polynomial in every processing unit. N-adalines are adalines with second order non-linear preprocessors the weights of which correspond to the polynomial coefficients of the original GMDH network. Furthermore, the Widrow-Hoff learning rule and not the least square method is applied to estimate the weights.

Similarly, neural network with active neurons is another example of the combination between neural networks and GMDH theory. The modelling ability of neural networks and GMDH has been tested in a number of applications but the conclusions do not indicate a clear winner. Madala [3] observed that GMDH produce better results in an identification task of linear systems but logistic neural networks were better in the non-linear case. On the contrary, Parker and Tumala [69] claim that GMDH performed better in non-linear systems than neural networks do.

It is clear that there is no universal accepted method and the choice of the most appropriate technique depends on the application itself, the level of noise within the data and the fuzziness of the object. Muller in [41] strongly recommends a union of deductive and inductive methodologies as the only pragmatic solution to the model building problem.

5. GMDH type Neural Networks

The basic features of GMDH theory are the partition of data sample into two subsamples, the evolution of a partial description to the final model and the application of an external criterion, which value passes its minimum when the optimum model have been found. The application of external criteria is important in order to objectively identify the optimum model and Ivakhnenko in [10] point out in a strict sense that "no external criterion, no GMDH". Japanese and American researchers have argued that this last feature is narrowing the range of applications for GMDH and attempt to apply criteria that do not require data partition [70]. Ivakhnenko [32] criticises this work claiming that the application of selection criteria that do not require partition of the data (i.e. internal criteria) can identify only physical models. Ivakhnenko et al. [71] point out that these algorithms are deductive methods whose solutions will coincide to that of GMDH only in case of complete input vector. On the other hand GMDH algorithms as inductive algorithms can be applied in cases of noisy data and select non-physical models in accordance to the Shannon's theorem in communication. Muller et al. [1] classifies these algorithms in the category of GMDH type NN since they satisfy the expression "the more complex the model is the more accurate it is".

On the other hand Tamura et al. [72] questions those basic features of GMDH (i.e. heuristics) characterising them, as drawbacks that should be eliminated. A heuristic free GMDH algorithm is proposed where the division of data is avoided, the number of selected intermediate variables in every layer is not predefined and four different partial descriptions are applied to increase flexibility. The Akaike's Information Criterion (AIC) in the whole set of data is used as selection criterion while the number of selected models is only limited by the computer resolution. Four different generators of partial descriptions are used which include linear, second and high order polynomial as well as optimal partial polynomials. The last type of generators can be either optimal partial polynomials in which parametrically unstable terms in the second order polynomial are eliminated, or polynomials in which redundant terms in the second order equation are eliminated by applying a stepwise regression method using a test of significance for the training set.

Sawaragi et al. [73] proposed a slight modified "revised GMDH" algorithm where the AIC is replaced by the prediction sum of squared criterion (PSS), while Yoshimura et al. [74] suggested its replacement by the residual sum squared criterion (RSS). Testing the revised GMDH over linear regression methods

it was proved that revised GMDH with RSS produced better results, while in revised GMDH with either AIC or PSS, a number of inputs were eliminated despite their importance and high correlation to the output. The revised GMDH has the advantage of using flexible partial descriptions but is also computational expensive due to the large number of candidate models for every pair of variables and the stepwise regression analysis for parameter identification. In [75] a stepwise regression recursive algorithm is used for parameter estimation reducing the computational load. Furthermore, the weighting error sum of squared criterions in forecasting horizon is used as a selection criterion, which can identify a good property in each layer in the whole forecasting field and not only in the case that the number of layers is prespecified as AIC and PESS do. The revised GMDH algorithm has been compared to neural networks for the prediction of the magnitude of voltage of harmonics but failed to produce better result [76]. On the other hand, in comparison to statistical methods and an optimal error predictor (OEP) over long-range and short-range predictions, it was proved that GMDH provides better results in short range prediction only. However, it does not show features of robustness to the change of prediction lead times where OEP is superior [77].

Kondo [78] presents a modification of the revised GMDH with respect to the type of partial descriptions generators. Two linear and non-linear generators are applied in the place of every processing unit. Second order polynomials are replaced by third order polynomials and in one case logistic function, which coefficients are estimated by logistic regression analysis. The new revised GMDH algorithms is investigated further by applying two different learning algorithms to update the weights between the neurons of the GMDH type NN [79]. In the first algorithm logistic and multivariate regression analysis are used to update the weights in accordance to the type of partial generator. Both old (i.e. used to find the structure of network) and new (i.e. obtained after the determination of structure) data are used. In the second algorithm only new data are used with backpropagation algorithm to recalculate the coefficients between the neurons.

The algorithm has been tested against NN in a financial application proving that a 10-15% improvement was achieved [80]. Additionally, Kondo et al. [81] introduce the utilisation of a feedback loop in every layer that reduces the complexity of the final model. Using the feedback loop the output neurons are not only combined with each other to the next layer causing a rapidly increase in complexity but they are also combined with the input variables of the system so that complexity increases gradually as well as the accuracy. Additionally, the number of partial generators is increased to seven. In [82] Kondo et al. propose to extend the number of arguments of partial generators beyond two, proving that the accuracy in a medical recognition application have increased.

6. Modifications of GMDH

The objectiveness of GMDH algorithm and its satisfactory performance as a non-linear modelling approach has driven a number of researchers to investigate it further as well as test it in a broad spectrum of applications. Their work has pointed either in the direction of providing a theoretical background for the method or in modifications that address some of its weaknesses and problems. Ivakhnenko admits that the theoretic base of GMDH has not been fully explored but similar to Monte Carlo method its experimental results provide a satisfactory justification [10]. The weakness of a strong theoretic foundation have been criticised by Van Zyl et al. [83] arguing that several aspects of the basic GMDH approach seems to be unconvincing and unfounded. One of the problems discussed in [83], is the justification of the functional form of the auxiliary variable in the balance of variables criterion, which is a better selection criterion in cases that the table of initial data contains information about several interrelated variables. The auxiliary function is a sum of system variables with given numerical weights. It was proved that the results depend on the form of auxiliary function and even for the same formula the results were different for minor changes in the coefficients. Additionally, there is a lack in the rationale behind the choice of the linear form for that variable.

Despite the proof of the statement made by Ivakhnenko about the existence of a minimum in external criteria as the complexity increases, Van Zyl points out that there is no satisfactory justification in the question: 'why this prediction to be more reliable than the others". Therefore the theoretical basis should be developed before the empirical data are introduced into the modelling procedure. Ivakhnenko's claim that short sequence of data and the 'environment" (i.e. input arguments – activation function) are enough is unfounded without its theoretic justification. The correct selection of the input vector, the complexity of final model, its consistency, the reduction of computational load are some of the aspects in the basic GMDH that have been investigated in order to improve its accuracy

and broaden its horizon of applications. Most of the proposed solutions have been designed and tested for the multilayer GMDH algorithms but their application can be also extended to the combinatorial algorithms.

Selection of Input Arguments

One of the main features of GMDH is its ability to objectively select the most appropriate input arguments amongst a set of candidates. However, the identification of these candidate input arguments is not straightforward and may effect its performance [84]. A large input vector will increase computation time while the inclusion of redundant variables may reduce accuracy. The elements of input vectors will be decided by the researchers and they should be as relevant as possible. The degree of relevance to the target can be estimated by the correlation analysis [85]. The cross and auto correlation function will not only reduce the input dimension but also indicate the number of lagging arguments that should be included in the modelling procedure [86]. Krotov and Kozubovskiy [87] accept the importance of cross correlation function pointing out that not only the Pearson correlation coefficient that presuppose normal distribution of deviations but the rank correlation should be also used [88]. Karnazes et al. [89] propose the selection of input arguments to take place according to a clusterization procedure. An algorithm will sequentially examine a set of input data vectors and grows clusters in a n-dimensional space. The final clusterization will detect operating regions in the independent spaces, as well as noisy and non-stationary variables. Small clusters that lies outside of large clusters will be attributed either to noise or non-stationary conditions and therefore will be eliminated.

Ryoubou et al. [90] propose the introduction of Interpretative Structural Modelling (ISM) theory into the GMDH algorithm in order to control the selection of input variables. The authors do not follow the idea of considering the system as a black box concept and claim that the theory, experience and intuition behind the system should be taken into account. Lemke and Muller [64] propose the application of objective systems analysis to reduce the variable set in cases where the model is in form of systems of equations. A parametric GMDH algorithm will be applied and a discrimination criterion will decide on which of these equations should be used in the system. In that way a nuclei of the dynamic variables of the system will be established. The variables outside the nuclei will be modelled individually by parametric GMDH algorithms. The above procedure has been also known as the OSA GMDH algorithm.

Generally, the identification problem can be subdivided into two other problems: the identification of structure and the estimation of parameters. Bastian and Gasos in [91] examine the identification of structure as two separate problems. The identification of input variables of the model (i.e. type I) and the identification of the input-output relations (i.e. type II). GMDH algorithms can be applied for the solution of type II problem as well as the parameter estimation. For the solution of type I problem, a MLP network is suggested. All the inputs are fed to the network and the regularity criterion is calculated. Then one of the input is replaced by a random signal and if the new value of regularity criterion remains small the excluded variable is considered redundant and deleted. However, such procedure is computationally demanding since every time the efficiency of one input argument is checked the network should be retrained. In [92] an initial single training is proposed but the algorithm may not always detect the redundant variables.

One of the features of inductive approaches is that the structural and parameter estimations are not happening separately. Van Welden et al. [93] argue that such a procedure may cause problems in cases of ill-defined tasks and it is better to distinguish these two operations. The "system approach problem solver" (SAPS) is used to identify those variables that are closer to the output and then GMDH can be applied to obtain the optimal model. The computation time is reduced with SAPS being a preprocessor to find the structural relationships and GMDH serving the purpose of estimating the parameters and obtaining a more refined structure.

Ikeda et al. [94] adopt a different approach for the selection of input vector. The input variables N are chosen from a wide set of variables N_0 according to the value of the mean square $\overline{(y - z_i)}^2$ where, y the output and $z_i = a_{i0} + a_{i1}x_i + ... + a_{in}x_i^n$ with x the input variable and n the order of a polynomial chosen subjectively by experts. The proposed modification attempts to eliminate the weakness of linear correlation coefficient to identify non linear relations however, Billings and Zhu [95] introduce higher order correlation functions that are able to compensate for non-linear relationships.

Inaccuracies in Parameter Estimation

The method of least square estimates is the most popular method to calculate the coefficients of partial descriptions. If the data matrix is well defined its estimates will be accurate however, in the majority of real world systems the data matrix is ill-defined and the least squares biased. Yaremenko [96] proposed the utilisation of orthogonal Chabyshev polynomials to overcome it. Tumanov [97] pointed out that 'GMDH algorithm with mutually orthogonalised partial descriptions ensures stability of the model structure as well as the estimates of its coefficient during the identification procedure of complex objects with changing dependent variables". The proposed algorithm allows the utilisation of dependent variables in the set of experimental data, obtains estimates of all model coefficients and perform adaptation of the model coefficients by refining their estimates over the entire sequence of experimental points. Duffy et al. [85] propose the utilisation of stepwise multiple regression techniques as well as the re-estimation of all the terms in the final equation using both training and testing set of data.

