5th International Conference on Experiments/Process/System Modeling/Simulation/Optimization
5th IC-EpsMsO
Athens, 3-6 July, 2013

A HYBRID MODEL FOR WIND SPEED FORECASTING USING ARMA MODELS AND SUPPORT VECTOR MACHINES (SVM)

Stylianos Sp. Pappas[1]*, Georgios E. Chatzarakis[1], Christos C. Pappas[2], Vassilios C. Moussas[3]

[1]Department of Electrical Engineering Eduacators Aspete – School of Pedagogical and Technological Education N. Heraklion, 141 21 Athens, Greece.

[2]University of Greenwich, School of Engineering, Medway Campus, Pembroke Building, Central Avenue, Chatham Maritime, Kent ME4 4TB.

[3]School of Technological Applications, Technological Educational Institute of Athens, 122 Ag. Spyridonas St., Egaleo 12210, Greece
*e-mail: spappas@aegean.gr, ksiroval@otenet.gr

Keywords: Instructions Wind power, time series prediction, ARMA, Support Vector Machine, Neural Networks, Multi-Model Partition Algorithm.

Abstract. *An alternative electric power source, such as wind power, has to be both reliable and autonomous. An accurate wind speed forecasting method plays the key role in achieving the aforementioned properties and also is a valuable tool in overcoming a variety of economic and technical problems connected to wind power production.*
As it is known ARMA (AutoRegressive Moving Average) models have been widely used for linear time series forecasting. One of their major disadvantages is the difficulty they have in identifying the non linear characteristics of the data. Recenlty, another Neural Network (NN) architecture called Support Vector Machine (SVM), was introduced and successfully applied in predicting the behaviour of non linear time series.
The aim of this work is to combine the benefits of both methods and apply them in order to achieve a reliable wind speed forecasting hybrid method. The ARMA model identification and parameter estimation was accomplished using the Multi-Model Partition Algorithm (MMPA). The method proposed is based on the reformulation of the problem in the standard state space form and on implementing a bank of Kalman filters (KF), each fitting an ARMA model of different order.
Real data were used and real cases were tested based on the measurements of the wind speed provided by Vestas Hellas®. The parameters of the wind speed forecasting problem are complex and unique, however their appropriate modelling can lead to very promising results.

1 INTRODUCTION

Energy is considered amongst the most significant factors that are closely related to both economic and social development. It is also a fact that nowadays the majority of the electrical energy production is based on the fossil fuels, which on one hand, are without any doubt, highly efficient but on the other are responsible for the emission of greenhouse gases and their reserves are limited.

Consequently renewable sources of energy, such as wind, biomass, solar power, wave power etc, have been already adopted for electric power production. It is well known that the wind power generation raises issues of reliability due to the fact that the wind speed is significantly and directly affected by various factors such as the type of the terrain, the height, season of the year, atmospheric conditions, obstacles present and many more. This leads to the conclusion that unless the reliability of the wind power generation is at an acceptable level, wind power is not eligible for constant electrical energy supply to the power system. [1, 2].

Recent studies have shown that combined forecasting methods can offer robust solutions and can be efficiently implemented to various real life problems in diverging fields such as chemical processes, economics, load forecasting, tourism demand, environmental issues, medicine and many more [3, 4, 5, 6, 7].

In this study a hybrid model is presented, that reveals the advantages of an ARMA and SVM model in wind speed modelling and prediction problem. Initially successful model identification and parameter estimation has to be performed in order to choose the most appropriate ARMA models. For tackling this task the well established MMPA was used. This approach was introduced by Lainiotis [8-9] and summarizes the parametric model uncertainty into an unknown, finite dimensional parameter vector whose values are assumed to lie within a

known set of finite cardinality. A non-exhaustive list of the reformulation, extension and application of the MMPA approach as well as its application to a variety of problems can be found in [10-15].

In this research real data were used provided by Vestas Hellas® the simulation results appear to be very promising.

2 HYBRID MODEL PRESENTATION

2.1 ARMA Model

The problem of fitting an ARMA model in a given time series is present for more than half a century and is still appearing in many different fields such as in remote monitoring of civil infrastructure, predicting the demand for auto spare parts in China due to the fierce market competition, forecasting equipment failures in order to adjust maintenance policies in manufacturing plants, estimating retail sales volumes, predicting the outbreak and development of animal infectious diseases and many more [16-17].

