Prediction of the long-term electrical energy consumption in Greece using adaptive algorithms

STYLIANOS SP. PAPPAS, SPYRIDON ADAM

Department of Electrical and Electronic Engineering Educators,
A.S.PE.T.E. - School of Pedagogical and Technological Education,
N. Heraklion, 141 21 Athens, GREECE

spappas@aegean.gr, spadam@aspete.gr

Abstract: - Planning the electricity grid is an ahead-looking process that requires long term prediction for a time interval greater than one year. The importance of accurate of long-term load forecasting cannot be overlooked, since it provides the future load demand; a crucial factor that is considered in scheduling the generation, transmission and distribution of the electrical energy, reliably and economically. In this study real data is used and the performance of the combination of the well-established multimodel partitioning filter (MMPF) implementing extended Kalman filters (EKF) with Support Vector Machines (SVM), is compared to the one of an artificial multilayer layer feed-forward neural network (ANN). The results indicate that both methods are reliable, however the combination of MMPF and SVM provides a more accurate long-term load forecasting. The proposed method is a useful tool since the electric system administrator based on its forecasts will able to use efficiently the current resources in order to meet the forecasted demand using a least-cost plan.

Key-Words: - Artificial neural networks, energy consumption, gross domestic product, extended kalman filters, multi model partitioning filter, support vector machines, installed capacity.

1 Introduction

Forecasting models are widely applied in many areas. For example in economics they are used in real time GDP forecast, using mixed frequency data that includes industrial production, employment, private consumption and exports, [1] and in long-term GDP per capita growth applying the S-shaped logistic pattern [2].

In [3] parametric and non-parametric models are tested and compared upon their forecasting power on implied volatility indices. Another application concerning financial time series forecasting is called the random walk dilemma and a possible solution is a novel method called increasing decreasing linear neuron as presented in [4].

In the area of business large international data sets are used to analyze whether business cycle forecasts herd or antiherd with strong indications for the second [5]. Additionally in [6], the authors propose a new modelling methodology for forecasting the spare part demand for electronic commodities in the spare parts logistics services.

Another active research area is medicine and concerns the public health planning, especially during outbreaks via real time forecasts of infectious diseases [7].

In agricultural a combination of short term weather forecasts are applied in order to accurately predict certain factors, such as maximum and minimum daily temperatures, precipitation and radiation, whose knowledge greatly benefit cost effective decision making [8].

One more application concerns aviation and more specifically the improvement of the accuracy of airport weather forecasts by learning from the relationships between previously modelled and observed data, based on a new machine learning methodology [9].

Wind speed forecasting is another important active research area. In [10] a novel hybrid approach is presented, able of tackling the problem successfully.

Finally new efficient methods for short term or long term electricity demand load prediction are presented in [11-16] using ARMA, ARIMA and SARIMAX models and in [17-23] by applying artificial neural networks, support vector machines and hybrid methods.

It is well known that the electrical energy consumption has been significantly increased worldwide due to industrial development and also due to economic and population growth.

Therefore it is essential for a country's power system to include a careful and strategic planning in terms of equipment and facilities expansion, in order to meet its customer's present and future electric demand reliably.

An accurate long-term energy consumption forecasting is essential because it can lead to a successful schedule as far as the maintenance and development of the existing power generation plants is concerned and also to a low cost economic plan, in terms of buying new equipment, investing to the appropriate electrical power production technologies (either renewable or non-renewable), expanding and modernize the existing distribution network.

The problem with long-term forecasting is that is being influenced by a number of factors, such as the overall existing capacity, the average ambient temperature, humidity, the energy consumption per person, the gross domestic product (GDP) and many more.

In Greece the economic recession has led to a 20% decrease of GDP and at the same time to an increase of 20% in the price of oil. This had as a result the increase of the electric demand, especially for heating during the winter, since the KWh price was more economical [24]. Finally the increase of the highest temperature during the summer months has led to extensive use of air conditions and other cooling devices increasing further the load demand [25-27].

