A HYBRID GMDH AND LEAST SQUARES SUPPORT VECTOR MACHINES IN TIME SERIES FORECASTING

R. Samsudin[*], *P. Saad*[*], *A. Shabri*[†]

Abstract: Time series consists of complex nonlinear and chaotic patterns that are difficult to forecast. This paper proposes a novel hybrid forecasting model which combines the group method of data handling (GMDH) and the least squares support vector machine (LSSVM), known as GLSSVM. The GMDH is used to determine the useful input variables for the LSSVM model and the LSSVM model that works as time series forecasting. Three well-known time series data sets are used in this study to demonstrate the effectiveness of the forecasting model. These data are utilized to forecast through an application aimed to handle real life time series. The results found by the proposed model were compared with the results of the GMDH and LSSVM models. Experiment result indicates that the hybrid model was a powerful tool to model time series data and provides a promising technique in time series forecasting methods .

Key words: *Group method of data handling, least square support vector machine, autoregressive integrated moving average, neural network, Box-Jenkins method*

Received: August 18, 2010
Revised and accepted: June 13, 2011

1. Introduction

Accurate forecasting of time series data has been one of the most important issues in hydrological research. The most comprehensive of all popular and widely known statistical models, which have been utilized in the last four decades for time series forecasting, are the Autoregressive Integrated Moving Average (ARIMA) model

[*]R. Samsudin, P. Saad
Faculty of Computer Science and Information System, 81310, Skudai, Johor, Malaysia, E-mail: ruhaidah@utm.my, Tel: +6019-7444268, Fax : +607-5565044, E-mail: puteh@utm.my, Tel: +607-5532344, Fax : +607-5565044
[†]A. Shabri
Faculty of Science, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia, E-mail: ani@utm.my, Tel: +607-5534238, Fax : +607-5566162

[39,40,41,42]. The ARIMA model is only a class of linear model and thus it can only capture linear feature of data time series.

More advanced AI is support vector machine (SVM), which was proposed by Vapnik and his co-workers in 1995 through statistical learning theory [1]. The SVM is a powerful methodology and has become a hot topic of intensive study due to its successful application to solve most non-linear regression and time series problems and its growing use in the modeling and forecasting of time series processes. The standard SVM is solved by using quadratic programming methods. However, this method is often time-consuming and has a higher computational burden because of the required constrained optimization programming.

Least squares support vector machine (LSSVM), as a modification of SVM, was introduced by Suykens [2]. The method uses equality constraints instead of inequality constraints and adopts the least squares linear system as its loss function, which is computationally attractive. The LSSVM also has good convergence and high precision. Hence, this method is easier to use than quadratic programming solvers in the SVM method. Extensive empirical studies [3] have shown that the LSSVM is comparable to the SVM in terms of generalization performance. The major advantage of the LSSVM is that it is computationally very cheap while it still possesses some important properties of the SVM.

One sub-model of ANN is a group method of data handling (GMDH) algorithm. It was first developed by Ivakhnenko [4] as a multivariate analysis method for modeling and identification of complex systems. The main idea of the GMDH is to build an analytical function in a feed-forward network based on a quadratic node transfer function whose coefficients are obtained by using a regression technique. This model has been successfully used to deal with uncertainty, linear or nonlinearity of systems in a wide range of disciplines, such as engineering, science, economy, medical diagnostics, signal processing and control systems [5,6,7,8,9].

There have been several studies suggesting hybrid models, combining the ARIMA and ANN model [10,11,12,13,38], the GMDH and ANN model [14], GMDH and differential evolution [9], ARIMA and support vector machine (SVM) [15], ANN and Fuzzy system [16], ANN and SVM [37], ANN and Genetic Algorithm [43,46], Particle Swarm Optimization and SVM [44,45]. Their results showed that the hybrid model can be an effective way to improving predictions achieved by either of the models used separately.

In this paper, a novel hybrid GMDH-type algorithm is proposed by integrating simple the GMDH with the LSSVM to forecast time series data. The hybrid model combines the GMDH and the LSSVM into one methodology, known as the GLSSVM. To verify the application of this approach, three well-known data sets that always handled in real life time series application are used in this study. There are the Canadian lynx data, Wolf's sunspot data and the international airline passengers.

2. Individual Forecasting Models

This section presents the GMDH, LSSVM and combination of the GMDH and LSSVM models used for modeling time series. The models were chosen in this study because these methods have been widely and successfully used in forecasting time series.

2.1 The least square vector machines (LSSVM) model

The LSSVM is a new technique for regression. The LSSVM predictor is trained using a set of time series historic values as inputs and a single output as the target value. In the following, we briefly introduce the LSSVM, which can be used for time series forecasting.

Consider a given training set of n data points $\{x_i, y_i\}_{i=1}^{n}$ with input data $x_i \in R^n$, p is the total number of data patterns and output $y_i \in R$. SVM approximate the function in the following form

$$y(x) = w^T \phi(x) + b, \tag{1}$$

where $\phi(x)$ represents the high dimensional feature spaces, which is nonlinearly mapped from the input space x. In the LSSVM, for function estimation, the optimization problem is formulated (Suykens et al., 2002) as follows:

$$\min\ J(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{n} e_i^2. \tag{2}$$

Subject to the equality constraints

$$y(x) = w^T \phi(x_i) + b + e_i, \quad i = 1,\ 2,\ ...,\ n.$$

The solution is obtained after constructing the Lagrange

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^{N} \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\}.$$

With Lagrange multipliers α_i. The conditions for optimality are given by

$$\frac{\partial L}{\partial w} = 0 \to w = \sum_{i=1}^{N} \alpha_i \phi(x_i),$$

$$\frac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{N} \alpha_i = 0,$$

$$\frac{\partial L}{\partial e_i} = 0 \to \alpha_i = \gamma e_i,$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \to w^T \phi(x_i) + b + e_i - y_i = 0,$$

for $i = 1,\ 2,\ ...,\ n$. After elimination of e_i and w the solution is given by the following set of linear equations:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \phi(x_i)^T \phi(x_l) + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix},$$

where $y = [y_1; \ldots; y_n]$, $\mathbf{1} = [1; \ldots; 1]$, $\alpha = [\alpha_1; \ldots; \alpha_n]$. According to Mercer's condition, the kernel function can be defined as