Considering the general case an m-variate (i.e. multivariate) ARMA model of order (p, q) [ARMA (p, q)] for a stationary time series of vectors y observed at equally spaced instants k = 1, 2, …,n is defined as:

$$\mathbf{y}_k = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{k-i} + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{v}_{k-j} + \mathbf{v}_{k,} \qquad \mathrm{E}[\mathbf{v}_k \mathbf{v}_k^{\,T}] = \mathbf{R} \tag{1}$$

where the m-dimensional vector v_k is uncorrelated random noise, not necessarily Gaussian, with zero mean and covariance matrix R, $\theta = (p, q)$ is the order of the predictor and $A_1,\ldots,A_p$, $B_1,\ldots,B_q$ are the m x m coefficient matrices of the multivariate (MV) ARMA model.

It is obvious that the problem requires both the predictor's order $\theta = (p, q)$ determination and computation of the predictor's matrix coefficients $\{A_i, B_j\}$.

The major disadvantage of the ARMA models is that their performance can be limited by any significant data non-linearities.

2.2 Multimodel Partition Algorithm (*MMPA*)

Due to the fact that the wind speed has not a constant or periodic behaviour, it was noted, by trial and error, that not a single ARMA model that was able to describe the whole data set satisfactory. It is actually the combination of various ARMA models, each one used for different time intervals and applied for different time duration that describes in the best manner the existing data. So instead of having various ARMA models of different order θ running in parallel with the SVM it was decided to load all the data to an adaptive filter programmed with the MMPA, and it will be the job of that filter to decide which ARMA model will be used each time. The work presented in this paper is an extension and combination of the MMPA with SVM, so an analytical presentation of the MMPA and all the associated equations are presented in [18].

2.3 Suuport Vector Machines (SVM)

In support vector machines, as they were proposed in [19], the training data set $x_i \in R^d, i = 1,\ldots,N$ is mapped into a higher dimensional feature space, via an operator Φ.

A mathematical representation of the SVM function is:

$$y = \omega \cdot \Phi(x) + b \tag{2}$$

where ω and b can be found by the minimization of the following equations:

$$S(C) = C\frac{1}{M}\sum_{j=1}^{M} Lf_e(h_j - y_j) + \frac{1}{2}\|\omega\|^2, \tag{3}$$

$$Lf(h, y) = \begin{cases} |h - y| - e & |h - y| \geq e, \\ 0 & others, \end{cases} \tag{4}$$

where parameters C and e are user defined. The term h_i is the actual wind speed at the time instant j and term $Lf_e(h_i - y_i)$ is the *loss function*. By looking at equation (4) it is obvious that there is any penalty for errors below e. The width of the function is given by the term $\frac{1}{2}\|\omega\|^2$ and finally the training error term is given by $C\frac{1}{M}\sum_{j=1}^{M} Lf_e(h_i - y_i)$, where C is the trade-off between the width of the function and the minimum training

error. For dealing with non linear cases, like wind speed data, one may introduce slack variables ξ and ξ* into equation (2) such that :

$$\omega \cdot \Phi(_j) + b_j - h_j \leq e + \xi_j^* \tag{5}$$

$$-(\omega \cdot \Phi(_j) + b_j) + h_j \leq e + \xi_j \tag{6}$$

where $\xi_j, \xi_j^* \geq 0,$ and $j = 1,2,\ldots M$.

By considering the above slack variables and in order to include any extra cost of the training errors, equation (3) which represents the objective function to be minimized is rearranged to:

$$S(\omega, \xi, \xi^*) = C^* \left(\sum_{j=1}^{M} \xi_j + \xi_j^*) \right) + \frac{1}{2} \omega \omega^{\mathrm{T}} \tag{7}$$

where again C^* is user defined and is the trade off between the maximum margin defined by $\|\omega\|$ and the minimum training error as defined by $\sum_{j=1}^{M} \xi_j + \xi_j^*)$.

