

Comparing NARX and NARMAX models using ANN and SVM for cash demand forecasting for ATM

Gonzalo Acuña, *Senior Member, IEEE* and Cristián
Ramirez
Universidad de Santiago de Chile, USACH
Departamento de Ingeniería Informática
Santiago, Chile
gonzalo.acuna@usach.cl

Millaray Curilem *Member, IEEE*
Universidad de la Frontera
Departamento de Ingeniería Eléctrica
Temuco, Chile
millaray@ufro.cl

Abstract—a comparative study between NARMAX and NARX models developed with ANN and SVM when used to forecast cash demand for ATMs is conducted. A simple methodology for developing SVM-NARMAX models is proposed. The best results were obtained with NARX-ANN models. In addition no significant differences were found between NARX and NARMAX for both ANN and SVM. Hence it seems advisable to choose simpler models, such as NARX and a user-friendly tool like ANN at least for this particular application.

Keywords—component; NARMAX; NARX; ANN; SVM; ATM; forecasting

I. INTRODUCTION

The need to have good predictive models is becoming a relevant issue because of the importance of anticipating different phenomena such as climatic variations, changes in stock values or the evolution of variables in complex industrial processes. Unfortunately, the complexity of all these phenomena makes the development of suitable dynamic models based on laws very difficult. A successful alternative approach to address this problem is to design data-driven models. In that sense, in the field of linear systems there is a vast literature originated in the pioneering work of Box and Jenkins [1]. They proposed the development of models that consider that the evolution of a phenomenon can be explained by its previous behavior and by the effect of exogenous variables when they exist.

In case of non-linear phenomena, nonlinear autoregressive models, NAR, or nonlinear autoregressive with exogenous variables (NARX) are proposed. The predictive power of these models can be increased when previous errors are incorporated as regressors [2]. This results in the so called nonlinear autoregressive moving average model NARMA or NARMAX when also exogenous variables are included. These models may be better predictors because they use information about past errors in order to improve the prediction. The price to pay, as will be seen later, is that they are more difficult to identify.

Extensive use of Artificial Neural Networks (ANN) in the field of system identification, predictive control, design of observers and predictors can be found in the literature [3, 4]. Despite the good results achieved with ANN some difficulties

remain in its design, such as choosing the number of neurons in the hidden layers, the problem of overfitting, the existence of local minima in the objective function and the low capacity of generalization, among others.

Support Vector Machine (SVM) are considered as very efficient tools for classification and regression. They have many advantages such as good generalization ability and an optimization process based on a convex function with no local minima [5]. In the case of dynamical systems, almost all works that use SVM focus on NARX type models [6] and few works report the use of SVM to develop NARMAX dynamic models. Suykens et al. [7], established the equations needed to train Least-Square SVM (LS-SVM) -a tool similar to traditional SVM but with some features that add greater simplicity of use- to develop NOE dynamic models (non-linear, with output error) with similar characteristics to NARMAX. However, they also established the great difficulty of handling such equations for practical implementations. Other authors mention NARMAX-SVM modeling applications but they are not clear enough about how SVM was used in the development of such dynamic models [8, 9]. In [10] a non linear response was simulated using SVM in a two stage implementation. In the first stage an ARMAX model is defined and its output is used to train the SVM to simulate the non linear response under study, in a one-step ahead prediction situation. No explanation is given about how the prediction behaves for multiple-step-ahead prediction. That is the reason why for this type of models ANN are still the preferred tool [11].

In a previous related work Simutis et al [12] compared ANN and SVM when forecasting cash demand for ATMs (Automatic Teller Machines) using only NARX type models while Ramirez and Acuña [13] also compare ANN and SVM for cash demand forecasting but using LS-SVM. In that paper NARMAX models were only used for ANN not for SVM.

In the present paper we propose a simple methodology for developing NARMAX-SVM models in order to conduct a comparative study with NARX equivalent models when used to forecast cash demand for ATMs. Additional comparisons with ANN performance are also studied.