The reasons of inadequacy in least square estimates are explained by Sarychev in [98]. According to that the problem is based on the false assumption that the distribution of the error vector is normal and the author argues that the assumption of a binary exponential distribution is more suitable. This argument is justified by the different nature of the error distributions in different selected intermediate variables, the simple structure of the partial descriptions in the first layer with respect to the true model and the difference of the contribution of the individual descriptions in the previous layer to the total change in the output of the current layer. The method of least moduli is introduced where noise immunity is increased producing more stable and unbiased coefficients. According to that, the type of partial descriptions is simpler and a direct monitoring of the complexity and composition of the terms is used for describing the intermediate models. Ivanchenko et al. [99] propose the iterative least square method as a more robust approach, which increase the noise immunity of GMDH algorithms. The ability of least squares and instrumental variables method to produce unbiased estimates is tested by Ivakhnenko and Zholnarskiyi [18]. They conclude that in case of noise free data or normally distributed noisy data both of them produce good results but instrumental variable method is superior in different conditions.

Multicollinearity

Another problem found exclusively in multilayer algorithm, which effects the stability of coefficients, is that of multicollinearity. The selected variables in one layer may be highly correlated to those selected in previous layers, which will result to the appearance of multilayerness error. Duffy and Franklin [85] attempt to solve the problem by applying a stepwise multiple regression technique for the formulation of partial descriptions in place of least squares. Svetal'skiy et al. [100] propose instead of transferring in the next layer, the outputs of the selected partial descriptions to use the input arguments that generate these outputs. Mamedov [101] follows a similar approach and describe constraints to reduce complexity. The degree of polynomial and number of parameters in the approximating models should not exceed a certain natural number controlling in that way its complexity. Additionally the testing regions are narrowed so that the probability of finding a true model is increased.

The ridge regression analysis is another effective approach for stabilising the coefficients of models and solves the multicollinearity phenomenon [102]. A similar approach is followed in [103] where the ridge bias parameter is determined according to the data and the algorithm is called "adaptive GMDH". The optimum value of the ridge bias parameter is estimated by a special formula known as "the generalised ridge estimator of the adaptive GMDH" but precautions should be taken to avoid the bias phenomenon. Jirina [104] claims that the multilayerness error is caused by the polynomial form of partial descriptions, which increases the error as the complexity increases. Sigmoid functions are proposed for every generator of partial descriptions, which minimum and maximum value will be assigned with respect to the spread of the output values. Ikeda et al. [105] in order to stabilise the structure of the prediction model introduce an adaptive mechanism for estimating the coefficients of the model every time new data are added while the structure of model is preserved. An additional advantage of this procedure is the significant reduction of computation time.

Reduction of Complexity

Ivakhnenko [4] claims that if the number of selected models in every layer is as large as possible the optimum model will be never lost. On the other hand, following that procedure the complexity of the model as well as its computation time is increased. Triseyev [106] reduces the complexity of GMDH

algorithms by following a different approach for the selection of intermediate variables, which is based on the diversity of variables criterion and the structural number of partial descriptions. An intermediate variable will be selected as optimum in layer K only if its performance on the diversity criterion of variables is improved in the K+1 layer. In the final layer it is possible to have a set of optimal models so that the structural number will select the single model. In case of multivalueness for the structural number the minimum selection criterion will decide the optimum. The structural number is a binary vector, which indicates the presence or absence of characteristic input variables in the partial descriptions. Precautions should be considered for large input dimensions since it may become computationally expensive. The applications of lower order partial descriptions or the combination of low and high order equations can reduce the complexity. Parker et al. [69] in order to avoid an increase of the model order use second order polynomial in the first layer but only linear forms in subsequent layers. A pruning type technique is adopted by Dolenko et al. [107] where the final model is checked and insignificant terms are reduced. Additionally, the requirement of normalisation of data priori to GMDH is highlighted since unnormalised data were proved to produce worst results as well as computational unstable models.

Multiplicative-Additive GMDH Algorithm

It was mentioned above that the form of partial descriptions might affect the model complexity. The choice of partial descriptions is closely related to the field of applications [108]. The different types of descriptions and different complexing methods have been driven in a wide range of GMDH algorithms. Generally, partial descriptions of parametric polynomial models can be divided into four main categories according to the combination of their terms. Additive where new terms are added to the partial descriptions, multiplicative of unit power of factor, generalised multiplicative-additive and descriptions where the power of factors is replaced by a number p which can be either prespecified or not [109]. Ivakhnenko et al. [110] classify the GMDH algorithms with respect to reference function of the system, proposing the development of an algorithm that is not based on the sum of polynomials (e.g. combinatorial) but on the multiplication of polynomials of one variable. These algorithms are characterised as multiplicative-additive GMDH since use both summation and multiplication of partial descriptions. Multiplicative-additive algorithms have expanded the functional space of GMDH algorithms but face the problem of unstable coefficients due to least square method. Least squares method have been applied to the logarithms of their values after taking the powers but the minimum mean square is not satisfied [111]. The introduction of multiplicative-additive algorithm with optimum power factors has eliminated that problem [109]. Krotov and Kostenko in order to improve its accuracy propose the expansion of input arguments by using fractional transformation of the original variable [112]. The expediency of this modification will be decided by the decrease in the minimum of external criterion. The algorithm include two steps where in the first step the powers of polynomials are found by Combi GMDH and at the second step a complete sorting of the related models takes place. The algorithm has been tested successfully in the prediction of ecological systems [87].

Formulas of Partial Descriptions

Despite the wide range of partial descriptions the majority of the researchers follows the argument that Volterra series are capable of identifying any non-linear system and therefore have adopted polynomial partial descriptions similar to Ivakhnenko polynomial [7]-[9]. However, due to the complexity of the model and the requirement of including the theory behind the object, many modifications have been designed in order to adapt to system's properties. Duffy et al. [85] in order to increase the spectrum of partial descriptions in every layer, introduce the linear combination of all input variables as an additional partial description to the second order polynomial. Ikeda et al. [94] proposed the introduction of each input variable into a polynomial prior to their application in the partial generators in the expense of increasing complexity. Additionally, the squared terms in the second order polynomial are eliminated. The elimination of squared terms is also proposed in [103]. Similarly, Hara et al. [113] propose a linear description of two inputs (i.e. constant term is not included) and then its output will be fed to a second order polynomial. In that way, less than six data are required for the calculation of its parameters however, it is pointed out that Ivakhnenko polynomial has a high degree of freedom and is superior in fitting the input-output relations. In [84] an algorithm developed by Ihara is presented where partial descriptions are replaced by a filter function and the second order polynomial increasing the stability of long term predictions. Sarychev [98] proposes the exclusion of not only the square terms but also that the product of the two variables creating partial terms with only two unknown parameters. Ivakhnenko et al. [18] propose new short-term partial descriptions with quadratic term, which need a small number of observations (i.e. three to four). However in case of noisy data it is

recommended to use five to ten times this number of points. Park et al. [39] propose a wide range of partial descriptions like, linear, quadratic, cubic, bilinear, bicubic, trilinear and tricubic.

Overfitting

A consequence of complexity is the overfitting problem and poor generalisation. The partition of the data into two subsamples and the selection of the optimum model according to its accuracy on an unknown set of data may ensure the good generalisation. However, the large number of parameters in the final model could create overfitting problems and therefore techniques, which eliminate the number of parameters, should be adopted. Mehra in [114] adopts the application of stepwise regression method for parameter estimation that is capable of eliminating the multicollinearity problem as well. Additionally, the Stern estimator is proposed for parameter identification but it is based on Akaike' s Information Criterion. Hara et al. [115] propose the replacement of least squares by a BDS estimator, which is similar to the weighted least squares (WLS) estimator. The BDS estimator needs more computation time and therefore the WLS is finally preferred. Due to high computational load of both BDS and WLS it is proposed to apply the least squares in first place and use a number of statistics (e.g. chi-square tests, run tests, variance values of data, etc.) to guess the possibility of overfitting so that switch to BDS only if it is high. Based on experiments the overfitting was not totally reduced but the results were better [116].

Partition of Data

The objectiveness of GMDH algorithm is based on the utilisation of an external criterion to select the optimum model, which require the partition of the data. The subsamples should cover the operating regions of the system and have similar properties in order to avoid poor generalisation. The requirement of splitting data into two groups will lead to different models for different subsamples and researchers have investigated a number of techniques to overcome it [97]. A simple technique will include the most recent observations on the checking set with the rest data being in the training set. Another technique may involve the variance of the data, where a mix of low and high variance data will be included in both subsamples. Duffy et al. in [85] propose two different approaches, which ensure a proper distribution of the data in both sets. The first suggest a fixed selection of pattern such as putting alternative points in time in the training and testing set. The second ensures a better spread of data and based on a random function which binary output (0–1) will indicate the data used in the training and checking set. Muller [49] and Yurachkovskiy [84] introduce an alternative approach, which is appropriate in ill-defined objects with limited number of observations. According to that all the data are used in the training and checking set following the principle of cross-validation and the averaged regularity criterion will decide on the optimum model. Jirina [104] based on the conclusions driven in [70] proposes the utilisation of PSE criterion that can be expressed as a function of the trained squared error. In that case the iterative procedure will take place until the minimum of PSE is found. Ivakhnenko in [117] claims that the subsmaples of data should be as different as possible since that ensures the deepest minimum for the external criterion. The superposition of noise in the initial data is introduced to change the nature of data and increase the noise immunity of model.

Low Accuracy in GMDH Method

All these modifications have been tested and proved to improve the accuracy of GMDH. However, in many cases and particularly in applications of long range prediction the GMDH was inaccurate. Ivakhnenko in [118] recognising this failure of GMDH, summarises its causes in the existence of a short delta form correlation between output and predictors, the insufficient functional variety of the model candidates, the immoderate use of a sequence of external criteria for choosing the optimal complexity and the overcomplication of individual models. In addition, GMDH has been primary developed for the solution of small and modest problems, which is not the case for real world systems. The application of correlation analysis prior to GMDH algorithm as well as the development of a combined criterion in the place of external criterion could solve these problems and therefore improve accuracy.

Molnar in [119] justifies the low accuracy to the loss of correlation information between elementary polynomial during the estimation of coefficients. The re-estimation of the optimum parameters using the whole amount of data and the utilisation of Akaike's selection criterion to remove unnecessary coefficients is proposed. This method is also known as "extended Kalman filtering" which basic idea is similar to Newton's theory where small adjustments to the coefficients are made until convergence is obtained. Another cause of low accuracy is the possibility of eliminating important variables during the sorting out procedure. GMDH is geared to minimise the mean square error of the resulting model so it

takes into account average tendencies only. Any variable, which causes the function values to fall out that average tendency will be characterised as noise and therefore eliminated despite its importance. Styblinski et al. [120] propose the combination of a theoretic model and an inductive modelling approach as a proper solution to the above problem. A single layer interpolation technique known as "maximally flat quadratic interpolation" is proposed to produce a physical model for a circuit performance system. However, due to the high complexity of MFQI models a combination of MFQI and GMDH is finally chosen. According to that the input parameters and the output of physical models are fed in a GMDH algorithm where GMDH will select the best input variables and correct the physical model. In the application of GMDH algorithm the variables are trasfmormed by a second order polynomial or a bilinear logistic function prior to their introduction in the partial generators in the form of a rational function is used as partial generators [108].

GMDH Algorithm for Discrete Process

The majority of GMDH algorithms has been developed for continuous variables and cannot be applied to binary or discrete problems. A rebinarization technique, which will be used for the transition from binary to continuous attributes with the subsequent use of well known GMDH algorithms, is a potential solution. Ivakhnenko et al. in [121] introduce such an algorithm, which reconstructs with a sufficient precision an unknown harmonic function that is represented by a binary code. The sliding control criterion is applied to improve the parameter estimation since least square estimations sometimes provide imprecise estimates. An extensive review of such algorithms can be found in [122]. The harmonic rebinarization or rediscretization algorithm can be also applied to discrete pattern recognition problems allowing the application of parametric GMDH algorithms to find the optimum space of features, the structure of a decision rule and estimate its coefficients [45].