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j), \quad i,\ j = 1,\ 2,\ \ldots,\ n. \tag{3}$$

This finally leads to the following LSSVM model for function estimation:

$$y(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b, \tag{4}$$

where α_i, b are the solution to the linear system. Any function that satisfies Mercer's condition can be used as the kernel function. The choice of the kernel function $K(.,.)$ has several possibilities. $K(x_i, x_j)$ is defined as the kernel function. The value of the kernel is equal to the inner product of two vectors $\mathbf{X}_i$ and $\mathbf{X}_j$ in the feature space $\phi(x_i)$ and $\phi(x_j)$, that is, $K(x_i, x_j) = \phi(x_i) * \phi(x_j)$. The structure of an SVM is shown in Fig. 1.

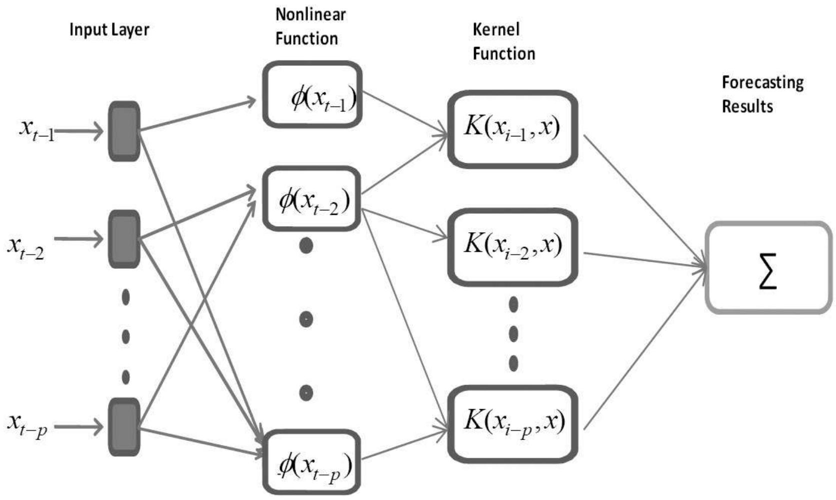

Fig. 1 *Structure of an SVM.*

The typical examples of the kernel function are as follows:

Linear:	$K(x_i, x_j) = x_i^T x_j$
Sigmoid:	$K(x_i, x_j) = \tanh(\gamma\, x_i^T x_j + r)$
Polynomial:	$K(x_i, x_j) = (\gamma\, x_i^T x_j + r)^d, \quad \gamma > 0$
Radial basis function (RBF):	$K(x_i, x_j) = \exp(-\gamma\, \lVert x_i - x_j \rVert^2), \quad \gamma > 0 \qquad (5)$

Here γ, r and d are kernel parameters. The kernel parameters should be carefully chosen as they implicitly define the structure of the high dimensional feature space $\phi(x)$ and thus control the complexity of the final solution.

2.2 The group method of the data handling (GMDH) model

The GMDH was introduced by Ivakhnenko in early 1970 as a multivariate analysis method for modeling and identification of complex systems. The GMDH method was originally formulated to solve higher order regression polynomials, specially for solving modeling and classification problem. General connection between inputs and output variables can be expressed by a complicated polynomial series in the form of the Volterra series, known as the Kolmogorov-Gabor polynomial [4]:

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} x_i x_j x_k + \ldots, \tag{6}$$

where x is the input to the system, M is the number of inputs and a are coefficients or weights. However, for most applications the quadratic forms are called as partial descriptions (PD) for only two variables are used in the form

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{7}$$

to predict the output. To obtain the value of the coefficients a for each m models, a system of Gauss normal equations is solved. The coefficient a_i of nodes in each layer is expressed in the form

$$\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-\mathbf{1}}\mathbf{X}^T\mathbf{Y}, \tag{8}$$

where $\mathbf{Y} = [y_1\ y_2 \ldots y_M]^T$, $\mathbf{A} = [a_0,\ a_1,\ a_2, a_3, a_4, a_5]$,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{Mp} & x_{Mq} & x_{Mp}x_{Mq} & x_{Mp}^2 & x_{Mq}^2 \end{bmatrix} \tag{9}$$

and M is the number of observations in the training set.

The main function of GMDH is based on the forward propagation of signal through nodes of the net similar to principal used in classical neural nets. Every layer consists of simple nodes each of which performs its own polynomial transfer function and passes its output to nodes in the next layer. The basic steps involved in the conventional GMDH modeling [18] are as follows:

Step 1: Select normalized data $X = \{x_1, x_2, \ldots, x_M\}$ as input variables. Divide the available data into training and testing data sets.

Step 2: Construct $MC_2 = M(M-1)/2$ new variables in the training data set and construct the regression polynomial for first layer by forming the quadratic expression which approximates the output y in Eq. (7).

Step 3: Identify the contributing nodes at each hidden layer according to the value of mean root square error (RMSE). Eliminate the least effective variable with replace the columns of X (old columns) by the new columns of Z.

Step 4: The GMDH algorithm is carried out by repeating steps 2 and 3 of the algorithm. When the errors of the test data in each layer stop decreasing, the iterative computation is terminated.

The configuration of the conventional GMDH structure is shown in Fig. 2.