Finally by introducing positive Lagranian multipliers and maximizing equation (7) the latter equation is reformed to:

$$\begin{aligned} S(a_j - a_j^*) &= \sum_{j=1}^{M} h_i (a_j - a_j^*) - e \sum_{j=1}^{M} (a_j - a_j^*) \\ &- \frac{1}{2} \sum_{j=1}^{M} \sum_{i=1}^{M} (a_j - a_j^*) \times (a_i - a_i^*) K(x_j - x_i) \end{aligned} \tag{8}$$

subject to

$$\begin{aligned} &\sum_{j=1}^{N} (a_j - a_j^*) = 0, \\ &0 \leq a_{j,} a_j^* \leq C, \end{aligned} \tag{9}$$

where j = 1,2,…M .

The Lagranian multipliers a_j, a_j^*, satisfy $a_j * a_j^* = 0$ and also

$$f(x, a, a^*) = \sum_{j=1}^{l} (a_j - a_j^*) K(x - x_i) + b \tag{10}$$

The ***Kernel*** function $K(x - x_i)$ introduced in equation (10) is defined such that $K(x_j - x_i) = \Phi(x_i) \cdot \Phi(x_j)$, meaning that its value is equal to the inner product of the vectors x_i and x_j, included in the featured space $\Phi(x_i)$ *and* $\Phi(x_j)$.

In this study the Gaussian Kernel function (11), also known as Radial Basis Function (RBF), is used.

$$K(x_j - x_i) = e^{\left(\frac{-\|x_j - x_i\|^2}{2\sigma^2} \right)} \tag{11}$$

The most significant feature of the SVM compared to other similar algorithms is that they manage to achieve optimum performance by restricting the decision's function complexity so that is the most suitable according to the quantity of the data present.

2.4 The hybrid model

In The wind speed behaviour is unpredictable and it is difficult to be represented. This is the reason for combining two different techniques for modelling the linear and the non-linear part of the series. The hybrid model proposed is based on a linear pattern, *L(t)* produced by the MMPA and a non-linear one, *NL(t)* produced by the SVM. It can be represented as:

$$Q(t) = L(t) + NL(t) \tag{12}$$

Both parts are directly calculated from the wind speed time series.

If *e(t)* is the MMPA estimation error at any time instant *t*, then

$$e(t) = Z(t) - \tilde{L}(t) \tag{13}$$

It is now the SVM that models these residuals as

$$e(t) = f\,(e_{(t-1)}, e_{(t-2)},\ldots, e_{(t-n)}) + \Delta t \tag{14}$$

where f is non linear and Δt is random error. It is obvious that $\widetilde{NL}(t)$ is the forecast of (14).

Consequently the forecast of the hybrid model is

$$\tilde{Q}(t) = \tilde{L}(t) + \widetilde{NL}(t) \tag{15}$$

support vector machines, as they were proposed in [25], the training data set $x_i \in R^d, i = 1,\ldots,N$ is mapped into a higher dimensional feature space, via an operator Φ. Figure 1 shows a schematic representation of the hybrid model.

3 RESULTS

In this method the weighted average of the estimates produced by the elemental ARMA filters were used as a data pre-processor in order to detect the data's linearities. This was succeeded using a bank of 10 Kalman filters of order (*1*,*1*), (*2*,*2*), (*3*,*3*), …,(*10*,*10*) programmed with the MMPA. Then the MMPA's estimation error was applied as input to the SVM that were able to achieve a further error reduction and come up with a better forecasting outcome. As far as the SVM are concerned the three parameters (C, σ, e) had to be carefully adjusted. Unsuitable values for these parameters may lead to either over fitting or under fitting of the training data. The values used in this research were $C = 35$, $\sigma = 2.5$ and $e = 0.6$.

This research was conducted based on the hour average of daily wind speed recorded by the Vestas Hellas from December 2010 up to February 2011. The obtained time series did not follow any periodic pattern and it was also presenting irregular amplitudes, making it hard to both model and predict (Figures 2-4 "Raw" data).

The aim of this work is to generate a single-step prediction based on past observations. The data were normalized to take values from zero to one, before using them as input data to the hybrid model.

From the available data points, 744 were for December, 744 for January and 517 for February. For each month 20% of the available data was used for training, 20% for validation and 60% for testing.