Model's Validation

A very important subject in every modelling procedure is that of model validation. It is significant to ensure that the selected model is adequate to reflect the causal relationships between input-output. Muller in [49] proposes the computation of the model with and without randomisation as one of the solutions to that problem. On the other hand, Krotov et al. [87] present a number of criteria, which can prove the verification of the forecast. The correlation coefficient, the mean square error S of the forecast and the mean squared deviation σ of the predicted process from the mean value of the entire series of observations (norm) can be used. In that case the reliability of the model could be characterised by the ratio S/σ.

Problems with Combi-GMDH Algorithm

Combinatorial GMDH algorithms in addition to the above drawbacks appear to have a high sensitivity to computation time. Muller et al. [123] propose two ways of speeding the sorting out procedure. The first one suggest the utilisation of information array in the formulae of $W^T W$ instead of the data sampling area $W=(XY)$ (i.e. X is the input vector and Y the output). The second propose the calculation of parameters with a recursive method of "framing". Ivakhnenko et al. [46] propose an alternative approach for the reduction of its computation time. According to that it is proposed to range variables with respect to a criterion during the complete sorting of models after a number of layers. In that way only F best variables will continue in the sorting out procedure until the minimum of criterion is achieved.

Ivakhnenko et al. [124] claim that Combi algorithm fails to produce the optimal non-physical model when the data are exact. The reason is the inability of external criterion to produce a single minimum. The criterion will produce an "uncertainty zone" where the optimal physical model corresponds to its midpoint. In that case a simplified non-physical model can be only found by searching for a model according to some external or internal criteria in that flat uncertainty zone. A threshold GMDH algorithm can solve the problem of exponential increase in computation time where an auxiliary criterion estimates the effectiveness of input variables. The effectiveness of continuous variables is estimated according to the absolute value of the correlation coefficient of the output variable and the feature to be estimated. On the other hand in the basic GMDH algorithm, the number of variables over which the external criterion search is performed is either constant (for exhaustive search) or reduced when adding new terms at the expense of eliminating non-effective inputs.

Genetic Algorithms in GMDH

Recent development in some of the GMDH aspects has involved the concept of Genetic Algorithms (GA). Robinson [125] points out that disadvantages of GMDH are its fixed structure and the

deterministic nature of the search for the best model. She uses MOGA optimisation algorithm to search the space of possible polynomials in order to optimise the performance of GMDH. The Ivakhnenko polynomial is replaced by a full four order polynomial and GA is used to identify the optimal partial description. The above modification have been characterised as Term Optimisation of GMDH (TOGMDH) since it only finds the optimum terms in partial descriptions and does not alter its structure. Robinson also proposed the Structure Optimisation of GMDH (SOGMDH), which optimises both the model's terms and the structure of the final model. In that algorithm, MOGA optimisation algorithm performs a wider stochastic search over a large range of possible models. SOGMDH uses the form of partial descriptions in TOGMDH but allows the evolution of the model in more than one layer allowing in that way the combination of two different partial descriptions in a latter stage. Both algorithms have been tested in regression and classification tasks where SOGMDH has shown a remarkable increase in accuracy.

7. Applications of GMDH Algorithms

The inability of deductive algorithms (neural networks, statistical models) to perform highly accurate identification and forecasting tasks has urged the development of other methods, which require less a priori information. Inductive sorting out procedure requires a small volume of a priori information for the system producing non physical models, which can approximate the ill-defined systems in a simpler and more accurate way than physical model do. GMDH algorithms as representative of inductive procedure have been modified in order to improve their performance and adapt on the peculiarities of every different object. Ivahnenko et al. [71] provide a long list of problems where GMDH has been successfully applied. Modifications in GMDH, which ensure the solution of some problems in these fields of applications, are also introduced.

Financial Systems
One of the preliminary areas of research where GMDH have been tested is the identification and forecasting task of financial systems. Macroeconomic systems have a limited number of observations and a dynamic behaviour, which cannot be identified by conventional methods. Scott et al. [126] claim that these methods fail, since they are not able to propose an objective technique for the selection of input argument, these systems are extremely non linear and methods are unable to take into account the minor interactions of their variables. GMDH on the other hand seems to address these problems and its performance is better in comparison to known economical systems. [96], [126], [127]. Ivakhnenko and Kostenko design a GMDH algorithm for long range predictions of the British economy. The principles of objective systems analysis have been applied to identify the optimum systems of equations. The algorithm is a two-level performing both annual and seasonal predictions. The number of equations in the system is defined by a discrimination criterion known as the criterion of accuracy of stepwise prediction. The criterion should be calculated on a different set of data because when exhaustive search is performed (i.e. combinatorial algorithm), it is transformed to an internal criterion selecting the more complex model [128], [129].

Brusilovskiy et al. [130] classify the prediction of economic indices into two categories. In the first group there is a partial description where information regarding only the dynamics of the predicted indices is used. Secondly, in the system prediction the dynamics of other indices, which interact with the predicted one may also be used but could be poor due to the incompleteness of input vector. Goleusov and Kondrasheva in [131] investigate the ability of GMDH to extract possible information about the interdependencies between the financial indices of the countries and therefore their economical systems. They conclude that GMDH is more than a simple modelling method and expands the possibilities of contextual interpretation of the results of economic modelling. Ivakhnenko et al. [132] apply a combination of two parametric GMDH algorithms in modelling macroeconomic models. Firstly OSA algorithm is developed to create nuclei of the system and then the Combi algorithm is applied to those variables outside the nuclei.

However, based on Nalimov's statement that models with a more fuzzy language should describe system with fuzzy characteristics, the utilisation of non-parametric GMDH algorithms is very common [133]. GMDH algorithm as a modelling and prediction approach can be also successively included in the development of integrated portfolio trading system. Lemke and Muller [64] in a first stage apply parametric and non parametric GMDH algorithms to the prediction of a portfolio of shares while in the

second stage they design a control process, which involves the transformation of the prediction into trading signals. Water et al. [134] are also applying the GMDH algorithm in the prediction of stock prices. F-statistics have been used as selection criteria but their sensitivity to the partition of the data sample effects their reliability. Their combination with the training score, which represents the number of times that the GMDH network correctly predicts the movements of the stock price in the training interval, is proposed as a potential and more effective selection approach.

Ecological Processes
Another area of modelling where GMDH algorithms have been popular is ecological processes. Generally, most of these processes are highly non linear with noisy data and statistical models fail to produce accurate results. Dolgopolov [135] proposes a combined method of the principle of inductive self-organisation and physical laws for the mathematical formulation of general dynamic models describing the evolution of the magnetic field of active solar regions. The optimal structure of the model is determined by the GMDH on the basis of a complete description in the form of a system of differential equations (i.e. the analogue of differential equations in GMDH is difference equations) of magnetohydrodynamics and a source function which takes into account forces of a potential nature. The concept of predicting the solar activity is discussed in [60] where both harmonic GMDH and the method of analogue complexing are discussed. The method of analogue complexing as a more robust approach is producing best results. Due to the long history a single analogue is found to produce more accurate results in comparison to the combination of analogues.

Krotov et al. [87] apply the multiplicattive-additive GMDH algorithm for the prediction of tree-growth rings while Valenca and Ludermir [47] compare the Box-Jenkins approach with neural networks with active neurons for the forecast of daily river flows. It was proved that NN with active neurons rise up the accuracy of short-term forecast and hence increase the lead-time of step by step long term forecast. Chang et al. [136] are dealing with another important ecological problem, the flood forecasting. A stepwise regression similar to the one introduced by Tamura and Kondo in [72] is proposed to tackle the multicollinearity problem and the algorithm is known as SGMDH. Additionally, a new recursive algorithm introduced by Catlin [137], which reduces computation time and estimates the coefficients at each step in an adaptive way is proposed for real time forecasting. The latter algorithm is known as RESGMDH and perform better in cases where the rainfall histograms between calibration event (training data) and verification event (testing data) are significantly different.

Control Applications
The inductive sorting out approach has been successfully applied in parts of control applications also. Kozubovskiy and Kupriyanov [138] apply the multilayer GMDH to calculate the pressure difference at the output of a differential pneumatic bridge. The bridge is part of a general pneumatic system, which controls the air environment inside a mine. Furthermore, GMDH can be applied for the calculation of optimum values of variables of Quasi-open loop systems. Many macroeconomics systems have long term feedback effects so researchers are forced to perform open-loop control looking at the optimal decision for each given year or quarter of economical planning. The available data can be divided into three subsets, the input manipulated variables, which optimal values should be found, the external input variables and the output values, which optimal values are known. The GMDH algorithm involves three steps where we have the calculation of deflection between the optimal reference value of the output and its real value, the development of Combi GMDH to obtain model of constraints and the solution of the systems of equations obtained by Combi. Ivakhnenko et al. [139]-[140] claim that when the number of manipulated variables and output variables is equal instead of using simplex method or other non linear programming optimisation technique, the optimal values for the manipulated inputs can be solved by simply solving a system of linear equation provided by Combi GMDH algorithm. The above concept is extended in [141]-[142] to cover the cases of dynamic objects, also. The proposed technique is simplified by avoiding complex non-linear programming methods and uses GMDH algorithms for the computation of the parameters of the equation for the constraints.

GMDH in Diagnostic Tasks
Expert systems have been designed for the development of advisory systems, which assist in making decisions in technical diagnostics tasks. However according to Kus et al. [143] their performance is poor due to the difficulties in formalising in a mathematical language the experts' knowledge, the mathematical transformation is not as good as pure expert knowledge and particularly in diagnostic tasks the input information is limited. A computerised-aid decision system is proposed that will overcome those problems and will be used for real world systems of high complexity, incomplete data

information where the information of their behaviour is coming solely from observations and the transformation of human knowledge is difficult. The proposed diagnostic system uses cluster analysis to accumulate and represents the knowledge as well as parametric GMDH algorithms, which will estimate the output of particular variables that play the role of a state identificator. The difference between the predicted value of the state identificators and their actual values will indicate an emergency state for a particular element of the investigated system [144]. The proposed system is expanded in [145] where the case of dynamical systems is considered. In that case the time t should be included in the partial descriptions of GMDH algorithm for continuous variables or past values of the target for discrete variables.

Other Applications
GMDH algorithms have been successfully designed in many more identification and forecasting process like the case of rare processes where combinatorial GMDH is successfully applied in connection to the average regularity criterion due to the limited number of observations [146]. GMDH algorithm has been also applied to the optimisation of physical model's structure and estimation of their parameters in agricultural systems [147]. Random processes in the form of a trend function and a remainder have been successfully modelled with GMDH algorithms [148]. The GMDH algorithms are used in both the trend and remainder identification task. Buttner et al. [149] apply the principle of self-organisation in the extremely complex process of predicting the rate of births in a country (i.e. fifty factors may effect the demographic process). Sarychev [150] developed an iterative GMDH algorithm for a number of processes, which models constitute a superposition of two dimensional beta distributions where the coefficients and factors are unknown. In this case the development of Combi GMDH is unacceptable due to the large amount of admissible combinations of the model parameters.

Due to the chaotic behaviour of most climate systems, Lin et al. [151] propose the combination of GMDH and chaotic theory as a more appropriate method for the prediction of the mean temperature in two cities in China. The new algorithm is known as Group Method of Phase Space Component have the same advantage as GMDH and enables the adjustment of the selected model by the appropriate choice of some of its parameters. GMDH algorithms have been also designed for the identification of bilinear models in [152]. It was proved to be computationally demanded but produce better results when the structure of bilinear model is unknown however, a proposed bilinear predictor was superior in cases where the model structure is known a priori. Iwasaki et al. [153] introduce an autonomous modelling strategy for the non-linear friction on the table motion. Using GMDH in place of NN, an optimal non-linear model is found without requiring the knowledge of the exact friction properties. Modelling the friction allow the achievement of higher motion performances since the table motion accuracy is effected by the friction. Xue and Watton [154] design a GMDH type NN to model fluid power systems. Due to unstable result from the second order polynomial for every pair of variables they applied it for every set of three variables. Based on their results it is clear that GMDH network performs well producing more accurate models.