II. AUTOMATIC TELLER MACHINES (ATM)

ATMs are funded and administered by financial institutions that make available to customers a simple method for conducting financial transactions in a public space with almost no human intervention. According to estimates developed by ATMIA (ATM Industry Association) the number of ATMs worldwide for 2007 exceeded 1.6 million units [12].

Some banks typically maintain an excess of up to 40% more cash in their terminals (ATM) of what they really need. In this regard, many experts believe that holding excess money appropriated is roughly 15% to 20% [12].

Costs related to keeping cash in an ATM represent among 35% to 60% of total maintenance costs [12]. Through improvements in management and cash management, banks can avoid falling into losses in new business opportunities due to having high cash assets. This is why it is necessary to develop new methods and advanced ways of estimating the demand for money at an ATM. Based on more accurate predictions and actual adjusted cash flow, financial institutions can lower their operating costs.

Banks and financial services assume that the demand for cash can be associated with certain variables that can have substantial effects on the level of demand for cash. Some of these variables we consider are the following [14]:

- ATM Location
- Seasonal factors such as weekends, holidays, etc.
- Historical data from the ATM.

Previous works demonstrate that the demand for cash is a non linear problem, thus it is necessary to use non linear modeling tools.

III. MATERIALS

A. Data processing

Data come from the NN5 competition [15] and correspond to a set of 30 series of ATM's withdrawal on a daily basis from ATMs located in different parts of England. All series show a strong cyclical component of 7 days, as well as some recurring seasonal periods such as summer holidays or Christmas. Almost all series contain empty values (missing values) and some of these series show long-term trends or irregularities such as outliers or "gaps" [16].

For each of the 30 ATMs a training set (387 examples), a validation set (300 examples) and a test set (100 examples) were considered.

B. Outliers and Gaps

All series include 3 types of "Gaps" or singularities.

- Observations equal to 0, indicating that no withdrawals have taken place due to "cash out" of the ATM.
- "Missing Values" indicating that on that day the client's transaction was not recorded.

- Outliers, indicating which data is above or below the normal behavior of withdrawals at the ATM.

This research addresses the 3 types of abnormalities, detecting outliers, missing values and values equal to 0.

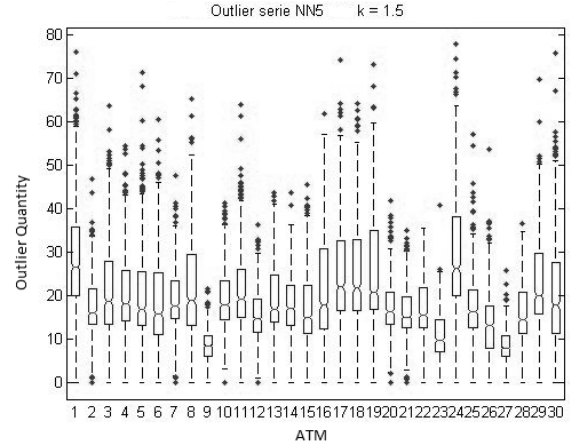


Figure 1. Identification of outliers by the Boxplot method. (*): Outlier data. In the X axis 30 ATM are shown while in the Y-axis the amount earned by each ATM can be observed. All Outlier along the series were replaced by cubic spline interpolation.

To detect outliers the boxplot method by quartiles is used with $k = 1.5$ where the lowest quartile is $Q1 = x_f$, with the f -th ordered observation. f is defined as (Eq. 1):

$$f = \frac{\frac{n+1}{2} + 1}{2} \quad (1)$$

If f involves a fraction, $Q1$ is the average of x_f and x_{f+1} . For obtaining $Q3$, the f observations are counted from the beginning, e.g. $Q3 = x_{n+1-f}$ [11]. Then an outlier is one that meets the following condition (Eq. 2):

$$\begin{aligned} x &< Q1 - k(Q3 - Q1) \\ x &> Q3 + k(Q3 - Q1) \end{aligned} \quad (2)$$

On the other hand, once each of the anomalous data is identified (a total of 870) it is replaced by cubic spline interpolation with a polynomial form $P(x) = ax^3 + bx^2 + cx + d$. Figure 1 shows the amount of outlier identified by each ATM [13].