The power generation network was able to respond and cover the extra load demand mainly due to the increased power production using natural gas and renewable sources and also through imports of electrical energy [28].

Figures 1 and 2 indicate the above mentioned features.

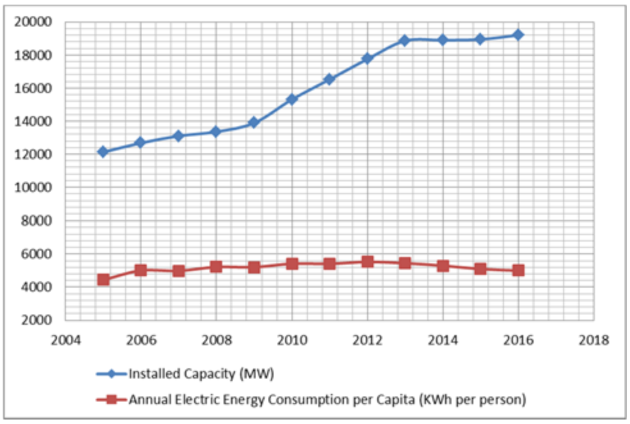


Figure. 1: The Greek installed capacity and Annual Electric Energy per person from 2004 – 2018.

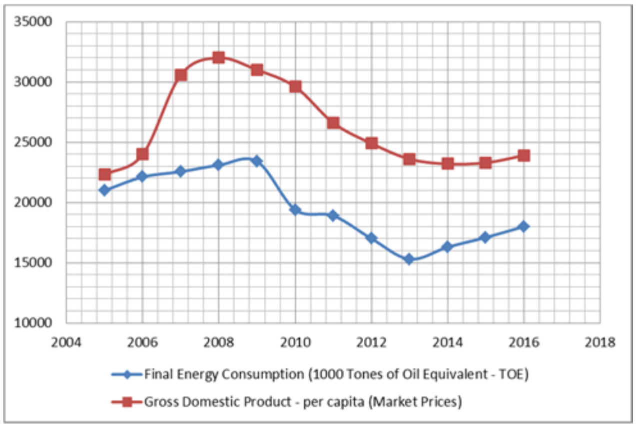


Figure 2: The Greek final energy consumption and gross domestic product from 2004 – 2018.

2 Adaptive techniques

2.1 The MMPF method

The multi model partitioning filter (MMPF) was initially proposed and presented by Lainiotis in [29-31] and since then it has been successfully applied in a numerous applications such as the modelling of the grounding resistance variation [32], order and parameter estimation of multivariate (MV) ARMA models [33-34], electric load modelling and forecasting [11-13], wind speed prediction combined with support vector machines (SVM) [10], network anomaly detection [35], multiple source detection [36], towed array shape estimation [37] and finally combined with genetic algorithms for data mining [21, 38-39].

The method proposed is based on a hybrid model that combines the adaptive MMPF [11-13], known for its stability, with SVM. The idea of combing these two methods for long-term electric load forecasting came from the fact that it was successfully applied for wind speed prediction [10], as well as mid-term and short-term load forecasting [21]. The data used now is not subjected to any prior – offline manipulation in order to remove weekly and annual seasonality as was done in previous cases [13]. To tackle this problem the MMPF implements a bank of extended Kalman filters (EKF) with ARMA models instead of simple Kalman filters (KF) with ARMA models in order to handle data's non-linearities. MMPF with EKF combined with genetic algorithms (GA) were successfully applied in prediction of epilepsy and in the evolution of stock values using biomedical and financial data respectively [39].

The method is analytically presented in [10] and [21] but a brief description will be also presented here. Figure. 3 represents a block diagram of the adaptive method proposed.

After several trials it was noted that only one ARMA model is not able to accurately describe the input data series. However if someone combines several ARMA models of different order θ, for different time periods and for different time duration, then the existing data will be satisfactorily described.