Software Packages with GMDH Algorithms
The wide application of GMDH algorithms has urged a number of researchers to design software packages, which enable the development of one or more GMDH algorithms. Semenov and Malinovskaya [155] proposed the software package 'PARIS', which combined the Brandon's algorithm and GMDH theory. It is a multi-step procedure where Brandon's algorithm is firstly used to isolate the most significant repressors and Combi GMDH algorithm to find their optimum complexity polynomial. Then the residual component vector of the output variable is calculated and the next set of remaining repressors is identified. That procedure is repeated until all the input arguments have been used. The final model will be the sum of all these optimal polynomial. Tetko et al. [156] developed a software package suitable for Quantitative Structure Activity Relationships studies (QSAR). These tasks are characterised by limited observations and a large number of input variables with some of them irrelevant and highly correlated. Krotov et al. [157] describe the main parts of a computerised package for the modelling procedure of complex systems while Lemke in [158] designs a software package for the revised GMDH algorithm. An important feature of that package is the precautions taken by the designer in order to increase its accuracy. Firstly it is avoided to select only a number of intermediate variables in every layer and exclude other with the same value of selection criterion. Secondly in case of system of equations it is avoided to include variables in the equations at the same time t.

Muller and Lemke [133] propose the "KnowledgeMiner", which is a self-organising modelling software tool. It is a powerful and easy to use modelling tool that is designed to support the knowledge extraction process on a highly automated level [67]. It is able to implement advanced self-organising modelling techniques like multilayer GMDH, NN with active neurons and analogue complexing method [159]. The "KnowledgeMiner" has been successfully applied in financial systems for predictions tasks or decision systems like the design of a system responsible for the decision over the solvency of a company [67], [160].

8. External Criteria

The objectiveness of GMDH theory in the selection of the optimum model is based on the principle of external complement, according to which when a model's complexity increases then certain criteria, which hold the property of external complement go through their minimum [4]. Failure to achieve that may be caused by the existence of too noisy data, incomplete information in input arguments and improper choice of reference function. All these criteria that satisfy the above property are known as external in contrast to internal criteria, which have a monotonic behaviour choosing the model according to the degree of complexity following the property "the more complex a model is the more accurate will be" [1]. Generally, the modelling procedure may involve either exact or noisy data and the choice of criterion is based on the level of noise within the data. Ivakhnenko [32] points out that both external and internal criteria will choose the same model when the data are exact but their choice will not coincide if the level of noise increases. Internal criteria require supplementary information about the statistical characteristics of the data sample increasing the subjectiveness of the choice [148]. They can only choose physical models that can be used explicitly in short range predictions while external criteria are choosing non physical models that are less complex and more accurate in long range predictions.

A wide variety of external criteria exist and their choice depends on the type of the problem, the level of noise and the designer. Ivakhnenko [161] divides the external criteria into three categories: (a) exact criteria, (b) integral criteria like time-integral of exact data and (c) differential criteria. Long data sample can successfully apply exact criteria however, for short samples a more robust approach should be followed and differential criteria are preferable. Exact criteria are choosing models that are the most accurate in the given sample of data in contrast to differential criteria, which choice is based on the principle of consistency selecting a consistent model that does not depend on the given part of samples [162]. Obviously exact criteria will be appropriate to short range prediction while consistent criteria will provide models suitable for long range predictions.

Regularity criterion has been the main representative of exact criteria. Mehra [114] illustrates two of the main problems of regularity criterion and GMDH in general. Firstly, there are not clear rules for selecting the number of data points in each subsample and secondly the requirement for partition of data into two sets may cause problems in short sequences of data. Yurachkovskiy [84] in order to solve the above drawbacks proposes the averaged regularity criterion, which use all data in both training and checking set of data. On the other hand Goleusov et al. [131] present the symmetric regularity criterion to construct regular models which are slightly sensitive to small variations in the initial data. Sarychev in [163]-[164] produces a theoretic background for the regularity criterion and symmetric regularity criterion. It is proved that following the scheme of repeated observations for the partition of the data sample the regularity criterion as well as the symmetric regularity criterion hold the two basic properties that GMDH theory requires in order to be classified as external criteria. According to those, their mathematical expectation has a minimum that is attained for some small variances of the observation error on the actual structure and that shifts when the variance of the observation error increases in the region of simpler models.

Despite the ability of regularity criterion to satisfy the basic properties of external criteria it has the disadvantage of depending on the partition of the data. Sarychev [165]-[166] proves that indeed regularity criterion depends on partition of the data and therefore there is a problem in identifying the J optimal set of regressors in general and not for a specific set of data. A J optimal set of regressors is a group of regressors between a number of candidates, which in terms of contents and number correspond to the true model. It is proposed that the repeated scheme of observations may be the answer to the problem of partition of data and a new form of the regularity criterion seems to be able to identify it.

The problem of data partition have urged researchers to develop a more robust modelling approach where the choice of selection criterion will be as less as possible relevant to the given sample of data. The minimum bias or unbiasedness criterion was one of the first criteria following this robust approach [4], [114]. Aim of such criterion is the solution of one of the main problems in regression analysis, which is the choice of the most unbiased structure [167]. In case of multi-output systems or multi-level algorithms the balance of variables criterion is the equivalent to consistency criterion [118]. Ivakhnenko et al. [5] present different forms of consistency criteria according to the number of division of data space. However, in these types of criteria there is an imperfection of the mathematical form, which leads to a multivalueness problem. Consistency criteria may have more than one model with the same minimum value and the extraction of the true one amongst several fakes is necessary. In order to solve that problem the principle of regularisation should be applied. Ivakhnenko et al. in [5], [161] introduce two approaches that will identify the true model. Firstly it is proposed to sort the variants under the sum of a number of different consistency criteria applied in different partition of data sample. Alternatively, in cases of systems of equations the external equation will be calculated every time a new equation is added to the system and the extrapolation of the characteristics created by the minimum values of the criterion will indicate the true model. The extrapolation of that characteristic will indicate the true physical model.

Stepashko in [168] investigating the theoretic background of consistency criteria highlights four of their shortcomings. They tend to select simplified models, have a low selectivity due to their multimodal nature, are sensitive to the segmentation of data and finally face the multivalueness problem. The source of all these shortcomings is that consistency criteria do not consider the error of reproduction of the original data. According to Stepashko the problem of low selectivity could be solved by applying a new consistency criterion based on the partition of the data into three subsamples. This criterion will be able to look beyond the limits of estimation intervals and could find the best predicting model. However, the most appropriate solution to overcome all these shortcomings is the application of a combined criterion. Ivakhnenko et al. [51] claim that the problem of regularisation is effected by the level of noise within the data so that the more accurate the data are, the wider will be the ellipsoidal region of the locus of points of the minimum of a criterion which value is zero. In order to solve it, is proposed to the superposition of noise in the data sample. The noise will be random and with a large variance we shall be able to obtain a unique solution.

The properties of both regularity and consistency criteria have been investigated by Stepashko in [169] according to which as the variance of noise in the data increases consistent criteria tend to shift their minimum into simpler models proving in that way the applicability of Shannon's theorem in communication. However, in [170] where the asymptotic properties of six discriminating criteria are examined it is proved that in case of infinite number of data by plotting the graph complexity vs. regularity criterion its behaviour is monotonic without a clear minimum so it is proposed to interact them with precautions in case of very large set of data.

The limitation and problems of single criteria have indicated the introduction of combined criteria. The application of combined criterion will answer some problems of the individual criteria as well as combine both their properties. This procedure have been initial proposed by Tikhonov who claims that the choice of models should be done according to a differential (consistent) and accuracy criterion [51], [117]. Mehra in [114] proposes the utilisation of either regularity or unbiasedness criterion as auxiliary criteria first to indicate models that approaches to the optimum and then the application of balance of variables criterion as the final criterion to obtain the best discrimination between the candidate solutions. Additionally, he introduces the BIC criterion that is better than AIC for the improvement of GMDH reducing the number of parameters.

A number of different proposals has been made for the composition as well as the order of the subcriteria. Ivakhnenko [4] claims that their composition depends on the nature of the problem and the type of models. The application of unbiasedness plus regularity criterion is proposed for polynomial models in short range prediction without time variable. Harmonic and algebraic (polynomials) models for long term prediction with time variable should be identified by the unbiasedness plus the balance of variables criterion. Finally, for differential models (i.e. their analogue is finite difference models) in which long-term prediction is obtained by multistep integration, the unbiasedness plus stability of the prediction procedure criterion should be adopted. The choice of optimum model will depend on the indication of the non dominant solution in the plane of the two criteria. In case of long range predictions Ivakhnenko [32] is also proposing the combination of minimum bias criterion and the

criterion of stepwise prediction which is calculated on the entire interval of data. The advantage of the latter criterion is its ability to remove the bias of the parameter due to least square method.

Another important issue is that of the order with which the criteria should be used. Ivakhnenko [148] connect their order to the noise immunity of the individual criteria. The less noise-immune criterion should be used first and the more noise-immune should follow. The consistency criterion should be applied in the beginning but due to its low noise immunity [168] must be followed by an auxiliary criterion like the regularity criterion, the criterion of accuracy of prediction, the integral criterion or the symmetric regularity criteria. Additionally, the number of optimal models to be chosen by the first criterion and being subject to sorting by the auxiliary criterion is not clear. An alternative technique, which eliminates that problem allowing the simultaneous application of both criteria, is given in [13]. In that case the combined criterion will be a weighted sum of both criteria. However, the choice of appropriate value for weighted coefficient is not clear increasing in that way the subjectiveness of choice. In both cases there is the problem of the optimum choice for either the number of models or the weighted coefficients, which will also effect the final choice. Stepashko [168] in order to solve the problem of optimum number of models selected by the consistent criteria propose a simpler approach that includes the application of consistency criteria first by selecting all the model, which criteria value is smaller than ten times its minimum value and then apply to them a regularisation criterion.

Zaychenko et al. [40] propose the introduction of weight factors in front of the regularity criterion, which values should be within a range [0 1]. Different values of this factor are tested in different initial training data giving a permissible domain for the values of weight coefficient. The choice of regularity criterion for short range prediction is based on the assumption that good approximation in the past guarantees the good approximation in the immediate future. Chang et al. [136] propose the utilisation of more than two criteria in the combined criterion. The choice of optimum values corresponds to the minimum value of the objective function of three criteria. The objective function is a weighted sum of the three criteria with the weights to be allotted so that all of them have about the same level of influence (weight) in the final choice. A similar approach was followed by Robinson [125] where the MOGMDH is presented. According to that the Pareto front of all the selected models is found and all these models will be considered as optimum in the final layer. However, this procedure does not result in a final solution but rather a set of solutions, which final choice will be highly subjective.

In analogy to parametric modelling methods both external and internal criterion can be used in clusterization but internal criteria will select complex clusters without forming a minimum and therefore exclude the possibility of determining a unique clusterization of optimal complexity. Ivakhnenko [53] argues that in clusterization, accuracy and consistency are totally different things and the property of consistency is more important so the consistency criterion should be used in all cases. In analogy auxiliary criterion could be developed to solve the 'non-final' problem. The utilisation of balance of variables criterion is proposed when the data sample do not allow its partition into two subsamples [57]. Ivakhnenko et al. [162] claim that the problem of regularisation could be solved by the balance of discretization criterion. According to that, the clusterization is repeated for some intermediate number of levels choosing the clustering with the minimum value of the criterion. In case of applying the consistency criterion in clustering the regularisation problem can be overcome by 'inverting' certain dipoles and calculating the overall criterion of consistency.