IV. MODEL IDENTIFICATION METHODS

A. NARX

A non-linear regressive model with exogenous input (NARX) is the extension of an ARX model and is given by equation (3) [17]:

$$y_{(t)} = f(y_{t-1}, \dots, y_{t-n}, u_{t-1}, \dots, u_{t-m}) + e_t \quad (3)$$

where $e(k)$, is the prediction error at time k and is modeled as a Gaussian white noise zero mean process with variance σ . It represents the model uncertainty and the noise of the experimental data. The predictor associated with this type of models is given by equation (4) and is outlined in Figure 2, where \hat{y} is the prediction of the autoregressive variable y from previous experimental data of itself and of the exogenous variable u in times $t-1$, $t-2$, ... considering the nonlinear function \square [17].

$$\hat{y}_{(t)} = \psi(y_{t-1}, \dots, y_{t-n}, u_{t-1}, \dots, u_{t-m}) \quad (4)$$

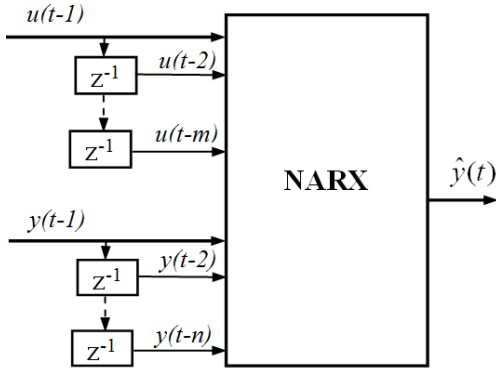


Figure 2. Associated predictor for NARX-type models. u is the exogenous input and y the autoregressive variable. \hat{y} corresponds to the prediction of the autoregressive variable.

From Figure 2 it can be seen that for identification of such models the simple series-parallel identification method can be used. Indeed, it is enough to provide the following to the chosen approximator of the prediction function -ANN or SVM in the case of this work.

As input:

- experimental data for the autoregressive variables from $t-1$ to $t-n$ (n has to do with the order of the system model).
- experimental data of the exogenous variables from $t-1$ to $t-m$.

As output:

- experimental data of autoregressive variables at a later time, t .

In the case of ANN the required training algorithm consists of the well known backpropagation. In the case of SVM it is enough to provide adequate input and output to the optimization method used to train SVM for regression [5].

B. NARMAX

A non-linear auto regressive moving average model with exogenous input (NARMAX) is the extension of an ARMAX model and is given by equation (5) [17].

$$y_{(t)} = f(y_{t-1}, \dots, y_{t-n}, u_{t-1}, \dots, u_{t-m}, e_{t-1}, \dots, e_{t-p}) + e_t \quad (5)$$

where $e(k)$ correspond to the random variable described above for Eq. (1). This term correspond to the prediction error from $t-1$ to $t-p$ (e.g. $e(k-1) = y(k-1) - \hat{y}(k-1)$). The predictor associated with this kind of model is given by Equation 6 [17] and is outlined in Figure 3:

$$\hat{y}_{(t)} = \psi(y_{t-1}, \dots, y_{t-n}, u_{t-1}, \dots, u_{t-m}, e_{t-1}, \dots, e_{t-p}) \quad (6)$$

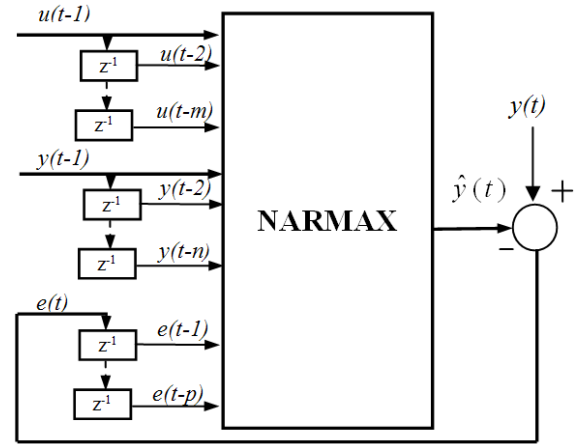


Figure 3. Associated predictor for NARMAX-type models. u is the exogenous input, y the autoregressive variable and e the predictive error. \hat{y} corresponds to the prediction of the autoregressive variable.