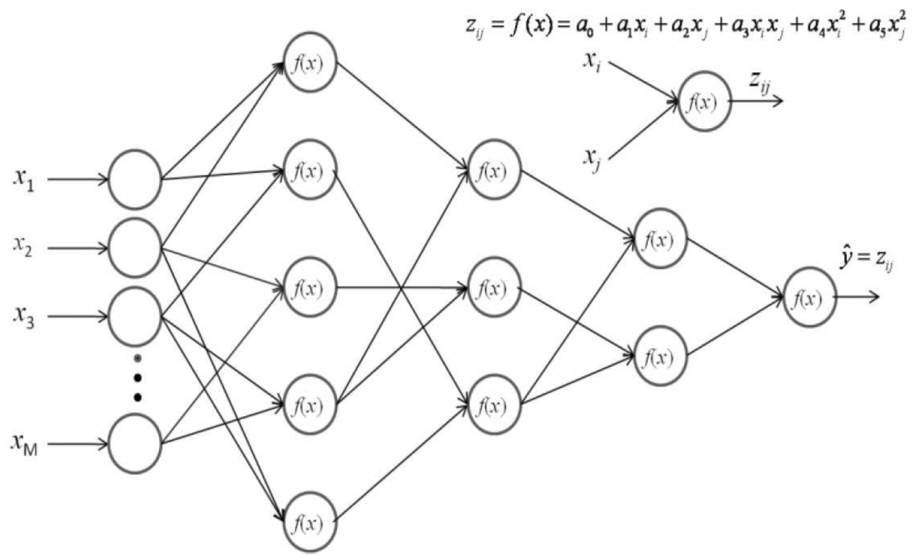

Fig. 2 *Basic structure of conventional GMDH.*

2.3 The hybrid model

In the proposed method, the combination of GMDH and LSSVM (GLSSVM) is applied to enhance the capability of the hybrid model. As input variables are selected by the decision made by GMDH and LSSVM model is used as time series forecasting. The hybrid model procedure is carried out in the following step:

Step 1: The normalized data are separated into the training and testing sets data.

Step 2: All combinations of two input variables (x_i, x_j) are generated in each layer. The number of input variables are $MC_2 = M(M-1)/2$. Construct the regression polynomial for this layer by forming the quadratic expression which approximates the output y in Eq. (10). The coefficient vector of the PD is determined by the least square estimation approach.

Step 3: Determine new input variables for the next layer. The output x' variable which gives the smallest of root mean square error (RMSE) for the train data set is combined with the input variables $\{x_1, x_2, \ldots, x_M, x'\}$ with $M = M + 1$. The new input $\{x_1, x_2, \ldots, x_M, x'\}$ of the neurons in the hidden layers is used as input for the LSSVM model.

Step 4: The GLSSVM algorithm is carried out by repeating steps 2 to 3 until $k = 5$ iteration. The GLSSVM model with the minimum value of the RMSE is selected as the output model. The configuration of the GLSSVM structure is shown in Fig. 3.

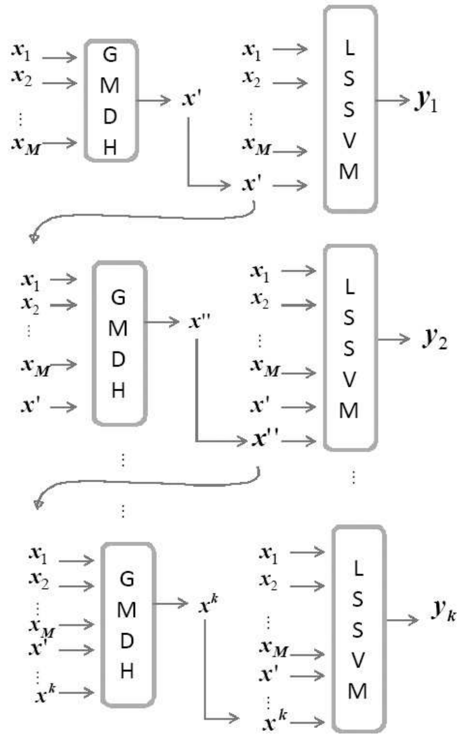

Fig. 3 *Structure of the GLSSVM model for time series forecasting.*

3. Empirical Results

In this section, we illustrate the hybrid GMDH-type algorithm and show its performance for a number of well-known and widely used datasets. The first one is a time series of the Canadian lynx data which was studied previously by Moran [19], Kajitani et al. [20], Subba Rao and Gabr [21], Zhang [10], Aladag et al. [22], and Khashei and Bijari [23]. The other one containing airline passenger data deals with nonlinear behavior and shows multiplicative seasonal behavior being already exploited in time series [24,25,26,27]. The third one is concerned with Wolf's sunspot data [28,21,10,23]. These time series come from different areas and have different statistical characteristics.

A. Lynx Series The first series that is considered is the lynx series which contains the number of lynx trapped per year in the Mackenzie River district of Northern Canada. The data set has 114 observations, corresponding to the period of 1821-1934. It has also been extensively analyzed in the time series literature with a focus on nonlinear modeling. These lynx data are one of the most frequently used time series. The data are plotted in Fig. 4, which shows a periodicity of approximately 10 years.

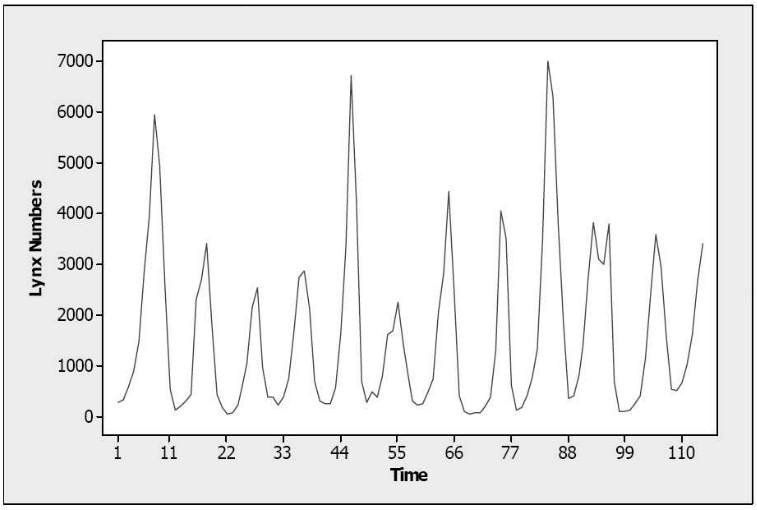

Fig. 4 *Canadian lynx data series (1821–1934).*

In designing the LSSVM, GMDH and GLSSVM models, one must determine the following variables: the number of input nodes and the number of layers. The selection of the number of input corresponds to the number of variables which play important roles in many successful applications of the ANN and GMDH models.