The requirement of external criteria for data partition in GMDH restricts its application to underdetermined tasks of experimental system analysis. In that case the length of data is small and partition is forbidden. The concept of cross-validation could be viewed as a powerful tool to overcome such problem. The balance of variables and PESS criterion could be applied successively in that cases [171], [49].

Despite the application of exact and consistent criteria in the majority of the cases a number of alternative criteria have been designed by researchers. Most of them are suitable for particular applications taking into consideration special conditions of the individual tasks. Braverman et al. [172] propose a matrix condition criterion for the selection of best models in basic multilayer GMDH. It is proved that the threshold of the condition number of the normal equations is based not on the computer resolution but to the reliability of the original data. An increase in error variance lead to reduction of condition number while an increment of the output measurable interval leads to its increment. Ivakhnenko et al. [16] show that the characteristic 'selection criterion vs. complexity' for the multilayer GMDH algorithm with calculation of remainders has a monotonic behaviour and the 'rule

of left angle" will determine the optimal complexity. However, this choice will be always more subjective.

The application of different criteria in the modelling procedure will produce contradictory solutions and therefore experts should make the final decision. Silov et al. [173] design a linguistic criterion that takes into consideration that knowledge and its choice will be based on the maximum of the fuzzy measure of closeness to a hypothetical or utopian model. Belogurov [174] designs a criterion that classifies the model into two categories those that are accurate and good for forecasting and those that are not. This criterion is essential to objective systems analysis (OSA) algorithms or for the creation of nuclei of input variables. The criterion will provide a boundary that separates the models into two classes with respect to their informational base. Any model with a value larger than the predefined boundary should be excluded by further investigation. The value of boundary is based on the application itself and the level of required accuracy. However, a value of 1 is used in most cases. A list of similar criteria can be found in [1] and [170].

Hild et al. [175] divide the selection criteria into two main categories. In the first category all the criteria that require partition of the data are included while in the second the 'information criteria' that do not require such partition. Ivakhnenko [10] claims that strictly speaking any criterion that do not require partition of the data sample cannot be considered as external and in that case there is no GMDH method. Such criteria require either the assignment of some parameters increasing their subjectiveness or are too crude and therefore forbidden. The only exception in that case is for the underdetermined objects where the cross-validation concept would be applied. However, Hild and Bozdogan point out that such criteria eliminate the drawbacks of regularity criterion like its sensitivity to outliers and their ignorance of models complexity while taking into account parametric and model uncertainty. AIC selection criterion is the first such criterion selecting the optimum model among a number of competitors by using the information quantity as measure of goodness of fit. Nagasaka et al. [34] proved that both AIC and cross-validation criteria would be able to identify the optimum model. According to Hild and Bozdogan the failure of finding the optimum model may be caused by the lack of fit, the lack of parsimony or the profusion of complexity. AIC is capable of compensating for the first two only and proposes the other criteria that could compensate for all these reasons. Based on the information-theoretic statistical theory they introduce the ICOMP and ICOMP(IFIM) that take into account both parametric uncertainty as well as complexity, eliminating the need for division of data sample. Muller et al. [1] point out that many times variables with the necessary information may be rejected because of weak interrelations to the output as well as variables with repeated information being inserted into the models. Common criteria fail to identify such situations and therefore the 'Innovation-Contribution' criterion in order to select independent variables automatically should be considered.

Lange in [176] proves that GMDH multilayer algorithms force the problem of bias in the coefficients when the Residual Sum of Squares (RSS) is used as structure criterion. A new structural criterion that will compensate for the existence of bias is designed. The new criteria is known as 'Local Data Uncertainty Criterion' (LDUC) and is a combination of two different subcriteria Q_1 and Q_2 with the first corresponding to the precision and the second to bias. This criterion is able to describe the resistance of the evaluated output and the resistance of the evaluated coefficients against changes of data values at a fixed given data set. The value of Q_1 will be larger when irrelevant inputs are included and the value of Q_2 will be larger in case those essential variables are omitted. Finally, Dolenko et al. in [107] apply the "coefficient of multiple determination R" as the selection criterion in multilayer GMDH. Mueller et al. [133] present a comprehensive table of the most common criteria in the parametric GMDH algorithms. These criteria are distinguished into two groups according to their feature of partition of the data sample. Both categories can be found in tables 2-3.

Name	Criterion	Given
$s_{e,p}^2$	$\dfrac{N}{N-p}s_e^2$	
Final Prediction Error (FPE)	$\dfrac{N+p}{N-p}s_e^2$	
Predicted Square Error (PSE)	$s_e^2 + 2\sigma_p^2\dfrac{p}{N}$	σ_p^2 - est. of σ_e^2
Mallow C_p	$\dfrac{Ns_e^2}{\sigma_c^2} + 2p - N$	σ_c^2 - upper bound of σ_e^2
AIC	$N\ln s_e^2 + 2p + c$	
BIC	$N\ln s_{e,p}^2 + p\ln\left(\dfrac{s_{y^m}^2}{s_{e,p}^2}\dfrac{N}{p}\right)$	
Schwarz	$s_e^2 + s_y^2\ln N\dfrac{p}{N}$	
Rissanen	$N\ln s_e^2 + p\ln N$	
Kocerka	$\dfrac{2ps_e^2}{N-p} - Ns_{y^m}^2$	
Vapnik	$\dfrac{s_e^2}{\left[1-\lambda\sqrt{\dfrac{k\left(\ln\left(\dfrac{N}{k}\right)+1\right)-\ln\alpha}{N}\left(1-\dfrac{1}{4}\dfrac{\ln\left(\dfrac{N}{k}\right)+1-\ln\alpha}{N}\right)}\right]}$	α
$i^2(N)$	$\dfrac{\sum_{t=1}^{N}\left(\overline{y_t^m} - y_t\right)^2}{\sum_{t=1}^{N}y_t^2}$	
Weighting Error Sum of Squared	$MWSS = \sum_{i=1}^{N}\sum_{\lambda=1}^{N}W(\lambda)e_i^2(\lambda)$	
Where	N: number of observations; y_t^m: model output; y_t: value of observation; p=1; $s_e^2 = \dfrac{1}{N}\sum_{t=1}^{N}\left(y_t - y_t^m\right)^2 = \dfrac{1}{N}\sum_{t=1}^{N}e_t^2$; $s_y^2 = \dfrac{1}{N}\sum_{t=1}^{N}\left(y_t - \overline{y}\right)^2$; $s_{y^m}^2 = \dfrac{1}{N}\sum_{t=1}^{N}\left(y_t^m\right)^2$; $W(\lambda)$: the weight of λ-step ahead prediction whose form depends on the prediction property at each lead time; $e_i(\lambda)$: denotes the predictions error of λ-step ahead at time i	

Table 2: Criteria of selection without partition of observations

Criteria of Accuracy	Mathematical Formula
Unsymmetrical Regularity criterion	$\Delta^2(B) = \sum_{t \in N_B} \left(y_t - y_t^m(A) \right)^2$ $\Delta^2(A) = \sum_{t \in N_A} \left(y_t - y_t^m(B) \right)^2$
Symmetrical Regularity criterion	$d^2 = \Delta^2(A) + \Delta^2(B)$
Unsymmetrical Stability criterion	$x^2 = \sum_{t \in N_A \cup N_B} \left(y_t - y_t^m(A) \right)^2$
Symmetrical Stability criterion	$S^2 = \sum_{t \in N_A \cup N_B} \left[\left(y_t - y_t^m(A) \right)^2 + \left(y_t - y_t^m(B) \right)^2 \right]$
Criteria of consistency	
Unbiased criterion for coefficients	$n_{unb,1}^2 = \sum_{i=1}^{M} \left(\alpha_i^A - \alpha_i^B \right)^2$
Unbiased criterion for outputs	$n_{unb,2}^2 = \sum_{t \in N_A \cup N_B} \left(y_t^m(A) - y_t^m(B) \right)^2$
Absolutely stable criterion	$V^2 = \sum_{t \in N_A \cup N_B} \left(y_t^m(A \cup B) - y_t^m(A) \right) \left(y_t^m(B) - y_t^m(A \cup B) \right)$
Where	$y_t^m(A)$ $\left(y_t^m(B) \right)$ - model output estimated on N_A (N_B) α_i^A $\left(\alpha_i^B \right)$ - parameters of a model estimated on N_A (N_B)

Table 3: Criteria of selection using training and testing data sets (N_A and N_B)

9. Conclusions

GMDH theory has been introduced to the scientific community in late sixties but it was the last two decades that it was thoroughly explored and expanded. A huge amount of research is undertaken in Russia and Ukraine where new algorithms as well as their theoretic base are established. The wide range of applications have confirmed and strengthened its position as a non linear modelling technique. Despite the contradictory results between neural networks and GMDH in some cases, it is no exaggeration that GMDH can successfully replace neural networks in identification and forecasting tasks. Its ability to address most of neural networks drawbacks have established her as a successful rival in modelling procedure of real world systems.

Recent developments in GMDH theory like neural networks with active neurons and the introduction of the concept of evolution (GA) have improved its accuracy and successfully address some of its disadvantages. However, both of them have not been thoroughly explored and further investigation is demanded. Additionally, most of the research in the selection criterion has focused on single objective optimisation, which is rarely the case in real world systems. It will be more appropriate to say that the optimisation of structure should take place according to a number of conflict objects in a multidimensional space and therefore multiobjective optimisation techniques should be explored.

Acknowledgement

References:

[1] J.A. Muller, A.G. Ivachnenko and F. Lemke, 'GMDH algorithms for complex systems modelling",
Mathematical and Computer Modelling of Dynamical Systems, vol.4. no.4. pp.275-316, 1998.

[2] W.S. Sarle, "Neural networks and statistical models", in *Proceedings of the Annual SAS User Group International Conference,* Dallas, pp.1538-1549, 1994.

[3] H. Madala, "Comparison of inductive versus deductive learning networks", *Complex Systems,* vol.5, no.2, pp.239-258, 1991.

[4] A.G. Ivakhnenko, "The group method of data handling in prediction problems", *Soviet Automatic Control c/c of Avtomatika,* vol.9, no.6, pp.21-30, 1976.

[5] A.G. Inakhnenko and N.L. Ivakhnenko, "Self-organization of mathematical models for creating an artificial intelligence system", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.24-33, 1986.

[6] J.H. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor, University of Michigan Press, MI, 1975.

[7] A.G. Ivakhnenko, "The group method of data handling – a rival of the method of stochastic approximation", *Soviet Automatic Control c/c of Avtomatika,* vol.1, no.3, pp.43-55, 1968.

[8] J.A. Muller and A.G. Ivakhnenko, "Self-organizing modelling in analysis and prediction of stock market", in *Proceedings of the Second International Conference on Application of Fuzzy Systems and Soft Computing – ICAFS'96,* pp.491-500, Siegen, Germany, 1996.

[9] A.G. Ivakhnenko, "Heuristic self-organization in problems of engineering cybernetics", *Automatica,* vol.6, pp.207-219, 1970.

[10] A.G. Ivakhnenko, "Sorting methods for modelling and clusterization (survey of the GMDH papers for the years 1983-1988). The present stage of GMDH development", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.4, pp.1-13, 1988.

[11] A.G. Ivakhnenko, "Polynomial theory of complex systems", *IEEE Transactions on Systems, Man and Cybernetics,* vol.SMC-1, no.4, pp.364-378, 1971.

[12] H.R. Madala and A.G. Ivakhnenko, *"Inductive learning algorithms for complex systems modelling",* Boca Raton, Florida, USA, CRC Press, 1994.

[13] V.S. Stepashko and Yu.P. Yurachkovskiy, "The present state of the theory of the group method of data handling", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.4, pp.36-46, 1986.