To identify this kind of model it is necessary to have the values of previous prediction errors. The predictor has then to be used during training in order to obtain those previous prediction errors. This makes the identification, also known as parallel method of system identification, much more complex to perform than the series-parallel identification previously presented for NARX-type models.

In the case of ANN the required training algorithm consists now of the backpropagation-through-time method [18]. However, as explained in the introduction, there are no clearly described methods to implement the NARMAX model with SVM.

C. Proposed method to implement NARMAX with SVM

In this work the following methodology is proposed:

- Divide the data set into three sets: training, validation and test.
- First identify a NARX model, as described in Section II-A, using the training set. Get the OSA prediction error (one-step-ahead) for NARX model using the validation data set.
- Use the prediction error as an additional input of a second NARX model trained on the validation set. As this second model includes the previous prediction error as input information it becomes a NARMAX type model.
- The generalization capability of this NARMAX model is evaluated on the test set, for multiple-step-ahead predictions, also known as a Model Predictive Output (MPO) configuration. Here, the model is evaluated when predicting the cash demand for the next 100 days.

D. Performance Indices

Two indices were used to quantify the prediction performance of the models.

The index of agreement (IA) developed by Willmott [19] as a standardized measure of the degree of model prediction error and varies between 0 and 1. A value of 1 indicates a perfect match, and 0 indicates no agreement at all. It is a dimensionless statistics and its value should be evaluated based on (a) the studied phenomenon, (b) measurement accuracy and (c) the model employed. IA becomes intuitively meaningful after repeated use in a variety of problems. (Eq. 7)

$$IA = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|y_i| + |\hat{y}_i|)^2} \quad \text{with} \quad \begin{aligned} y'_i &= y_i - y_m \\ \hat{y}'_i &= \hat{y}_i - y_m \end{aligned} \quad (7)$$

The symmetric mean absolute percentage error (SMAPE) is employed in the NN5 competition as the criteria to determine the winner [15][20]. Is an average of the absolute percent errors but these errors are computed using a denominator representing the average of the forecast and observed values. SMAPE has an upper limit of 200% and offers a well designed range to judge the level of accuracy. It should be less influenced by extreme values [21]. (Eq. 8)

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} * 100 \quad (8)$$

V. SIMULATIONS

A. ANN training

The multilayer perceptron (MLP) ANN was used with hyperbolic tangent activation function and Levenberg-

Marquard method for minimization of mean-square error criteria. The NNSYSID Toolbox for Matlab [22] was used for training NARX and NARMAX models using ANN. To determine the amount of autoregressors a Lipschitz function is applied [22] founding that 4 delays are needed for the output variable (cash demand (y)). One delay was considered for the exogenous variables and also one delay was chosen for the error for NARMAX models. Data were normalized in the range [0 1] and 150 different trainings for each model beginning with random weights were performed in order to avoid local minima. The training process was based on IA maximization for MPO predictions on the validation set.

Several configurations were tested including different number of neurons in the hidden layer. The following variables were considered as exogenous inputs: the day of the month (u1), day of week (u2), week (u3), month (u4) and a dummy variable (u5) to indicate special dates such as month-end, holidays and other calendar effects of interest. The final architectures for ANN NARX and NARMAX models are summarized in Table 1.