Therefore instead of having several ARMA models, of different order θ running in parallel with the SVM, the authors decided to load the data into a single MMPF. The job of this filter is to decide which ARMA model is suitable each time,

Let's assume that the order of the ARMA model that fits the data is $\theta = (p.q)$ then our problem in state-space form can be written as:

$$\mathbf{x}(k+1) = \mathbf{x}(k) \tag{1}$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k) \tag{2}$$

Now a new variable λ is assigned as λ=max (p, q). Then $\mathbf{x}(k)$ is an $m^2(\lambda+\lambda) \times 1$ vector made up from the coefficients of the matrices $\{\mathbf{A}_1, ..., \mathbf{A}_\lambda, \mathbf{B}_1, ..., \mathbf{B}_\lambda,\}$, $\mathbf{H}(k)$ is an $m \times m^2 (\lambda+\lambda)$ observation history matrix of the process $\{\mathbf{y}(k)\}$ up to time k-($\lambda+\lambda$).

If the general form of the matrices $\mathbf{A}_\lambda$ and $\mathbf{B}_\lambda$ is respectively:

$$\begin{bmatrix} a_{11}^{\lambda} & \dots & a_{1m}^{\lambda} \\ \vdots & \ddots & \vdots \\ a_{m1}^{\lambda} & \cdots & a_{mm}^{\lambda} \end{bmatrix}, \begin{bmatrix} b_{11}^{\lambda} & \dots & b_{1m}^{\lambda} \\ \vdots & \ddots & \vdots \\ b_{m1}^{\lambda} & \cdots & b_{mm}^{\lambda} \end{bmatrix} \tag{3}$$

then

$$\begin{aligned} \mathbf{x}(k) \triangleq [&\alpha_{11}^{1}\,\alpha_{21}^{1}\cdots\alpha_{m1}^{1}\vdots\cdots\vdots\cdots\alpha_{mm}^{1}\vdots\cdots\alpha_{mm}^{\lambda}\vdots \\ &b_{11}^{1}\,b_{21}^{1}\cdots b_{m1}^{1}\vdots\cdots\vdots\cdots b_{mm}^{1}\vdots\cdots b_{mm}^{\lambda}]^{T} \end{aligned} \tag{4}$$

If $p>q$ then $\lambda=p$, the last (m^2 (p-q)) MA coefficients are zero. If $q>p$ then $\lambda=q$, the last (m^2 (q-p)) AR coefficients are zero.

$$\begin{aligned} \mathbf{H}(k) \triangleq [&y_1(k-1)\mathbf{I}\cdots y_m(k-1)\mathbf{I}\vdots\cdots\vdots \\ &y_1(k-\lambda)\mathbf{I}\cdots y_m(k-\lambda)\mathbf{I}\vdots \\ &v_1(k-1)\mathbf{I}\cdots v_m(k-1)\mathbf{I}\vdots\cdots\vdots \\ &v_1(k-\lambda)\mathbf{I}\cdots v_m(k-\lambda)\mathbf{I}\ ] \end{aligned} \tag{5}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix.

Assuming that the system model and its statistics were completely known, the Kalman filter (KF) is the optimal estimator in the minimum variance sense.

However if the system model is not completely known the MMPF, is one of the most widely used approaches for similar problems [33-38].

In the case under consideration assume that the model uncertainty is the lack of knowledge of the model order θ. Let us further assume that the model order θ lies within a known sample space of finite cardinality, i.e. that $1 \leq \theta \leq M$, $\theta \in \mathfrak{I}$, where $\mathfrak{I}$ denotes the set of integers. The MMPF operates on the following discrete-time model:

$$\mathbf{x}(k+1) = \mathbf{F}(k+1, k/\theta)\,\mathbf{x}(k) + \mathbf{w}(k) \tag{6}$$

$$\mathbf{y}(k) = \mathbf{H}(k/\theta)\,\mathbf{x}(k) + \mathbf{v}(k) \tag{7}$$

where θ is the unknown parameter - the model order in this case- $\mathbf{F}$ is the state transition matrix and $\mathbf{w}(k)$ is independent, zero mean, white noise not necessarily Gaussian with covariance $\mathbf{Q}$ which is usually set to a small positive non zero constant. The optimal *MMSE* (Minimum Mean Square Error) estimate of $\mathbf{x}(k)$ is given by:

$$\hat{\mathbf{x}}(k/k) = \sum_{j=1}^{M} \hat{\mathbf{x}}(k/k;\theta_j)\, p(\theta_j/k) \tag{8}$$