To make these models simple and to reduce some of the computational burden, only the lagged variables obtained from the Box-Jenkins are used as input layers. The lynx series was studied by many researchers and the first time series analysis was carried out by Moran [19] and then recently by Kajitani et al. [20] who fit an AR(2) model to the logged data. Subba Rao and Gabr [21], Zhang [10] and Khashei and Bijari [23] found the best-fitted model is AR(12) model.

Hence, in this study, based on Box-Jenkins methodology, the AR(2) and AR(12) models are linear modeling and are considered as nonlinear function of several past observations, respectively as follows

$$x_t = f(x_{t-1}, x_{t-2}) + a_t \tag{10}$$

and

$$x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-12}) + a_t, \tag{11}$$

where f is a nonlinear function determined by the LSSVM, GMDH and GLSSVM models. In the training and testing of these models, the same input structures of the data set are used. The precision and convergence of the LSSVM is also affected by (γ, σ^2). In the LSSVM model, parameter values for γ and σ^2 need to be first specified. There is no structured way to choose the optimal parameters of the LSSVM. In order to obtain the optimal model parameters of the LSSVM, a grid search algorithm was employed in the parameter space. Cross-validation is a popular technique for estimating generalization performance. To obtain good generalization ability, we conduct a validation process to decide parameters. In order to better evaluate the performance of the proposed approach, we consider a grid search of (γ, σ^2) with γ in the range 10 to 1000 and σ^2 in the range 0.01 to 1.0. For each hyperparameter pair (γ, σ^2) in the search space, 5-fold cross validation on the training set is performed to predict prediction error. The best fit model structure for each model is determined according to criteria of performance evaluation. In the study, the LSSVM model was implemented with software package LS-SVMlab1.5 (Pelckmans et al. 2002) using MATLAB. The LSSVM method is employed, so a kernel function has to be selected from the qualified function. Many works on the use of the LSSVM in time series modeling and forecasting have demonstrated the favorable performance of the RBF kernel (Liu & Wang, 2008, Yu et al., 2006; Gencoglu and Uyar, 2009). Therefore, the RBF kernel, which has a parameter γ as in Eq. (5), is adopted in this work. Tab. V shows the performance results obtained in the training and testing period of the LSSVM approach.

The GMDH works by building successive layers with complex connections that are created by using second-order polynomial function. The first layer created is made by computing regressions of the input variables. The second layer is created by computing regressions of the output value. Only the best are chosen at each layer and this process continues until a pre-specified selection criterion is found.

The proposed hybrid learning architecture is composed of two stages. In the first stage, the GMDH is used to determine the useful inputs for the LSSVM method. The estimated output values x' are used as the feedback value and are combined with the input variables $\{x_{t-1}, x_{t-2}, \ldots, x_{t-M}\}$ in the next loop calculations. In the second stage, the LSSVM mapping of the combination inputs variables $\{x_{t-1}, x_{t-2}, \ldots, x_{t-M}, x'\}$ seeks optimal solutions for determining the best output for forecasting.

The performances of the GMDH, LSSVM and GLSSVM for time series forecasting models for lynx data are given in Tab. I.

		Training		Testing	
Input	Model	MAE	MSE	MAE	MSE
M1	GMDH	0.1681	0.0444	0.0634	0.0082
	LSSVM	0.1613	0.0411	0.0657	0.0074
	GLSSVM	0.1681	0.0442	**0.0552**	**0.0056**
M2	GMDH	0.1522	0.0359	0.0623	0.0058
	LSSVM	**0.0898**	**0.0141**	0.1303	0.0301
	GLSSVM	0.1507	0.0346	0.0654	0.0067

Tab. I *Comparison of the GMDH, LSSVM and GLSSVM in training and testing.*

It is clear from Tab. I, in the training phase that the LSSVM obtained the best MSE and MAE statistic of 0.0898 and 0.0141, respectively. Analyzing the results during testing, considering the MSE and MAE being regarded here as a performance index, the experimental results clearly demonstrate that the GLSSVM outperforms the other models. Fig. 5 shows the actual and forecasted values respectively.

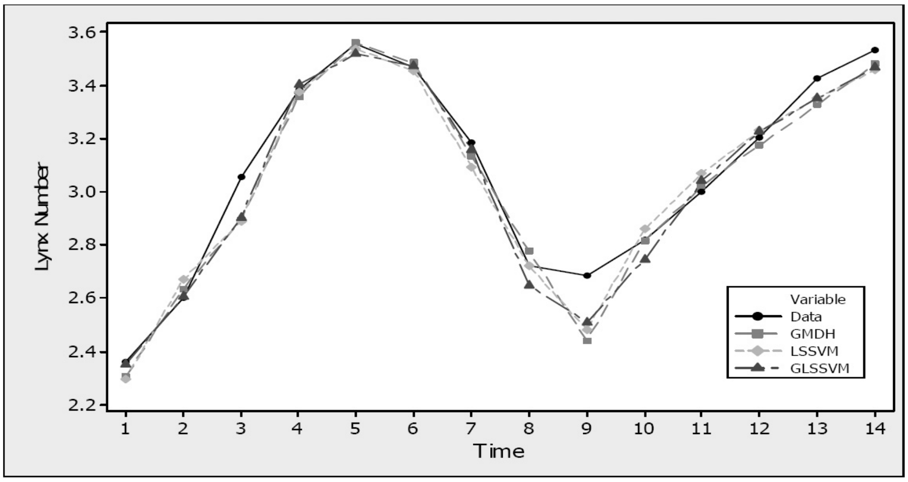

Fig. 5 *Comparison between observed and predicted for the GMDH, LSSVM and GLSSVM models for lynx time series (testing phase).*

Tab. II shows the performance of our proposed model and other models studied in the previous literature. The experiment results show that our proposed model offers encouraging advantages and has good performance.