The performance of the hybrid method is judged by (a) comparing the predicted and the observed (raw) wind time series, Figures 2-4, (b) drawing scatter diagrams of the predicted and the observed sequences, Figures 5-7 and (c) by computing the mean percentage absolute error (MAPE) for the testing data set, using the mathematical formula given in Eq. 16. Table 1 summarizes the results.

$$MAPE = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\left|\mathrm{P}_{real,j} - \mathrm{P}_{predicted,j}\right|}{\mathrm{P}_{real,j}}\right)\times 100 \tag{16}$$

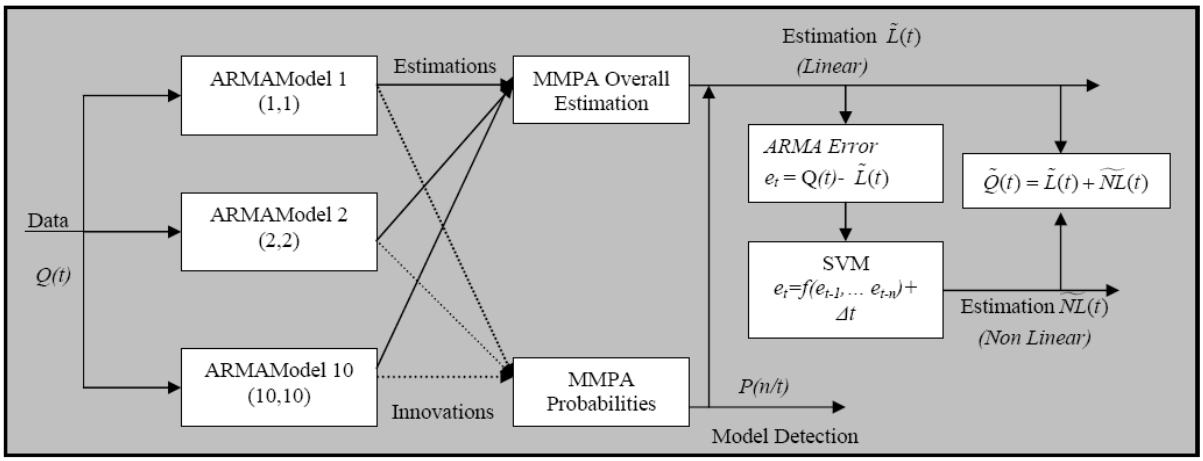

Figure 1. Schematic Representation of the Hybrid Model

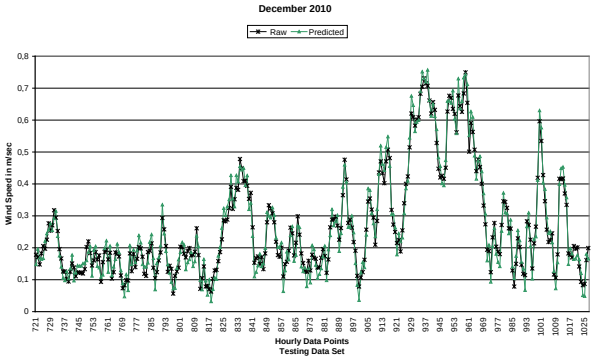

Figure 2. December 2010

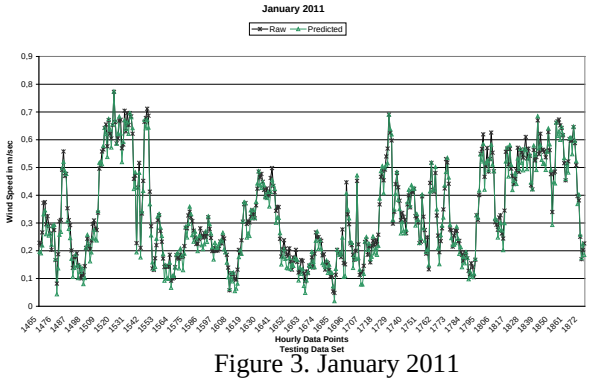

Figure 3. January 2011

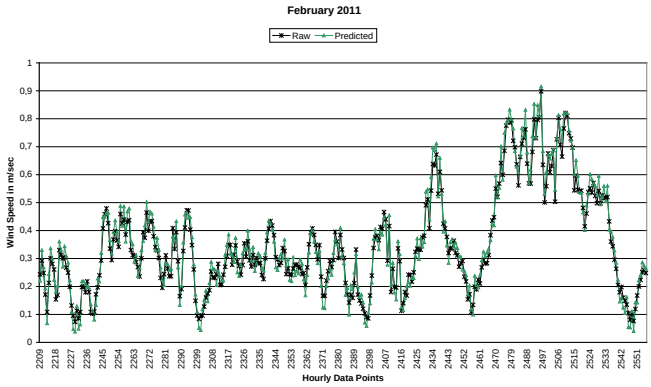

Figure 4. February 2011

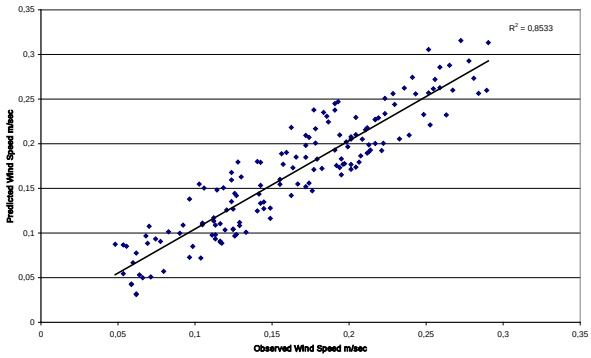

Figure 5. Dec 2010: Scatter of predicted and observed wind speed time series

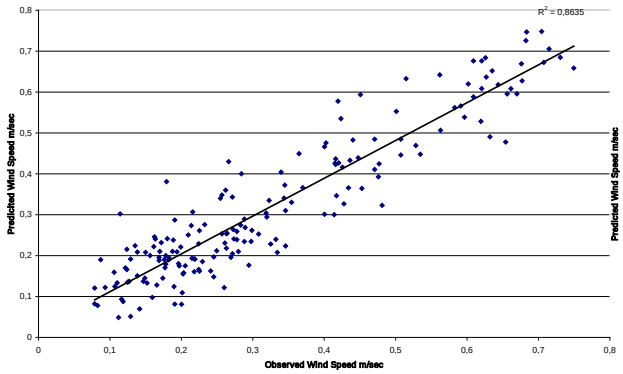

Figure 6. Jan 2011: Scatter of predicted and observed wind speed time series

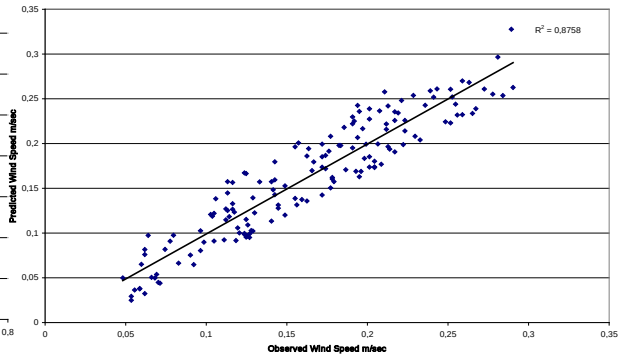

Figure 7. Jan 2011: Scatter of predicted and observed wind speed time series

Month	***MAPE %***	***R^2***
December	3.27	0.8533
January	3.19	0.8635
February	3.02	0.8758
Average	***3.01***	**0.8685**

Table 1. Example of the construction of a table

4 CONCLUSIONS

The area of forecasting is very demanding and is over than 50 years that ARMA models were exclusively used for tackling real life problems. Recently ANN were applied in difficult prediction problems showing very satisfactory results especially due to their ability of manipulating the non linearities of the data set. The aim of this work was not to just add yet another technique of wind speed prediction but to actually validate the fact that different forecasting methods fulfil each other and lead to accurate results. As it was shown the performance of the hybrid method proposed was satisfactory with an average error of 3.01% which can be considered quite small for wind speed forecasting. Future work can include adjustments for wind speed prediction for time intervals smaller than 1 hour, say ten minutes intervals and also for on-line wind speed prediction.