A set of M (10 in this work) models is designed, each matching one value of the parameter vector, $\{(1,1), (2,2), \ldots (M,M)\}$. The probabilities $p(\theta_j/k)$ for each model are set to $1/M$, where M is the cardinality of the model set. The number of the models to be designed is a trade-off between accurate estimation and computational time. The greater the number of the models designed, the more accurate the prediction, but at the same time the greater the computational burden. However for long-term predictions that are usually performed off-line, this is not a disadvantage. Literature shows that up to 10 models are adequate even for on-line applications. [33-39].

A bank of EKF is applied, one for each model, which can be run in parallel. This means that enormous computational time can be saved. At each iteration the MMPF selects the model which corresponds to the maximum posteriori probability as the correct one. This probability tends to one, while the others tend to zero. The overall optimal estimate can be taken either to be the individual estimate of the elemental filter exhibiting the maximum posterior probability (MAP) or the weighted average of the estimates produced by each (6) filter which the case used in this paper.

The probabilities are calculated in a recursive manner as it is shown in equations 1 and 2.

$$p(\theta_j/k) = \frac{L(k/k;\theta_j)}{\sum_{j=1}^{M} L(k/k;\theta_j)\, p(\theta_j/k\text{-}1)}\, p(\theta_j/k\text{-}1) \tag{9}$$

$$\begin{aligned} L(k/k;\theta_j) = &\left|\mathbf{P}_{\tilde{\mathbf{y}}}(k/k\text{-}1;\theta_j)\right|^{-1/2} \\ &\cdot \exp[-\frac{1}{2}\tilde{\mathbf{y}}^{T}(k/k\text{-}1;\theta_j) \\ &\mathbf{P}_{\tilde{\mathbf{y}}}^{-1}(k/k\text{-}1;\theta_j)\,\tilde{\mathbf{y}}(k/k\text{-}1;\theta_j)] \end{aligned} \tag{10}$$

where the innovations process

$$\tilde{\mathbf{y}}(k/k-1;\theta_j)=\mathbf{y}(k)-\mathbf{H}(k;\theta_j)\hat{\mathbf{x}}(k/k-1;\theta_j) \quad (11)$$

is a zero mean white process with covariance matrix

$$\mathbf{P}_{\tilde{y}}(k/k-1;\theta_j)=\mathbf{H}(k;\theta_j)\mathbf{P}(k/k;\theta_j)\mathbf{H}^{\mathrm{T}}(k;\theta_j)+\mathbf{R} \quad (12)$$

For all the above equations $j = 1,2,\ldots,M$.

2.2 Support Vector Machines

In support vector machines, 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:

$$y = \omega\cdot\Phi(x) + b \quad (13)$$

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

$$S(C)=C\frac{1}{M}\sum_{j=1}^{M} Lf_e(h_j-y_j)+\frac{1}{2}\|\omega\|^2, \quad (14)$$

$$Lf(h,y)=\begin{cases}|h-y|-e & |h-y|\geq e,\\ 0 & others,\end{cases} \quad (15)$$

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 (15) 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 (13) such that :

$$\omega\cdot\Phi(_j) + b_j - h_j \leq \mathrm{e} + \xi^*_j \quad (16)$$

$$-(\omega\cdot\Phi(_j) + b_j) + h_j \leq \mathrm{e} + \xi_{\mathrm{j}} \quad (17)$$