Model	MSE	MAE
Zhang' ARIMA [10]	0.02049	0.1123
Zhang' ANN [10]	0.02046	0.1121
Zhang' Hybrid [10]	0.01723	0.10397
Khashei & Bijari' ANN [23]	0.01361	0.089625
Kajitani' SETAR [20]	0.01400	-
Kajitani' FNN [20]	0.0090	-
Aladag' Hybrid [22]	0.0090	-
Proposed Model (GLSSVM)	**0.00560**	**0.0552**

Tab. II *Comparison of performance of the proposed model with those of other forecasting models.*

B. The Airline Passenger Data The airline passenger data set was first used by Brown [33] and then by Box and Jenkins [24]. The airline passenger data set consists of the total number (in thousands) of passengers on international airlines from January 1949 to December 1960. Fig. 6 shows that the data have an upward trend together with seasonal variation whose size is roughly proportional to the local mean level called multiplicative seasonality. The airline series exhibits nonlinear behavior and shows multiplicative seasonal behavior.

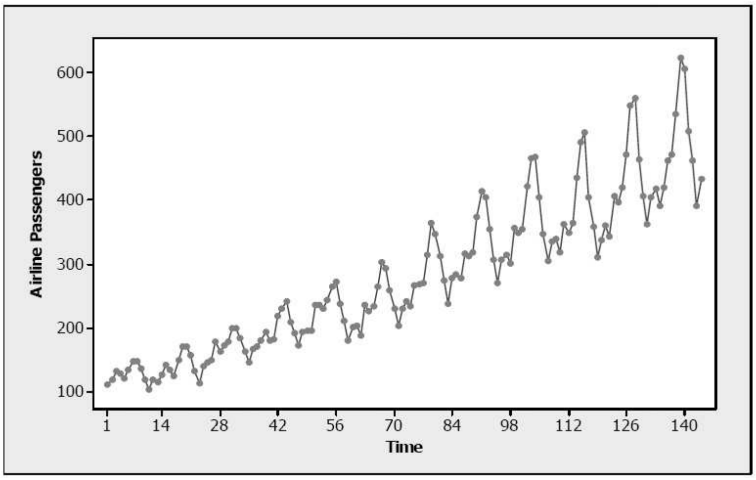

Fig. 6 *Airline passenger data series (Jan 1949 – Dec 1960).*

As in many other studies involving this time series, the data from the first 11 years (132 observations) are used for modeling and the 12 last observations are used for testing. For the airlines series, Box et al. [34] identified that the best model that fits the airline data is SARIMA(0,1,1)x$(0,1,1)_{12}$ after making logarithmic transformation same as that identified by other researchers, e.g. Newton (1988), Faraway and Chatfield [26].

Using the Box-Jenkins approach on the airlines data, the best model found is the SARIMA model of order (0,1,1)x$(0,1,1)_{12}$ given by

$$(1-B)(1-B^{12})x_t = (1-0.3484B)(1-0.5623B^{12})a_t$$

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} - 0.3484a_{t-1} - 0.5623a_{t-12} + 0.1959a_{t-13} + a_t. \quad (12)$$

For the airlines series presented in Eq. 12, the output x_t can be expressed as follows

$$x_t = f(x_{t-1}, x_{t-12}, x_{t-13}, a_{t-1}, a_{t-12}, a_{t-13}). \quad (13)$$

The nodes in the input layer consist of lagged variables x_{t-1}, x_{t-12}, x_{t-13} and the effects of random errors a_{t-1}, a_{t-12}, a_{t-13} on forecast. Hence, the LSSVM, GMDH and GLSSVM models have six input nodes in the input layer for the independent variables in the function f, and one output node in the out layer consists of the

	Training		Testing	
Model	MAE	MSE	MAE	MSE
GMDH	10.726	187.712	11.726	188.948
LSSVM	**4.786**	**37.376**	20.287	591.000
GLSSVM	7.426	103.371	**11.316**	**186.027**

Tab. III *Comparison of the GMDH, LSSVM, and GLSSVM in training and testing.*

prediction value at the next month. The results of the forecasting experiments for the LSSVM, GMDH and GLSSVM models are summarized in Tab. III.

Tab. III shows that in the training phase, the LSSVM model obtained the best MSE and MAE statistics of 37.376 and 4.786, respectively. However, analyzing the results during testing, it can be observed that the GLSSVM model outperforms all other models in terms of MSE and MAE. Fig. 7 shows the output results.

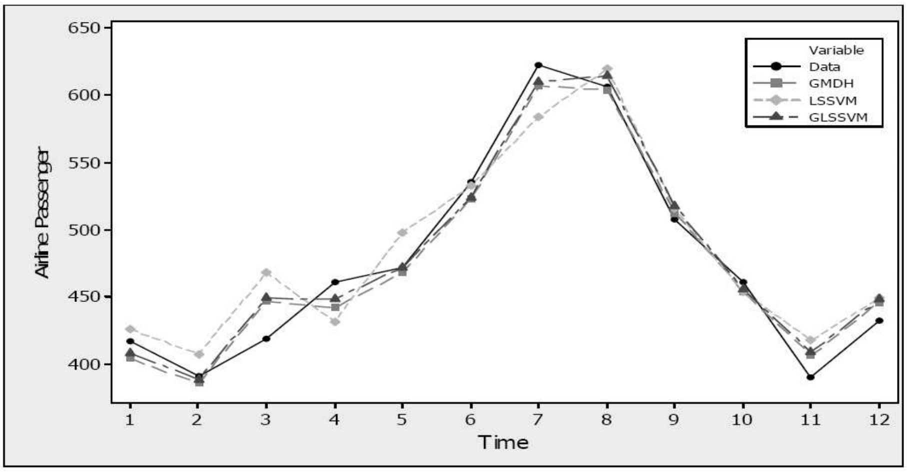

Fig. 7 *Comparison between observed and predicted for the GMDH, LSSVM and GLSSVM models for airlines passenger time series (testing phase).*

Tab. IV compares the GLSSVM with other models existing in the literature. For the purpose of comparison, the best value of sum of square error (SSE) for the ARIMA and ANN models found by Faraway and Chatfield [26] are also shown in Tab. I. From Tab. IV, using MSE (or SSE) as the performance index, the best SSE value found by the SARIMA method in the forecasting of this series is 4328, the best value of Faraway and Chatfield [26] found with ANN is 2900, and the best SSE value found by the GLSSVM is 2863. It is observed that the proposed GLSSVM model was significantly better than other models.