REFERENCES

[1] Lei, M, Shiyan, L., Chuanwen, J., Hongling, L. and Zhang, Y (2009)., “A review on the forecasting of wind speed and generated power”, Renewable and Sustainable Energy Reviews, Vol. 13 pp. 915-920.
[2] Marciukaitis, M., Katinas, V., Kavaliauskas A. (2008), “Wind power usage and prediction prospects in Lithuania”, Renewable and Sustainable Energy Reviews, 12, pp. 265-277.
[3] Chien-Hsien, W., Vassiliadis, C.A. (1998), “Applying hybrid artificial intelligence techniques in wastewater treatment”, Engineering Applications of Artificial Intelligence, 11(6), pp. 685-705.
[4] Shafie-khah, M. Parsa Moghaddam, M. Sheikh-El-Eslami, M.K. (2011), “Price forecasting of day-ahead electricity markets using a hybrid forecast method”, Energy Conversion and Management, 52(5), pp. 2165-2169.
[5] Kuan-Yu Chen (2011), “Combining linear and nonlinear model in forecasting tourism demand”, Expert Systems with Applications, 38(8), pp. 10368-10376.
[6] Bruno Baruque, Emilio Corchado, Aitor Mata, Juan M. Corchado. (2010), “A forecasting solution to the oil spill problem based on a hybrid intelligent system”, Information Sciences, 180(10), pp. 2029-2043.
[7] Wei-Chang Yeh, Wei-Wen Chang, Yuk Ying Chung. (2009), “A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method”, Expert Systems with Applications, 36(4), pp. 8204-8211.
[8] Lainiotis DG. (1976), “Partitioning: A unifying framework for adaptive systems I: Estimation”. Proc. of the IEEE, 64(8), pp. 1126-1143.
[9] Lainiotis DG (1976), “Partitioning: A unifying framework for adaptive systems II: Control”. Proc. of the IEEE, 64(8), pp. 1182-1198.
[10] Pappas S.Sp, Leros AK, Katsikas SK. (2006), ”Joint order and parameter estimation of multivariate autoregressive models using multi-model partitioning theory”, Digital Signal Processing, 16(6), pp.782-795.
[11] Lainiotis DG, Papaparaskeva P. (1998), “A partitioned adaptive approach to nonlinear channel equalization”, IEEE Trans on Communications 46(10), pp. 1325-1336.
[12] Moussas VC, Likothanassis SD, Katsikas SK, Leros AK. (2005), “Adaptive on-line multiple source detection”, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, (ICASSP’05), pp.1029-1032.
[13] Nikitakos NV, Leros AK, Katsikas SK. (1998), “Towed array shape estimation using multimodel partitioning filters”, IEEE J of Oceanic Engineering, 23(4), pp. 380-384.
[14] S.Sp. Pappas, L. Ekonomou, P. Karampelas, S.K. Katsikas, P. Liatsis. (2008), “Modeling of the Grounding Resistance Variation using ARMA Models”, International Journal of Simulation Modelling Practice and Theory, 16(5), pp. 560-570.
[15] Pappas S.Sp., Ekonomou L., Karamousantas D.Ch., Chatzarakis G.E., Katsikas S.K., Liatsis P. (2008), “Electricity Demand Loads Modeling using AutoRegressive Moving Average (ARMA) models”, Journal of Energy, 33(9), pp. 1353-1360.
[16] E. P. Carden and M. W. Brownjohn. (2008), “ARMA modelled time-series classification for structural health monitoring of civil infrastructure,” Mechanical Systems and Signal Processing, vol. 22, Issue 2, pp. 295-314.
[17] Yun Chen; Heng Zhao; Li Yu. (2010), “Demand Forecasting in Automotive Aftermarket Based on ARMA Model,” International Conference on Management and Service Science (MASS), pp. 1-4.
[18] Pappas, S.S.; Harkiolakis, N.; Karampelas, P.; Ekonomou, L.; Katsikas, S.K. (2008), “A New Algorithm for On-Line Multivariate ARMA Identification Using Multimodel Partitioning Theory,” Panhellenic Conference on Informatics, PCI '08, pp. 222 – 226.
[19] V. Vapnik, (1995), The nature of statistic learning theory, Springer-Verlag, New York, NY.