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 (14) which represents the objective function to be minimized is rearranged to:

$$S(\omega,\xi,\xi^*)=C^*\left(\sum_{j=1}^{M}\xi_j+\xi_j^*)\right)+\frac{1}{2}\omega\omega^{\mathrm{T}} \quad (18)$$

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 (19) the latter equation is reformed to:

$$S(a_j-a_j^*)=\sum_{j=1}^{M}h_i(a_j-a_j^*)-e\sum_{j=1}^{M}(a_j-a_j^*)$$
$$-\frac{1}{2}\sum_{j=1}^{M}\sum_{i=1}^{M}(a_j-a_j^*)\times(a_i-a_i^*)K(x_j-x_i) \quad (19),$$

subject to

$$\sum_{j=1}^{N}(a_j-a_j^*)=0,$$
$$0\leq a_{j,}\,a_j^*\leq C, \quad (20)$$

also where $j = 1,2,\ldots M$.

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

$$f(x,a,a^*)=\sum_{j=1}^{l}(a_j-a_j^*)K(x-x_i)+b \quad (21)$$

The ***Kernel*** function $K(x-x_i)$ introduced in equation (21) 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 Radial Basis Function (RBF), is used.

$$K(x_j-x_i)=e^{\left(\frac{-\|x_j-x_i\|^2}{2\sigma^2}\right)} \quad (22)$$

2.3 The proposed method

The weighted average of the estimates, L(t), 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 Extended Kalman filters of order (1,1), (2,2), (3,3), …,(10,10) programmed with the MMPF.

Then the MMPA's estimation error, e(t), was applied as input to the SVM , an analytical description can be found in [10], that was able to achieve a further error reduction and come up with a better forecasting outcome, NL(t).

As far as the SVM is concerned all of its parameters had to be carefully adjusted by trial and error. Unsuitable values for these parameters may lead to either over fitting or under fitting of the training data.

The most significant feature of the SVM compared to other similar algorithms is that it manages to achieve optimum performance by restricting the complexity of the objective function so that is the most suitable according to the quantity of the data present.

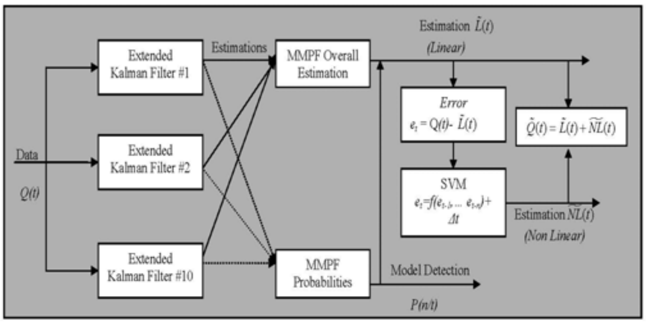


Figure 3: Schematic Representation of the proposed method **(for a better image resolution see at the end of the article)**

The output of hybrid model can be represented as:

$$Q_t = L_t + NL_t \tag{23}$$

Both parts are directly calculated from the electric load time series.

If e(t) is the MMPF estimation error at any time instant t, then:

$$e_t = Q_t - \hat{L}_t \tag{24}$$

It is now the SVM that models these residuals as:

$$e(t) = f(e_{(t-1)}, e_{(t-2)}, ..., e_{(t-n)}) + \Delta t \tag{25}$$

where f is non linear and Δt is random error. Consequently the forecast of the hybrid model is:

$$\hat{Q}_t = \hat{L}_t + \widehat{NL}_t \tag{26}$$

2.4 The ANN method

The structure of the artificial neural network applied in this work is analytically presented in [23] and a schematic diagram of its architecture is depicted in Figure 4.

It is a typical feedforward multilayer perceptron model (MLP), with an input layer of source neutrons, at least one hidden layer of computational neurons and an output layer of computational neurons. Although there are many types of ANN to be used, the one selected is appropriate due to its small solution network and quick computational speed, automatic generalization of knowledge enabling the recognitions of data sets, minimization of the mean squared error and its supervised training.