Model	SSE	MSE
Faraway' ARIMA [26]	4328	325.839
Faraway' ANN [26]	2900	241.670
Proposed model(GLSSVM)	2863	228.546

Tab. IV *Comparison of performance of the proposed model with those of other forecasting models.*

C. Box-Jenkins Furnace Time Series Finally, for further evaluation, we test a benchmark data set that is Box-Jenkins furnace time series [34]. This data set has been used extensively as a benchmark example for process identification. Box and Jenkins [24] used a time-series based approach to develop a model based on the gas furnace data set. This model has been referred to, repeatedly, as the linear model. The data set has since been used by many researchers including [35,36].

The data originally consist of 296 data points $[y(t), u(t)]$ from $t = 1$ to 296, where $y(t)$ represents the concentration of carbon dioxide in the gas mixture flowing out of a gas furnace under a steady air supply and u(t) represents the flow rate of the methane gas in a gas furnace. The data are plotted in Fig. 8. Here we are trying to predict $y(t)$ based on $\mathbf{x} = [x(t-p), .., x(t-1), y(t-p), \ldots, y(t-1)]$, where p is lag.

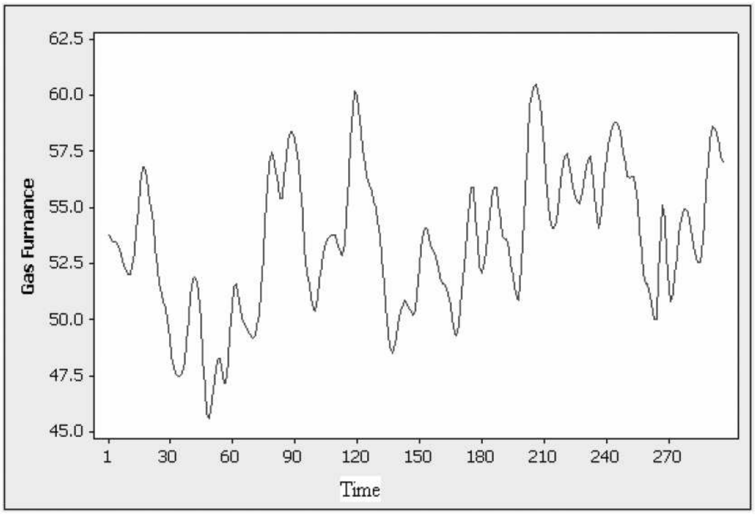

Fig. 8 *Time series of gas furnace.*

One of the most important steps in developing a satisfactory time series forecasting is the selection of the input variables. For the GMDH, LSSVM and GLSSVM models, there is no systematic approach which can be followed. The various input structures were tried. In this study, a number of lag (p), 1, 2 and 3 were considered. The input structures of forecasting models are shown in Tab. V.

Model	Input Structure
M1	$\mathbf{x} = [x(t-1), y(t-1)]$
M2	$\mathbf{x} = [x(t-2), x(t-1), y(t-2), y(t-1)]$
M3	$\mathbf{x} = [x(t-3), x(t-2), x(t-1), y(t-3), y(t-2), y(t-1)]$

Tab. V *Input structure of the models for Box-Jenkins furnace time series.*

In this study, 246 data samples are used for training and the remaining data samples are used for testing. Tab. VI reports the training and testing results of MAE and RMSE of different input structures for each of the individual forecasts and a hybrid forecasts for the furnace time series.

Input Structure	Model	Training MAE	Training MSE	Testing MAE	Testing MSE
M1	GMDH	0.3555	0.2034	0.5373	0.4814
	LSSVM	0.3456	0.1976	0.5474	0.4867
	GLSSVM	0.3426	0.1901	0.5454	0.4828
M2	GMDH	0.1146	0.0220	0.2624	0.1350
	LSSVM	0.0814	0.0106	0.2824	0.1478
	GLSSVM	0.0896	0.0129	**0.2560**	**0.1303**
M3	GMDH	0.1028	0.0174	0.3048	0.1937
	LSSVM	**0.0625**	**0.0071**	0.3480	0.2328
	GLSSVM	0.0699	0.0089	0.2885	0.1542

Tab. VI *The training and testing results of the GMDH, LSSVM and GLSSVM models.*

The results show that the best performance criteria (MAE = 0.0625, MSE = 0.0071) in training were obtained for the LSSVM model whose input structure is M3. For the testing phase, the best values of MAE and MSE (MAE = 0.2560, MSE = 0.1303) for the GLSSVM model was obtained using M2. The simulation results demonstrated that the new hybrid algorithm is more efficient than the conventional GMDH and LSSVM models. The observed and the best results obtained in predicted by GMDH, LSSVM and GLSSVM models are shown in Fig. 9.

Tab. VII contains a comparative analysis between the performances of the proposed GLSSVM model with other models studied in the literature. Compared with these models, the GLSSVM comes with a high accuracy and improved prediction capability.