[14] Yu.P. Yurachkovskiy, "Restoration of polynomial dependencies using self-organization", *Soviet Automatic Control c/c of Avtomatika,* vol.15, no.4, pp.17-21, 1982.

[15] A.G. Ivakhnenko and Yu.L. Kocherga, "Theory of two-level GMDH algorithms for long-range quantitative prediction", *Soviet Automatic Control c/c of Avtomatika,* vol.16, no.6, pp.7-12, 1983.

[16] A.G. Ivakhnenko, G.I. Krotov and V.S. Stepashko, "Harmonic and exponential harmonic GMDH algorithms. Part 2. Multilayer algorithms with and without calculation of remainders", *Soviet Automatic Control c/c of Avtomatika,* vol.16, no.1, pp.1-9, 1983.

[17] A.G. Ivakhnenko, G. Petrache and M.S. Krasyts'kyy, "A GMDH algorithm with random selection of pairs", *Soviet Automatic Control c/c of Avtomatika,* vol.5, no.4, pp.23-30, 1972.

[18] A.G. Ivakhnenko and A.A. Zholnarskiy, "Estimating the coefficients of polynomials in parametric GMDH algorithms by the improved instrumental variables method", *Journal of Automation and Information Sciences c/c of Avtomatika,* vol.25, no.3, pp.25-32, 1992.

[19] V.S. Stepashko, "GMDH algorithms as a basis for automating the process of modelling from empirical data", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.4, pp.42-52, 1988.

[20] V.D. Dimitrov, "A stochastic GMDH algorithm with successive introduction of pairs of variables", *Soviet Automatic Control c/c of Avtomatika,* vol.3, no.5, pp.61-63, 1970.

[21] S.G. Patereu and O.I. Sheluđ'ko, "A probabilistic GMDH algorithm with sequential discrimination of input features", *Soviet Automatic Control c/c of Avtomatika,* vol.6, no.3, pp.29-33, 1973.

[22] Y.P. Triseyev, "Modifications of GMDH algorithms for systemic forecasting", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.3, pp.35-38, 1987.

[23] A.G. Ivakhnenko, "On selecting a set of output variables and using the GMDH for passive and active design of an experiment", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.4, pp.94-95, 1988.

[24] V.S. Stepashko and Yu.V. Kostenko, "A GMDH algorithm for two-level modelling of multidimensional cyclic processes", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.4, pp.49-57, 1987.

[25] V.M. Vysotskiy, A.G. Ivakhnenko and V.I. Cheberkus, "Long-term prediction of oscillatory processes by finding a harmonic trend of optimum complexity by the balance-of-variables criterion", *Soviet Automatic Control c/c of Avtomatika,* vol.8, no.1, pp.18-24, 1975.

[26] A.P. Sarychev, "A multilayer nonlinear harmonic GMDH algorithm for self-organization of predicting models", *Soviet Automatic Control c/c of Avtomatika,* vol.17, no.4, pp.90-95, 1984.

[27] A.G. Ivakhnenko, G.I. Krotov and Yu.P. Yurachkovskiy, "An exponential-harmonic algorithm of the group method of data handling", *Soviet Automatic Control c/c of Avtomatika,* vol.14, no.2, pp.21-27, 1981.

[28] V.Y. Shelekhova, "Harmonic algorithm GMDH for large data volume", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.117-126, 1995.

[29] I. Hayashi and H. Tanaka, "The fuzzy GMDH algorithm by possibility models and its application", *Fuzzy Sets and Systems,* vol.36, no.2, pp.245-258, 1990.

[30] S. Matushita, A. Kuromiya, M. Yamaoka, T. Furuhashi and Y. Uchikawa, "A study on fizzy GMDH with comprehensible fuzzy rules", in *Proceedings of the 1994 IEEE Symposium on Emerging Technologies and Factory Automation – ETFA'94,* pp.192-198, 1994.

[31] K. Yokode, H. Tanaka and H. Ichibuchi, "Fuzzy if-then rules with certainty factors u sing multi-layer model of GMDH", *Japanese Journal of Fuzzy Theory and Systems,* vol.7, no.1, pp.47-63, 1995.

[32] A.G. Ivakhnenko, "New accents in the theory of self-organization of models", *Soviet Automatic Control c/c of Avtomatika,* vol.14, no.6, pp.43-53, 1981.

[33] H. Ichihashi, N. Harada and K. Nagasaka, "Selection of the optimum number of hidden layers in neurofuzzy GMDH", in *Proceedings of the 1995 IEEE International Conference on Fuzzy Systems,* vol.3, pp.1519-1526, 1995.

[34] K. Nagasaka, N. Harada, H. Ichihashi and R. Leonard, "Adaptive learning networks of multi-stage fuzzy production rules in expert system of grinding characteristics", *Computers & Industrial Engineering,* vol.27, no.1-4, pp.433-436, 1994.

[35] K. Nagasaka, H. Ichihashi and R. Leonard, "Neuro-fuzzy GMDH and its application to modelling grinding characteristics", *International Journal of Production Research,* vol.33, no.5, pp. 1229-1240, 1995.

[36] M. Brown and C. Harris, *Neurofuzzy Adaptive Modelling and Control,* Prentice Hall, New York, London, 1994.

[37] T. Ohtani, H. Ichihashi, T. Miyoshi and K. Nagasaka, 'Orthogonal and successive projection methods for the learning of neurofuzzy GMDH", *Information-Sciences,* vol.110, no.1-2, pp.5-24, 1998.

[38] T. Ohtani, H. Ichihashi, T. Miyoshi and K. Nagasaka, 'Structural learning with M-apoptosis in neurofuzzy GMDH", in *Proceedings of the 1998 IEEE International Conference on Fuzzy Systems,* vol.2, pp.1265-1270, 1998.

[39] H.S. Park, S.K. Oh, T.C. Ahn and W. Pedrycz, "A study on multi-layer fuzzy polynomial inference system based on extended GMDH algorithm", in *Proceedings of the 1999 IEEE International Conference on Fuzzy Systems - FUZZ-IEEE'99,* vol.1, pp.354-359, 1999.

[40] Y.P. Zaychenko, A.G. Kebkal and V.F. Krachkovskii, 'The fuzzy group method of data handling and its application to the problems of the macroeconomic indexes forecasting", available in URL address: *http://skyscraper.fortunecity.com/cray/793/articles/papers.html,* 2000.

[41] J.A. Muller, 'Self-organizing modelling present state and new problems", *Systems Analysis Modelling Simulation,* vol.18-19, pp.87-92, 1995.

[42] A.G. Ivakhnenko, G.A. Ivakhnenko and J.A. Muller, 'Self-organization of neural networks with active neurons", *Pattern Recognition and Image Analysis,* vol.4, no.2, pp.185-196, 1994.

[43] A.G. Ivakhnenko and J.A. Muller, 'Self-organization of nets of active neurons", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.93-106, 1995.

[44] A.G. Ivakhnenko, 'Self-organization of neuronet with active neurons for effects of nuclear tests explosions forecasting", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.107-116, 1995.

[45] A.G. Ivakhnenko and G.A. Ivakhnenko, "A comparison of discrete and continuous recognition systems", *Pattern Recognition and Image Analysis,* vol.6, no.3, pp.445-447, 1996.

[46] A.G. Ivakhnenko, D. Wunch and G.A. Ivakhnenko, 'Inductive sorting-out GMDH algorithms with polynomial complexity for active neurons of neural networks", in *Proceedings of the International Joint Conference on Neural Networks,* Piscataway, New Jersey, USA, IEEE, 1999.

[47] M. Valenca and T. Ludermir, 'Self-organizing modelling in forecasting daily river flows", in *Proceedings of the 5th Brazilian Symposium on Neural Networks,* pp.210-214, 1998.

[48] S. Beer, *Decision and Control: The Management of Operational Research and Management Cybernetics,* John Wiley & Sons, London, 1966.
[49] J.A. Muller, 'Self-organization of models – present state", available in URL address: *http://www.inf.kiev.ua/GMDH-home/articles/,* 1996.

[50] A.G. Ivakhnenko, G.A. Ivakhnenko and J.A. Muller, 'Self-organization of optimum physical clustering of a data sample for a weakened description and forecasting of fuzzy objects", *Pattern Recognition and Image Analysis,* vol.3, no.4, pp.417-422, 1993.

[51] A.G. Ivakhnenko and J.A. Muller, 'Problems of an objective computer clustering of a sample of observations", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.24, no.1, pp.54-62, 1991.

[52] A.G. Ivakhnenko, S.A. Petukhova and N.A. Ivakhnenko, 'Objective computerized clustering. Part 1. Theoretical questions", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.3, pp.1-9, 1986.

[53] A.G. Ivakhnenko, 'Objective clusterization on the basis of the theory of self-organization of models", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.5, pp.1-9, 1987.

[54] N.A. Ivakhnenko, L.P. Semina and T.A. Chikhradze, "A modified algorithm for objective clustering of data", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.9-18, 1986.

[55] N.A. Ivakhnenko, I. Lu, L.P. Semina and G.A. Ivakhnenko, "Objective computer clusterization. Part 2. Use of information about the goal function to reduce the amount of search", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.1, pp. 1-13, 1987.

[56] A.G. Ivachnenko and J.A. Muller, "Selection procedures and their application in economy and ecology", in *Proceedings of the 4th International Symposium on Systems Analysis and Simulation,* Berlin, Amsterdam, Elsevier, 1992.

[57] A.G. Ivakhnenko, Ye.N. Fateyeva and N.A. Ivakhnenko, "Nonparametric GMDH forecasting models. Part 1. Sorting the Bayes or Wald formulas", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.22, no.1, pp.1-8, 1989.

[58] A.G. Ivakhnenko, S.A. Petukhova, G.A. Ivakhnenko, V.M. Yudin and S.A. Kovbasyuk, "Objective selection of optimal clusterizing of a data sample during compensation of non-robust random interference", *Journal of Automation and Information Sciences c/c of Avtomatika,* vol.26, no.3, pp.45-58, 1993.

[59] A.G. Ivakhnenko and J.A. Muller, "Parametric and non parametric selection procedures in experimental systems analysis", *Systems Analysis Modelling Simulation,* vol.9, no.2, pp.157-175, 1992.

[60] A.G. Ivakhnenko, A.P. Sarychev, P.I. Zalevskiy and N.A. Ivakhnenko, "Experience of solving solar activity forecasting problems with precise and robust approaches", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.3, pp.31-42, 1988.

[61] A.G. Ivakhnenko, I.K. Timchenko and D.A. Ivakhnenko, "Nonparametric forecasting models. Part 4. Combining forecasts of multidimensional processes", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.23, no.1, pp.19-30, 1990.

[62] A.G. Ivakhnenko, "An inductive sorting method for the forecasting of multidimensional random processes and events, with the help of analogue forecast complexing", *Pattern Recognition and Image Analysis,* vol.1, no.1, pp.99-108, 1991.

[63] A.G. Ivakhnenko, V.A. Chainskaya and N.A. Ivakhnenko, "A non-parametric combinatorial GMDH algorithm for analogue search operators", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.23, no.5, pp.11-23, 1990.

[64] F. Lemke and J.A. Muller, "Self-organizing data mining for a portfolio trading system", *Journal of Computational Intelligence in Finance,* vol.5, no.3, pp.12-26, 1997.

[65] A.G. Ivakhnenko, N.N. Bogachenko and L.T. Min, "Model-free forecasting of random processes by complexing their analogues", *Pattern Recognition and Image Analysis,* vol.7, no.3, pp.309-314, 1997.

[66] J.A. Muller, "Self-organization of models – present state", in *Proceedings of the 1995 EUROSIM Conference. EUROSIM '95 Simulation Congress,* Amsterdam, Netherlands, Elsevier Science B.V., 1995.