Thousands of MLP ANN were designed and tested as a combination of five backpropagation learning algorithms, five transfer functions consisted of 1-5 hidden layers with 2 to 100 neurons in each hidden layer (Table 1). The Matlab neural toolbox was extensively used to train, develop and validate the ANN [42].

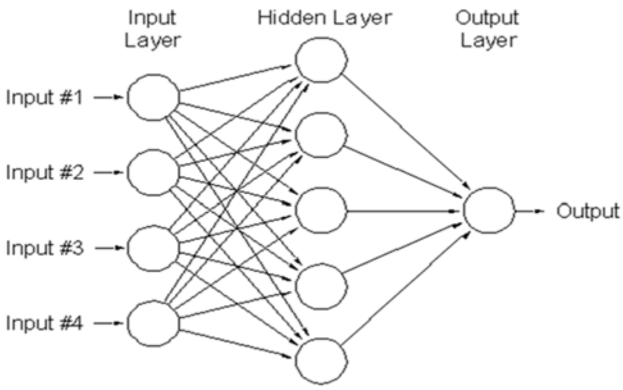


Figure 4: Schematic Representation of the ANN designed.

Table 1 Designed MLP ANN models

Structure	Learning Algorithm	Transfer Function
- 1 to 5 hidden layers - 2 to 100 neurons in each hidden layer	- Gradient Descent - Quasi-Newton - Levenberg-Marquardt - Random Order Incremental - Conjugate Gradient	- Hyperbolic Tangent Sigmoid - Logarithmic Sigmoid - Hard-Limit - Competitive - Linear

2.5 Data

The inputs for both proposed methods are namely the installed power capacity (IPC) and the annual electric energy consumption per capita, provided by the Independent Power Transmission Operator (ADMIE) [28], the yearly ambient temperature (YAT) provided by Greek national meteorological service [27] as well as the gross domestic product provided by Eurostat [40]. The output will be the final energy consumption (FEC) provided by the International Energy Agency [41].

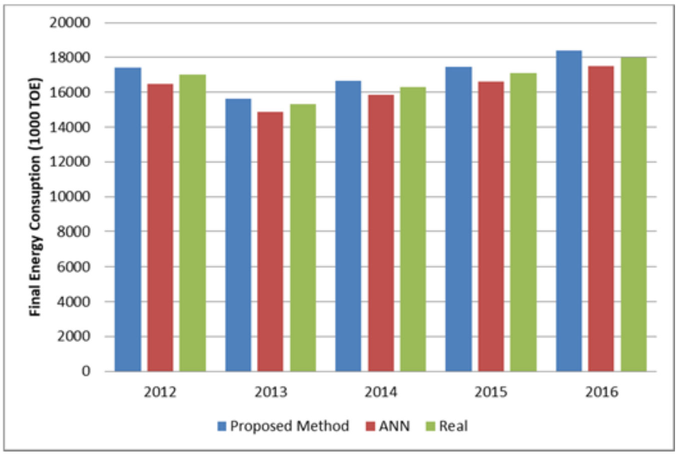
3 Results

Table 2 and Figure 5, indicate that both methods perform well and their predictions are very close to the real values since their absolute percentage error is not grater that 3%, with proposed method having the lowest absolute percentage error (APE(%)).

Table 2 Comparison Results

	Final Energy Consumption (1000 TOE)				
Year	**Proposed Method I**	**ANN II**	**Real**	**APE (100%) I**	**II**
2012	17420	16474	17000	2.47	3.09
2013	15624	14852	15300	2.12	2.93
2014	16667	15837	16300	2.25	2.84
2015	17456	16621	17100	2.08	2.80
2016	18391	17486	18000	2.17	2.86
			Average APE	**2.22**	**2.90**

Figure 5: Comparison of the predicted values



produced by the two adaptive methods with the real ones.