4. Conclusion

There are plenty of models used to predict time series data. In this paper, we have demonstrated how the time series data could be well represented by the hybrid models, combining the GMDH and LSSVM models. To illustrate the capability

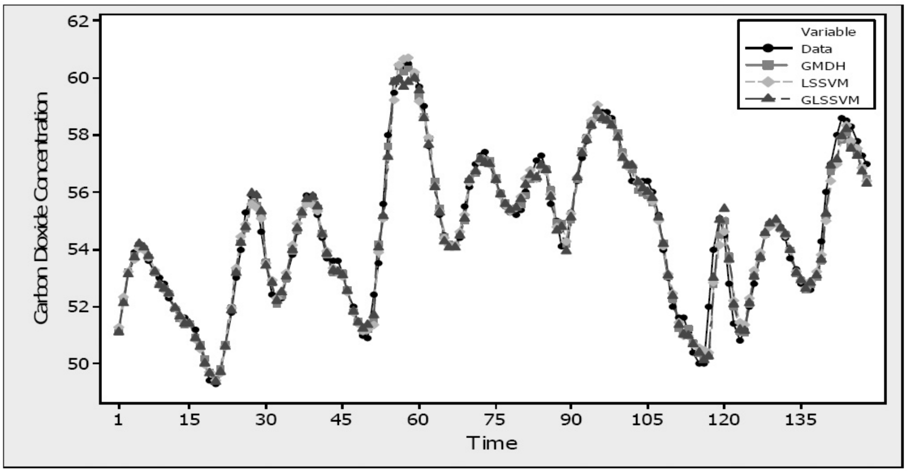

Fig. 9 *Comparison between observed and predicted for the GMDH, LSSVM and GLSSVM models for gas furnace problem (testing phase).*

	MSE	
Model	Training	Testing
Oh and Pedrycz's model [47]	0.020	0.271
Kim et al. model [48]	0.034	0.244
Lin and Cunningham's model [49]	0.071	0.261
Oh and Pedrycz's model Type 1 Basic case 1 [50]	0.017	0.148
Oh and Pedrycz's model Type 1 PNN case 2 [50]	0.017	0.147
Proposed model (GLSSVM)	**0.0129**	**0.1303**

Tab. VII *Comparison of performance of the proposed model with those of other forecasting models.*

of the LSSVM model, the lynx, airlines and gas furnace data were chosen as a case study. The time series data having various input structures are trained and tested to investigate the applicability of the GLSSVM compared with the GMDH and LSSVM models. One of the most important factor in developing a satisfactory forecasting model, such as the GMDH and LSSVM models, is the selection of the input variables. Empirical results on the three data sets using three different models clearly reveal the efficiency of the hybrid model. In terms of MSE and MAE values, for three data sets, the LSSVM model has the best ones in training, while proposed model has the best ones in testing. These results show that the hybrid model provides a robust modeling capable of capturing the nonlinear nature of the time series data and thus producing more accurate forecasts.

Acknowledgement

This work was supported by the Ministry of Science, Technology and Innovation, Malaysia.

References

[1] Vapnik V. N.: The nature of statistical learning theory. Springer-Verlag, New York, 1995.

[2] Suykens J. A. K., Vandewall J.: Least squares support vector machine classifiers, Neural Process. Lett., **9**, 3, 1999, pp. 293-300.

[3] Gestel T. V., Suykens J. A. K., Baesens B., Viaene S., Vanthienen J.: Benchmarking least squares support vector machine classifiers, Machine Learning, 54, 2004, pp. 5-32.

[4] Ivakhnenko A. G.: Polynomial theory of complex system, IEEE Trans. Syst., Man Cybern. SMCI-1, 1, 1971, pp. 364-378.

[5] Tamura H., Kondo T.: Heuristic free group method of data handling algorithm of generating optional partial polynomials with application to air pollution prediction, International Journal of Systems Science, 11, 1980, pp. 1095-1111.

[6] Ivakheneko A. G., Ivakheneko G. A.: A review of problems solved by algorithms of the GMDH, Pattern Recognition and Image Analysis, **5**, 4, 1995, pp. 527-535.

[7] Voss M. S., Feng X.: A new methodology for emergent system identification using particle swarm optimization (PSO) and the group method data handling (GMDH). In: Proc of the GECCO 2002, 2002, pp. 1227-1232.

[8] Lin Z. S., Liu J., He X. D.: The self organizing methods of long-term forecasting – GMDH and GMPSC model, Meteorology and Atmospheric Physics, 53, 1994, pp. 155-160.

[9] Onwubolu G. C.: Design of hybrid differential evolution and group method of data handling networks for modeling and prediction, Information Sciences, 178, 2008, pp. 3616-3634.

[10] Zhang G. P.: Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50, 2003, pp. 159-175.

[11] Jain A., Kumar A.: An evaluation of artificial neural network technique for the determination of infiltration model parameters, Applied Soft Computing, 6, 2006, pp. 272-282.

[12] Su C. T., Tong L. I., Leou C. M.: Combination of time series and neural network for reliability forecasting modeling, Journal of Chinese Industrial Engineering, 14, 1997, pp. 419-429.

[13] Wang W., Gelder P. V., Vrijling J. K.: Improving daily stream flow forecasts by combining ARMA and ANN models, In: Proc. Of the International Conference on Innovation Advances and Implementation of Flood Forecasting Technology, 2005.

[14] Wang X., Li L., Lockington D., Pullar D., Jeng D. S.: Self-organizing polynomial neural network for modeling complex hydrological processes, Research Report, R861, 2005, pp. 1-29.

[15] Chen K. Y., Wang C. H.: A hybrid ARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan, Expert Systems with Applications, 32, 2007, pp. 254-264.

[16] Yang Q., Lincang Ju L., Ge S., Shi R., Yuanli Cai Y.: Hybrid fuzzy neural network control for complex industrial process. In: Proc. of the International conference on intelligent computing, Kunming, CHINA, 2006, pp. 533-538.

[17] Suykens J. A. K., Van Gestel T., De Brabanter J., De Moor B., Vandewalle J.: Least squares support vector machines, World Scientific, Singapore, 2002.

[18] Zadeh N. N., Darvizeh A., Felezi M. E., Gharababaei H.: Polynomial Modelling of Exposive Process of Metalic Powders Using GMDH-type Neural Networks and Singular Value Decomposition. Modelling and Simulation in Materials Science and Engineering, 10, 6, 2002, pp. 727-744.