TABLE I. ANN ARCHITECTURES FOR NARX AND NARMAX MODELS

Model	Number of neurons (input-hidden-output)	Input variables	Output
NARX	9-4-1	$y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, u_{1,t-1}, u_{2,t-1}, u_{3,t-1}, u_{4,t-1}, u_{5,t-1}$	y_t
NARMAX	10-4-1	$y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, u_{1,t-1}, u_{2,t-1}, u_{3,t-1}, u_{4,t-1}, u_{5,t-1}, e_{t-1}$	y_t

B. SVM training

The library of tools developed at INSA de Rouen, France [23] was used to implement the SVM models. As mentioned above, the same sets of training, validation and testing were used for both ANN and SVM models. In this case, model training consisted in adjusting the parameters C, sigma and ϵ . In addition, λ , a parameter of the quadratic programming optimization method of SVM was also tuned. The same inputs than those used for ANN were considered to train and validate the SVM models. The training process was based on IA maximization for MPO predictions on the validation set.

SVM models were developed considering a radial basis function kernel. For NARX, the parameter C was increased by powers of 2 [24]. The exponents took values from -5 to 15 with step 1 and a more detailed search was carried out near the best values of C for each ATM, decreasing the step up to 0.1. The same method was used to find the values of sigma and λ , ranging from -5 to 5 in the first case and from -32 to -7 in the second case. In the first case, the step was decreasing the same way than C, from 1 to 0.1, while for λ from 4 to 0.25. Parameter ϵ varied between 0 and 0.2 with an initial step of 0.05 and finishing with 0.01.

For NARMAX the methodology described in section IV-C was applied. The NARX model previously trained was used for obtaining the OSA prediction error. This error was introduced

as an additional input of a second NARX model. For the second NARX the SVM was trained to find the values of C , σ and λ in the neighborhood of the values obtained for the first NARX model in a range of ± 1.5 with a step of 0.1. ϵ ranged from 0 to 0.2.

VI. RESULTS

Four models were tested when used as MPO predictors of cash demand for ATMs with a horizon of 100 days. The models are NARX and NARMAX type each one implemented with both ANN and SVM.

The MPO predictor for NARX models is described in Eq. 9.

$$\hat{y}_{MPO}(k) = f[\hat{y}(k-1), \dots, \hat{y}(k-4), u1(k-1), \dots, u5(k-1)] \quad (9)$$

While the MPO predictor for NARMAX models is described in Eq. 10:

$$\hat{y}_{MPO}(k) = f[\hat{y}(k-1), \dots, \hat{y}(k-4), u1(k-1), \dots, u5(k-1), 0] \quad (10)$$

For MPO prediction the error variable is fixed to 0 because it is impossible to *a priori* know future values of the error [25].

In Figure 4 a comparison of the IA indices between NARX and NARMAX ANN models for all the 30 ATMs can be seen. Solid and dashed lines are only used for better visualization of the results.

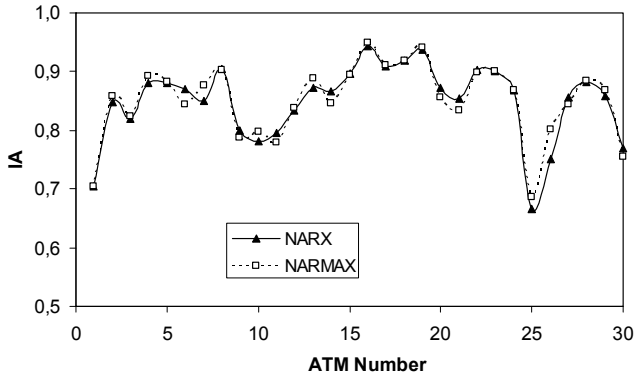


Figure 4. IA for the 30 ATMs when using ANN NARX and NARMAX type models as MPO predictors for an horizon of 100 days

For SVM, the same indices are depicted in Figure 5

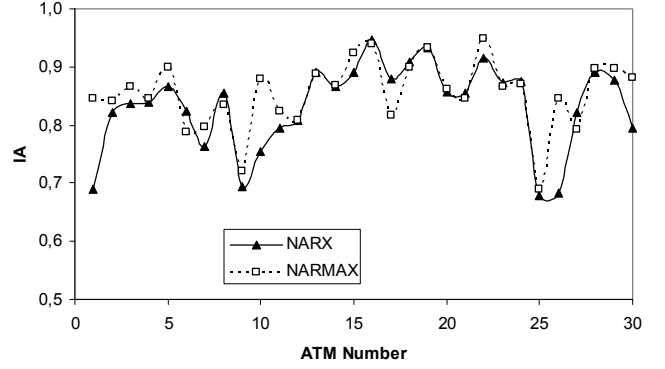


Figure 5. IA for the 30 ATMs when using SVM NARX and NARMAX type models as MPO predictors for an horizon of 100 days