[67] F. Lemke, "Knowledge extraction from data using self-organizing modelling technologies", *paper published on eSEAM'97 Conference, available in URL address: http://www.knowledgeminer.net/index3.htm,* 1997.

[68] D.T. Pham and X. Liu, "Modelling and prediction using GMDH networks of Adalines with nolinear preprocessors", *International Journal of Systems Science,* vol.25, no.11, pp.1743-1759, 1994.

[69] R.G.J. Parker and M. Tummala, "Identification of volterra systems with a polynomial neural network", in *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP'92,* vol.4, pp.561-564, 1992.

[70] S.J. Farlow, *Self-organizing methods in modelling. GMDH type algorithms,* New York and Basel, Marcel Dekker, Inc., 1984.

[71] A.G. Ivakhnenko and G.A. Ivakhnenko, "The review of problems solvable by algorithms of the group method of data handling (GMDH)", *Pattern Recognition and Image Analysis,* vol.5, no.4, pp.527-535, 1995.

[72] H. Tamura and T. Kondo, "Heuristics free group method of data handling algorithm of generating optimal partial polynomials with application to air pollution prediction", *International Journal of Systems Science,* vol.11, no.9, pp.1095-1011, 1980.

[73] Y. Sawaragi, T. Soeda, H. Tamura, T. Yoshimura, S. Ohe, Y. Chujo and H. Ishihara, "Statistical prediction of air pollution levels using non-physical models", *Automatica,* vol.15, no.4, pp.441-451, 1979.

[74] T. Yoshimura, U.S. Pandey, T. Takagi and T. Soeda, "Prediction of the peak flood using revised GMDH algorithms", *International Journal of Systems Science,* vol.13, no.5, pp.547-557, 1982.

[75] X.F. Wang and D. Liu, "A recursive algorithm for GMDH", *Systems Analysis Modelling Simulation,* vol.7, no.7, pp.533-542, 1990.

[76] H. Mori and S. Tsuzuki, "Comparison between backpropagation and revised GMDH techniques for predicting voltage harmonics", in *Proceedings of the 1990 IEEE International Symposium on Circuits and Systems,* vol.2, pp.1102-1105, 1990.

[77] A. Vicino, R. Tempo, R. Genesio and M. Milanese, "Optimal Error and GMDH predictors. A comparison with some statistical techniques", *International Journal of Forecasting,* vol.3, no.2, pp.313-328, 1987.

[78] T. Kondo, "The learning algorithm of the GMDH neural network and their application to the medical image recognition", in *Proceedings of the 37th SICE Annual Conference - SICE'98,* pp.1109-1114, 1998.

[79] T. Kondo, "GMDH neural network algorithm using the heuristic self-organization method and its application to the pattern identification problem", in *Proceedings of the 37th SICE Annual Conference - SICE'98,* pp.1143-1148, 1998.

[80] A.S. Pandya, T. Kondo, T.U. Shah and V.R. Gandhi, "Prediction of stock market characteristics using neural networks", in *Proceedings of the SPIE. The International Society for Optical Engineering,* vol.3722, pp.189-197, 1999.

[81] T. Kondo, A.S. Pandya and J.M. Zurada, "GMDH-type neural networks with a feedback loop and their application to nonlinear system identification", in *Proceedings of the ANNIE'99 Conference, International Session Paper,* St. Louis, Missouri, 1999.

[82] T. Kondo, A.S. Pandya and J.M. Zurada, "GMDH-type neural networks and their application to the medical image recognition of the lungs", in *Proceedings of the 38th SICE Annual Conference - SICE'99,* pp.1181-1186, 1999.

[83] J.G. Van Zyl and D.C.J. De Jongh, "Experiments in socio-economic forecasting using Ivakhnenko's approach", *Applied Mathematical Modelling,* vol.2, no.1, pp.49-56, 1978.

[84] Y.P. Yurachkovskiy, "Improved GMDH algorithms for process prediction", *Soviet Automatic Control c/c of Avtomatika,* vol.10, no.5, pp.61-71, 1977.

[85] J.J. Duffy and M.A. Franklin, "A learning identification algorithm and its application to an environmental system", *IEEE Transactions on Systems, Man and Cybernetics,* vol.SMC-5, no.2, pp.226-240, 1975.

[86] S.F. Kozubovskiy, "Determination of the optimal set of lagging arguments for a difference predicting model by correlation analysis", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.77-79, 1986.

[87] G.I. Krotov and S.F. Kozubovskiy, "Verification of dendroscale forecasting by a multiplicative GMDH algorithm", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.3, pp.1-7, 1987.

[88] M. Kendall and J.D. Gibbons, *Rank Correlation Methods,* Edward Arnold, London, 1990.

[89] P.A. Karnazes and R.D. Bonnell, "Systems identification techniques using the group method of data handling", in *Proceedings of the 6th Symposium on Identification and System Parameter Estimation,* vol.1, pp.713-718, Washington DC, Oxford, International Federation of Automatic Control, Pergamon, 1982.

[90] M. Ryoubu, S. Ikeda and Y. Sawaragi, "A prediction model for regional economic system by self-organization method", in *Proceedings of the International Conference on Cybernetics and Society,* pp.619-622, 1978.

[91] A. Bastian and J. Gasos, "A type I structure identification approach using feedforward neural networks", in *Proceedings of the 1994 IEEE International Conference on Neural Networks,* vol.5, pp.3256-3260, 1994.

[92] A. Bastian and J. Gasos, "Modeling using regularity criterion based constructed neural networks", *Computers & Industrial Engineering,* vol.27, no1-4, pp.441-444, 1994.

[93] D. Van Welden, G. Vasteenkiste and M. Abo-Elela, "Symbiosis of data analysis techniques for non-linear ill-defined systems", in *Proceedings of the 1991 Summer Computer Simulation Problem. Twenty-Third Annual Summer Computer Simulation Conference,* San Diego, Ca, USA, pp.32-38, 1991.

[94] S. Ikeda, S. Fujishige and Y. Sawaragi, "Non-linear prediction model of river flow by self-organization method", *International Journal of Systems Science,* vol.7, no.2, pp.165-176, 1976.

[95] S.A. Billings and Q.M. Zhu, "Nonlinear model validation using correlation tests", *International Journal of Control,* vol.60, no.6, pp.1107-1120, 1994.

[96] A.G. Yaremenko, "Synthesis of regression equation for gross product of South Carolina using GMDH algorithms", *Soviet Automatic Control c/c of Avtomatika,* vol.7, no.4, pp.70-73, 1974.

[97] N.V. Tumanov, "A GMDH algorithm with mutually orthogonal partial descriptions for synthesis of polynomial models of complex objects", *Soviet Automatic Control c/c of Avtomatika,* vol.11, no.3, pp.82-84, 1978.

[98] A.P. Sarychev, "Stable estimation of the coefficients in multilayer GMDH algorithms", *Soviet Automatic Control c/c of Avtomatika,* vol.17, no.5, pp.1-5, 1984.

[99] V.N. Ivanchenko, N.N. Lyabakh and A.N. Guda, "An algorithm of harmonic rebinarization of a data sample", *Journal of Automation and Information Sciences c/c of Avtomatika,* vol.25, no.3, pp.77-82, 1992.

[100] B.K. Svetal'skiy and P.I. Koval'chuk, "A multilayer algorithm of the group method of data handling with selection of primary arguments", *Soviet Automatic Control c/c of Avtomatika,* vol.12, no.4, pp.24-27, 1979.

[101] M.I. Mamedov, "An iterative polynomial GMDH algorithm with limitation of complexity of models being selected", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.1, pp.99-100, 1986.

[102] T. Nishikawa and S. Shimizu, "Identification and forecasting in management systems using the GMDH method", *Applied Mathematical Modelling,* vol.6, no.1, pp.7-15, 1982.

[103] T. Nishikawa and S. Shimizu, "The characteristics of a biased estimator applied to the adaptive GMDH", *Mathematical and Computer Modelling,* vol.17, no.1, pp.37-48, 1993.

[104] M. Jirina, "The modified GMDH sigmoidal and polynomial neural net", in *Proceedings of the Conference in System Identification - SYSID'94,* vol.2, pp.611-613, 1995.

[105] S. Ikeda, M. Ochiai and Y. Sawaragi, "Sequential GMDH algorithm and its application to river flow prediction", *IEEE Transactions on Systems, Man and Cybernetics,* vol.SMC-6, no.7, pp.473-479, 1976.

[106] Y.P. Triseyev, "GMDH algorithm with variable freedom of choice in selection layers based on criterion of diversity of variables", *Soviet Automatic Control c/c of Avtomatika,* vol.10, no.4, pp.30-33, 1977.

[107] S.A. Dolenko, Yu.V. Orlov and I.G. Persiantsev, "Practical implementation and use of group method of data handling (GMDH): prospects and problems", in *Proceedings of the 2nd International Conference on Adaptive Computing in Engineering Design and Control - ACEDC'96,* pp.291-293, PEDC, University of Plymouth, UK, 1996.

[108] M.A. Styblinski and S. Aftab, "Combination of interpolation and self-organizing approximation techniques – a new approach to circuit performance modeling", *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems,* vol.12, no.11, pp.1775-1785, 1993.

[109] A.G. Ivakhnenko and G.I. Krotov, "A multiplicative-additive non-linear GMDH with optimization of the power of factors", *Soviet Automatic Control c/c of Avtomatika,* vol.17, no.3, pp.10-13, 1984.

[110] A.G. Ivakhnenko, Ye.G. Deleur, A.G. Yaremenko, S.G. Patereu and M.M. Todua, "GMDH algorithm for synthesis of multiplicative models of complex systems", *Soviet Automatic Control c/c of Avtomatika,* vol.6, no.6, pp.52-57, 1973.

[111] A.G. Ivakhnenko, "Developing and applying the group method of data handling for modelling and long-range prediction – Part II causes of inaccurate predictions", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.18, no.6, pp.47-54, 1985.

[112] G.I. Krotov and Yu.V. Kostenko, "An algorithm for self-organization of additive-multiplicative models of complex systems", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.6, pp.73-78, 1987.
[113] K. Hara, T. Yamamoto and K. Terada, "Improved dual mode GMDH with automatic switch", *International Journal of Systems Science,* vol.21, no.8, pp.1553-1565, 1990.

[114] R.K. Mehra, "Group method of data handling (GMDH): review and experience", in *Proceedings of the IEEE Conference on Decision and Control,* pp.29-34, 1977.

[115] K. Hara, T. Yamamoto and K. Noguchi, "GMDH utilizing BDS with insufficient modelling data", *International Journal of Systems Science,* vol.17, no.12, pp.1677-1692, 1986.

[116] K. Hara, T. Yamamoto and K. Terada, "Prediction by dual mode GMDH", *International Journal of Systems Science,* vol.19, no.12, pp.2673-2681, 1988.

[117] A.G. Ivakhnenko, "Sorting methods in self-organization of models and clusterizations (review of new basic ideas). Iterative (multirow) polynomial GMDH algorithms", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.22, no.4, pp.88-99, 1989.

[118] A.G. Ivakhnenko, "Development and application of the group method of data handling for modelling and long-range prediction", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.18, no.3, pp.26-38, 1985.

[119] D.O. Molnar, "Modelling accuracy improvements for the group method of data handling", in P*roceedings of the 1st International Conference on Neural Networks,* vol.4, pp.839-846, 1987.

[120] M.A. Styblinski, S.A. Aftab and L.J. Opalski, "Modelling circuit performance functions by a combination of physical models and black-box approximation", in *Proceedings of the 1992 IEEE International Symposium on Circuits and Systems,* vol.2, pp.867-870, 1992.

[121] A.G. Ivakhnenko, A.A. Zholnarskiy and J.A. Muller, "An algorithm of harmonic rebinarization of a data sample", *Journal of Automation and information Sciences c/c of Avtomatika,* vol.25, no.6, pp.34-38, 1992.