[19] Moran P. A. P.: The statistical analysis of the sunspot and lynx cycles, J. Anim. Ecol., 18, 1953, pp. 115-116.

[20] Kajitani Y., Hipel W. K., Mcleod A. I.: Forecasting nonlinear time series with feedforward neural networks: A case study of Canadian lynx data, Journal of Forecasting, 24, 2005, pp. 105-117.

[21] Subba Rao T., Gabr M. M.: An Introduction to Bispectral Analysis and Bilinear Time Series Models. Springer, Berlin, 1984.

[22] Aladag C. H., Egrioglu E., Kadilar C.: Forecasting nonlinear time series with a hybrid methodology, Applied Mathematics Letters, 22, 2009, pp. 1467-1470.

[23] Khashei M., Bijari M.: An artificial neural network (p, d, q) model for time series forecasting, Expert Systems with Applications, **37**, 1, 2010, pp. 479-489.

[24] Box G. E. P., Jenkins G.: Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970.

[25] Newton H. J.: TIMESLAB: A Time Series Analysis Laboratory, Pacific Grove, 1998.

[26] Faraway J., Chatfield C.: Time series forecasting with neural networks: a comparative study using the airline data, Appl. Statist., 47, 1987, pp. 231-250.

[27] Hamzaçebi C.: Improving artificial neural networks' performance in seasonal time series forecasting, Information Sciences, **178**, 2, 2008, pp. 4550-4559.

[28] Hipel K. W., Mcleod A. I.: Time Series Modelling of Water Resources and Environmental Systems, Elsevier Science B. V., Amsterdam, The Netherlands, 1994.

[29] Pelckmans K., Suykens J., Van Gestel T., De Brabanter J., Lukas L., Hanmers B., De Moor B., Vandewalle J.: LS-SVMlab: A MATLAB/C Toolbox for Least Square Support Vector Machines, available at: www.esat.kuleuven.ac.be/sista/lssvmlab, 2003.

[30] Liu L., Wang W.: Exchange Rates Forecasting with Least Squares Support Vector Machines. In: Proc. Of the International Conference on Computer Science and Software Engineering, 2008, pp. 1017-1019.

[31] Yu P. S., Chen S. T., Chang I. F.: Support vector regression for real-time flood stage forecasting, Journal of Hydrology, **328**, 3-4, 2006, pp. 704-716.

[32] Gencoglu M. T., Uyar M.: Prediction of Flashover Voltage of Insulators Using Least Square Support Vector Machines. Expert Systems with Applications, 36, 2009, pp. 10789-10798.

[33] Brown R. G.: Smoothing Forecasting and Prediction of Discrete Time Series, Prentice Hall, Englewood Cliffs, 1962.

[34] Box G. E. P., Jenkins G. M., Reinsel G. C.: Time Series Analysis: Forecasting and Control. 3rd Edn., Prentice Hall, New Jersey, ISBN-13: 978-0130607744, 1994, p. 592.

[35] Yadav A., Mishra D., Yadav R. N., Ray S., Kalra P. K.: Time-series prediction with single integrate-and-fire neuron, Applied Soft Computing, 7, 2007, pp. 739-745.

[36] Chen Y., Yang B. , Dong J., Abraham A.: Time-series forecasting using flexible neural tree model, Information Sciences, 174, 2005, pp. 219-235.

[37] Salcedo Sanz S., Ortiz-García E. G., Pérez-Bellido Á. M., Portilla-Figueras A., Prieto L., Paredes D., Correoso F.: Performance comparison of multilayer perceptrons and support vector machines in a short-term wind speed prediction problem, Neural Network World, 2009, pp. 37-51.

[38] Sallehuddin R., Shamsuddin S. M., Hashim S. Z. M., Abraham A.: Forecasting time series data using hybrid grey relational artificial neural network and auto regressive integrated moving, Neural Network World, 2007, pp. 573-605.

[39] Ho S. L., Xie M.: The use of ARIMA models for reliability forecasting and analysis, Computers Industrial Engineering, **35**, 1-2, 1998, pp. 213-216.

[40] Zhang G. P.: Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50, 2001, pp. 159-175.

[41] Ho S. L., Xie M., Goh T. N.: A comparative study of neural network and Box-Jenkins ARIMA modeling in time series predictions, Computers and Industrial Engineering, 42, 2002, pp. 371-375.

[42] Pai P. F., Lin C. S.: A hybrid ARIMA and support vector machines model in stock price forecasting, Omega, 33, 2005, pp. 497-505.

[43] Irigoyen E., Larrea M., Valera J., Gómez V., Artaza F.: A Hybridized neuro-genetic solution for controlling industrial R^3 workspace, Neural Network World, 10, 2010, pp. 811-824.

[44] Chen M. Y.: A hybrid model for business failure prediction – Utilization of particle swarm optimization and support vector machines, Neural Network World, 21, 2, 2011, pp. 129-152.

[45] Peters S., Koenig A.: A hybrid texture analysis system based on non-linear & oriented kernels, particle swarm optimization, and kNN vs. support vector machines, Neural Network World, 17, 2007, pp. 507-527.

[46] Sadil J.: Artificial intelligence methods used for predicting traction power consumption, Neural Network World, 18, 2008, pp. 15-22.

[47] Oh S. K., Pedrycz W.: Identification of fuzzy systems by means of an auto-tuning algorithm and its application to nonlinear systems, Fuzzy Sets Syst., **115**, 2, 2000, pp. 205-230.

[48] Kim E., Lee H., Park M.: A simple identified Sugeno-type fuzzy model via double clustering, Information Science, 110, 1998, pp. 25-39.

[49] Lin Y., Cunningham G. A.: A new approach to fuzzy-neural modeling, IEEE Trans. Fuzzy System, **3**, 2, 1995, pp. 190-197.

[50] Oh S.-K., Pedrycz W.: The design of self-organizing polynomial neural networks, Information Science, 141, 2002, pp. 237-258.