TABLE II. MEAN OF QUALITY INDICES FOR MPO PREDICTIONS (100 DAYS) FOR THE 30 CONSIDERED ATMS

Model	Quality Indices	
	IA	SMAPE
	Mean \pm Std	Mean \pm Std
NARX-ANN	0,85 \pm 0,07	20,05 \pm 7,94
NARMAX-ANN	0,85 \pm 0,06	24,09 \pm 8,36
NARX-SVM	0,82 \pm 0,07	25,83 \pm 7,93
NARMAX-SVM	0,85 \pm 0,06	24,30 \pm 6,61

VII. CONCLUSIONS

From the analysis of the results shown in Figures 4 and 5 and in Table II it can be established that:

- For ANN (MLP) there are no significant differences between NARX and NARMAX models.
- Concerning SVM, NARMAX slightly improved the results obtained by NARX models at least allowing to obtain a mean IA equal to those obtained by ANN. Nevertheless only in the case of 4 ATMs (1, 10, 26, 30) NARMAX significantly improved the performance of NARX (from $IA=0.71 \pm 0.04$ to $IA=0.86 \pm 0.02$).

Unexpectedly, for this application, the predictive capability of NARMAX models did not outperform the NARX capability. SVM also showed no superiority over ANN. Moreover given that the best results were exhibited by NARX ANN models - which is even more remarkable considering the SMAPE index- and given the simplicity exhibited by ANN in their training - significantly lower computational effort to find weights than to calibrate parameters of SVM- it seems advisable to choose simpler models, such as NARX and a user-friendly tool like ANN at least for this particular application.

However it seems important to underline that the proposed method to implement NARMAX with SVM was easy to implement and allowed the SVM to predict in a MPO configuration. This is why we consider that the results of this work are encouraging to study this method and apply it to other model identification problems.

ACKNOWLEDGMENT

G.A. thanks partial funding of this work by Fondecyt project 1090316 and DICYT-USACH. M.C. thanks Dirección de Investigación Universidad de la Frontera DIUFRO.