[122] A.G. Ivakhnenko and J.A. Muller, "Present state and new problems of further GMDH development", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.3-16, 1995.

[123] J.A. Muller and G.A. Ivakhnenko, "Recent developments of self-organising modeling in prediction and analysis of stock market", available in URL address: *http://www.inf.kiev.ua/GMDH-home/articles/,* 1996.

[124] A.G. Ivakhnenko and G.A. Ivakhnenko, "Problems of further development of the group method of data handling algorithms. Part I", *Pattern Recognition and Image Analysis,* vol.10, no.2, pp.187-194, 2000.

[125] C. Robinson, *"Multi-objective optimisation of polynomial models for time series prediction using genetic algorithms and neural networks",* PhD Thesis in the Dept. of Automatic Control & Systems Engineering, University of Sheffield, UK, 1998.

[126] D.E. Scott and C.E. Hutchinson, "The GMDH algorithm – a technique for economic modelling", *Modelling and Simulation,* vol.7, pp.729-733, 1976.

[127] P.C. Parks, A.G. Ivakhnenko, L.M. Boichuk and B.K. Svetalsky, "A self-organizing model of the british economy for control with optimal prediction using the balance of variables criterion", *International Journal of Computer and Information Sciences,* vol.4, no.4, pp.349-379, 1975.

[128] A.G. Ivakhnenko and Yu.V. Kostenko, "Systems analysis and long-range quantitative prediction of quasi-static systems on the basis of self-organization of models. Part 1. Systems analysis at the level of trends", *Soviet Automatic Control c/c of Avtomatika,* vol.15, no.3, pp.9-17, 1982.

[129] A.G. Ivakhnenko, Yu.V. Kostenko and I.V. Goleusov, "Systems analysis and long-term quantitative prediction of quasi-static systems on the basis of self-organization of models. Part 2. Objective systems analysis without a priori specification of external influences", *Soviet Automatic Control c/c of Avtomatika,* vol.16, no.3, pp.1-8, 1983.

[130] B.Y. Brusilovskiy, N.A. Ivakhnenko, I.V. Shabalina and Yu.P. Yurachkovskiy, "Prediction of economic indices", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vo.18, no.4, pp.42-45, 1985.

[131] I.V. Goleusov and S.A. Kondrasheva, "Comparative analysis of the interdependence structure of the macro-economic indices of COMECON member-countries by the group method of data handling", *Soviet Journal of Automation and information Sciences c/c of Avtomatika,* vol.20, no.3, pp.39-43, 1987.

[132] A.G. Ivakhnenko, A.N. Ivakhnenko, Yu.V. Kostenko, J.A. Muller, A.P. Sarychev and Yu.P. Yurachkovskiy, "Nonparametric forecasting GMDH models. Part 3. Models in the pattern- and cluster-analysis language, for forecasting processes in economic macrosystems", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.22, no.3, pp.1-14, 1989.

[133] J.A. Muller and F. Lemke, *Self-Organising Data Mining. An Intelligent Approach to Extract Knowledge from Data,* Dresden, Berlin, 1999.

[134] P.R. Water, S. Wibier, E.J.H. Kerckhoffs and H. Koppelaar, "GMDH-based stock price prediction", *Neural Network World,* vol.7, no.4-5, pp.552-563, 1997.

[135] V.I. Dolgopolov, "A model of the dynamics of the magnetic field of active solar regions with determination of the source function by the group method of data handling", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.68-71, 1986.

[136] F-J. Chang and Y-Y. Yuan, "A self-organization algorithm for real-time flood forecast", *Hydrological Processes,* vol.13, pp. 123-138, 1999.

[137] D.E. Catlin, *Estimation Control and the Discrete Kalman Filter,* Springer-Verlag, New York, 1989.

[138] S.F. Kozubovskiy and V.V. Kupriyanov, "Design of automatic systems for the control and protection of the air environment on the basis of GMDH", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.3, pp.63-66, 1987.

[139] A.G. Ivakhnenko and G.A. Ivakhnenko, "Simplified linear programming algorithm as basic tool for open-loop control", *Systems Analysis Modelling Simulation,* vol.18-19, pp.315-319, 1995.

[140] A.G. Ivakhnenko and G.A. Ivakhnenko, "Simplified linear programming algorithm as basic tool for open-loop control", *Systems Analysis Modelling Simulation,* vol.22, no.3-4, pp.177-184, 1996.

[141] A.G. Ivakhnenko and G.A. Ivakhnenko, "Normative prediction and optimal control of multivariate objects by simplified linear programming", *Pattern Recognition and Image Analysis,* vol.7, no.2, pp.218-221, 1997.

[142] A.G. Ivakhnenko and G.A. Ivakhnenko, "Normative forecasting and optimal control for multidimensional objects using a self-organization of a system of non-physical models", *Journal of Automation and Information Sciences c/c of Avtomatika,* vol.29, no.4-5, pp.162-168, 1997.

[143] J. Kus and Z. Banaszak, "Group method of data handling in technical diagnostic tasks", *Applied Mathematics and Computer Science,* vol.3, no.3, pp.573-593, 1993.

[144] J. Korbicz and J. Kus, "Neural networks of the GMDH type and their applications in technical diagnosis", in *Proceedings of the Fourth International Symposium on Methods and Models in Automation and Robotics,* vol.3, pp.995-1000, Tech. Univ. Szczecin, Szczecin, Poland, 1997.

[145] J. Kus and J. Korbicz, "Self-organization of the GMDH-type neural networks in analysis of dynamical systems", in *Proceedings of the Fifth International Symposium on Methods and Models in Automation and Robotics,* vol.2, pp.639-644, Tech. Univ. Szczecin, Szczecin, Poland, 1998.
[146] A.G. Ivakhnenko and V.V. Osipenko, "Prediction of rare events on the basis of a GMDH algorithm", *Soviet Automatic Control c/c of Avtomatika,* vol.17, no.5, pp.6-10, 1984.

[147] V.V. Khubayev, "Logarithmic base functions with superimposed biological relationships for productivity models of agricultural systems", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.99-103, 1986.

[148] A.G. Ivakhnenko and V.S. Stepashko, "Use of the group method of data handling in predicting random processes", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.5, pp.1-10, 1986.

[149] T. Buttner, J.A. Muller and S.F. Kozubovskiy, "Application of self-organization theory to analysis and prediction of demographic processes", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.47-50, 1986.

[150] A.P. Sarychev, "A GMDH iteration algorithm for modelling random fields in a certain class of superposition of two-dimensional beta distributions", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.2, pp.25-29, 1987.

[151] Z.S. Lin, J. Liu and X.D. He, "The self-organizing methods of long-term forecasting (I) – GMDH and GMPSC model", *Meteorology and Atmospheric Physics,* vol.53, no.3-4, pp.155-160, 1994.

[152] E.M. Bielinska and J.E. Nabaglo, "Comparison of different method of bilinear time series prediction", in *Proceedings of the Third IEEE Conference on Control Applications,* vol.3, pp.1835-1839, 1994.

[153] M. Iwasaki, T. Shibata and N. Matsui, "GMDH-based autonomous modeling and compensation for nonlinear friction", in *Proceedings of the 1999 IEEE/ASME International Conference on Advanced Intelligent Mechatronics,* pp.860-865, Atlanta, USA, IEEE, 1999.

[154] Y. Xue and J. Watton, "A self-organizing neural network approach to data-based modelling of fluid power systems dynamics using the GMDH algorithm", in *Proceedings of the Institution of Mechanical Engineers. Part I (Journal of Systems and Control Engineering),* vol.209, no.I4, pp.229-240, 1995.

[155] N.A. Semenov and Ye.V. Malinovskaya, "Structural-parametric identification of polynomial models based on GMDH algorithms and Brandon's method", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.20, no.3, pp.17-19, 1987.

[156] I.V. Tetko, T.I. Aksenova, V.V. Volkovich, T.N. Kasheva, D.V. Filipov, W.J. Welsh, D.J. Livingstone and A.E.P. Villa, "Polynomial neural network for linear and non-linear model selection in quantitative-structure activity relationship studies on the internet", available in URL address: *http://skyscraper.fortunecity.com/cray/793/articles/application.html,* 2000.

[157] G.I. Krotov, Yu.V. Koppa and V.S. Stepashko, "Interactive modelling of complex objects based on GMDH algorithms", *Cybernetics and Computing Technology c/c of Kibernetika-I-Vychislitel'naya-Tekhnika,* vol.104, pp.49-53, 1994.

[158] F. Lemke, "SelfOrganize! – a software tool for modelling and prediction of complex systems", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.17-27, 1995.

[159] F. Lemke and J.A. Muller, "Self-organising modelling in financial risk control", *Proceedings of the 15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics,* vol.6 (Application on Modelling and Simulation), pp.733-738, 1997.

[160] J.A. Muller and F. Lemke, "Self-organizing modelling and decision support in economics", *Systems Analysis Modelling Simulation,* vol.18-19, pp.135-138, 1995.

[161] A.G. Ivakhnenko, "The problem of regularization and unimodality of the consistency criterion and its solution in the algorithms for objective system analysis and objective computer clustering", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.3, pp.11-15, 1988.

[162] A.G. Ivakhnenko and N.A. Ivakhnenko, "Nonparametric GMDH predicting models. Part 2. Indicative systems for selective modelling, clustering and pattern recognition", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.22, no.2, pp.1-10, 1989.

[163] A.P. Sarychev, "Solution of the GMDH partitioning problem when calculating the regularity criterion in an active experiment", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.22, no.4, pp.18-26, 1989.

[164] A.P. Sarychev, "An averaged regularity criterion for the group method of data handling in the problem of searching for the best regression", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.23, no.5, pp.24-29, 1990.

[165] A.P. Sarychev, "The J-optimal set of regressors determination by the repeated samples of observations in the group method of data handling", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.59-67, 1995.

[166] A.P. Sarychev, "Using repeated observation samples to determine a J-optimal set of regressors", *Journal of Automation and Information Sciences c/c of Avtomatika,* vol.26, no.3, pp.57-64, 1993.

[167] T.I. Aksenova and Yu.P. Yurachkovskiy, "A characterisation of unbiased structures and conditions for their J-optimality", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.4, pp.36-41, 1988.

[168] V.S. Stepashko, "Selective properties of the consistency criterion of models", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.19, no.2, pp.38-46, 1986.

[169] V.S. Stepashko, "Asymptotic properties of external criteria for model selection", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.21, no.6, pp.84-92, 1988.

[170] T.I. Aksenova, "Sufficient conditions and convergence rate using different criteria for model selection", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.69-78, 1995.

[171] A.G. Ivachnenko and J.A. Muller, "Parametric and non parametric selection procedures in experimental systems analysis", *Systems Analysis Modelling Simulation,* vol.9, no.2, pp.157-175, 1992.

[172] V.Y. Braverman and O.O. Fomychov, "Some aspects of the use of the group method of data handling in the case of small samples", *Soviet Automatic Control c/c of Avtomatika,* vol.7, no.2, pp.26-32, 1974.

[173] V.B. Silov and D.V. Vilenchic, "Linguistic decision-making methods in multicriterial selection of models", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.18, no.4, pp.92-94, 1985.

[174] V.P. Belogurov, "A criterion of model suitability for forecasting quantitative processes", *Soviet Journal of Automation and Information Sciences c/c of Avtomatika,* vol.23, no.3, pp.21-25, 1990.

[175] C. Hild and H. Bozdogan, "The use of information-based model evaluation criteria in the GMDH algorithm", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.29-50, 1995.

[176] T. Lange, "Structure criteria for automatic model selection in multilayered GMDH algorithms in case of uncertainty of data", *Systems Analysis Modelling Simulation,* vol.20, no.1-2, pp.79-91, 1995.