REFERENCES

- [1] Box, G. and Jenkins, G., 1970, Time series analysis: Forecasting and control, San Francisco: Holden-Day.
- [2] Hornstein, A. and Parlitz, U., 2004, Bias Reduction for Time Series Models Based on Support Vector Regression. *International Journal of Bifurcation and Chaos*, 14(6), pp. 1947-1956
- [3] Chai, M-I, S., Song, S., and Li, N-n, 2005, A review of some main improved models for neural network forecasting in time series. In: *Proceedings of the IEEE Intelligent Vehicles Symposium*, Issue 6-8, pp. 866--868
- [4] Woolley, J., Agarwal, P., Baker, J., 2010, Modeling and prediction of chaotic systems with artificial neural networks. *Internacional Journal for Numerical Methods in Fluids*, 63, pp. 989-1004.
- [5] Schölkopf, B., Smola, A., Williamson, R.C., Bartlett, P.L., 2000, New support vector algorithms, *Neural Computation*, 12, pp. 1083--121
- [6] Sapankevych, N and Sankar, R, 2009, Time series prediction using support vector machines: a survey, *IEEE Computational Intelligence Magazine*, May 2009, pp. 24-38.
- [7] Suykens, J.A.K., Van Gestel, T, De Brabanter, J., De Moor, B., Vandewalle, J. 2002, Least Squares Support Vector Machines, *World Scientific*, Singapore.
- [8] Martínez-Ramón, M., Rojo-Alvarez, J., Camps-Valls, G., Muñoz-Mari, J., Navia-Vásquez, A., Soria-Olivas, E, Figueiras-Vidal, A, 2006, Support vector machines for nonlinear kernel ARMA system identification, *IEEE Transactions on Neural Networks*, 17(6), pp. 1617-1622.
- [9] Guo, G., Wu, X., Zhuo, S., Xu, P., Cao, B., 2008, Prediction state of charge of Ni-MH battery pack using support vector machines for hybrid electric vehicles, *IEEE Vehicle Power and Propulsion Conference (VPPC)*, September3-5, 2008, Harbin, China
- [10] Yingmin, L. and Yinfeng, D., (2006), Support Vector Machines for Structure Nonlinear Response Simulation, *4th International Conference on Earthquake Engineering*, Taipei, Taiwan.
- [11] Gao, Y and Joo Er, M, 2005, NARMAX time series model and prediction: feedforward and recurrent fuzzy neural network approaches, *Fuzzy Sets and Systems*, 150, pp. 331-350.
- [12] Simutis, R., Dilijonas, D., Bastina, L. 2008, Cash demand forecasting for ATM using neural Networks and support vector regresión algorithms. *20th EURO Mini Conference*. Vilnius, pp 416-421.
- [13] Ramírez, C. and Acuña, G., 2011, "Forecasting cash demand in ATM using Neural Networks and Least Square Support Vector Machine", *Congreso Iberoamericano de Reconocimiento de Patrones, CIARP, Pucón, Chile, Noviembre 2011, Lecture Notes in Computer Science, LNCS 7042*, pp. 515-522.
- [14] Simutis, R., Dilijonas, D., Bastina L., Friman, J., Drobinov, P.: Optimization of Cash Management for ATM Network. *ISSN 1392 – 124x Information technology and control*, Vol.36, No.1A (2007)
- [15] NN5, "Forecasting Competition for Artificial Neural Networks & Computational Intelligence", <http://www.neural-forecastingcompetition.com/NN5> (2008)
- [16] Wichard, J. D., 2010, Forecasting the NN5 time series with hybrid models. *Leibniz-Institut für Molekulare Pharmakologie (FMP), Molecular Modeling Group, Robert-Roessle-Str. 10, 13125 Berlin, Germany*.
- [17] Ljung, L., 1987, *System Identification: Theory for the user*. Prentice-Hall, Englewood Cliffs, NJ.
- [18] Werbos P.J., 1990, Backpropagation Through Time: What It Does and How to do it. *Proc. IEEE*, Vol 78(10), pp 1150-1560.
- [19] Willmott, C. J. 1981. On the validation of models. *Physical Geography*, 2, 184-194. (1981)
- [20] Garcia-Pedrero, A., and Gomez-Gil, P., 2010, Time Series Forecasting using Recurrent Neural Networks and Wavelet Reconstructed Signals. *Proceedings of the 20th. International Conference on Electronics, Communications and Computers*. Puebla, Mexico.
- [21] Swanson, E., A., Tayman, J., Bryan, T. M.: MAPE-R: A rescaled measure of accuracy for Cross-Sectional Forecasts. *Department of Sociology. University of California Riverside. Riverside USA. Department of Economics, University of California. San Diego USA. McKibben Demographic Research. Rock Hill, South Carolina 29732*.
- [22] Norgaard, M., 2003, The NNSYSID Toolbox, for use with Matlab, <http://www.iau.dtu.dk/research/control/nnsysid.html>
- [23] Canu, S., Grandvalet, Y., Guigue, V., Rakotomamonjy, A, 2005, SVM and Kernel Methods Matlab Toolbox. *Perception Systèmes et Information*, INSA de Rouen, Rouen, France.
- [24] Frohlich, H., Zell, A., 2005, Efficient parameter selection for support vector machines in classification and regression via model-based global optimization. In: *Proceedings of the IEEE International Joint Conference on Neural Networks, IJCNN 2005*. pp. 1431-1436
- [25] Chen, P-H., Lee, Y-W, Chang, T-L, 2009. Predicting thermal instability in a closed loop pulsating heat pipe system. *Applied Thermal Engineering* 29 (8-9), pp 1566-1576.