The absolute percentage error is defined as:

$$APE(\%) = \frac{|FEC_{real} - FEC_{predicted}|}{FEC_{real}} . 100\%$$

The proposed method has constantly an APE(%) less than 2.5 while the ANN's is closer to 3. Another interesting point for further research, is that the proposed method overestimates the final value, while the ANN underestimates it.

The MMPF is indeed adaptive but it has a main disadvantage which is that in its initial structure is not able to handle non-linearities and seasonalities. Its implementation with the EKF tackles the successfully non-linearity problem but not the second one.

A solution to this comes with its combination with the SVM. The problem of complexity and computational burden is overcome since the separate EKF's can be parallel implemented, thus saving enormous computational time.

A decision of how many EKF's have to be implemented needs to taken. Literature shows [33, 39] that a maximum number of 10 parallel EKF's is adequate in producing a reliable and fast prediction.

At this point it should mentioned that there also other values affecting the energy consumption, such as electricity price, number of installed air-conditions, amount of CO_2 pollution and many more for which no real data was able to be found.

4 Conclusions

The paper presents two adaptive techniques, for long-term electric load demand prediction. The data used was real and was collected from Independent Power Transmission Operator, Eurostat, Authority of Hellenic Statistics, Greek National Meteorological Service and International Energy Service were used.

The results showed that both methods successfully tackled the problem, however the proposed method's performance is slightly better in terms of absolute percentage error.

Both methods considered can be a useful tool for the Greek electric utilities and regulation authorities.

References:

[1] S. Urasawa, Real-time GDP forecasting for Japan: A dynamic factor model approach, *Journal of the Japanese and International Economies*, Vol. 34, 2014, pp. 116-134.

[2] T. Modis, Long-term GDP forecasts and the prospects for growth, *Technological Forecasting and Social Change*, Vol. 80, No. 8, 2013, pp. 1557-1562.

[3] S. Degiannakis, G. Filis, H. Hassani, Forecasting global stock market implied volatility indices, *Journal of Empirical Finance,* Vol. 46, 2018, pp. 111-129.

[4] R. de A. Araújo, A.L.I. Oliveira, S. Meira, A hybrid model for high-frequency stock market forecasting, *Expert Systems with Applications*, Vol. 42, No. 8, 2015, pp. 4081-4096.

[5] J.C. Rülke, M. Silgoner, J. Wörz, Herding behavior of business cycle forecasters,

International Journal of Forecasting, Vol. 32, No. 1, 2016, pp.23-33.

[6] J. Dombi, T. Jónás, Z.E. Tóth, Modeling and long-term forecasting demand in spare parts logistics businesses, *International Journal of Production Economics*, Vol. 201, 2018, pp. 1-17.

[7] S. Funk, A. Camacho, A.J. Kucharski, R.M. Eggo, W.J. Edmunds, Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model, *Epidemics*, Vol. 22, 2018, pp. 56-61.

[8] K. Togliatti, S.V. Archontoulis, R. Dietzel, L. Puntel, A. VanLoocke, How does inclusion of weather forecasting impact in-season crop model predictions?, *Field Crops Research*, Vol. 214, 2017, pp. 261-272.

[9] P.R. Larraondo, I. Inza, J.A. Lozano, A system for airport weather forecasting based on circular regression trees, *Environmental Modelling & Software*, Vol. 100, 2018, pp. 24-32.

[10] S.Sp. Pappas, G.E. Chatzarakis, C.C. Pappas, V.C. Moussas, A hybrid model for wind speed forecasting using arma models and support vector macines (svm), *5th International Conference on Experiments/Process/System Modeling/Simulation/Optimization*, 2013.

[11] S.Sp. Pappas, L. Ekonomou, D.Ch. Karamousantas, G.E. Chatzarakis, S.K. Katsikas, P. Liatsis, Electricity demand loads modeling using autoregressive moving average (ARMA) models, *Energy*, Vol. 33, No. 9, 2008, pp. 1353-1360.

[12] S.Sp. Pappas, L. Ekonomou, V.C. Moussas, P. Karampelas, S.K. Katsikas, Adaptive load forecasting of the Hellenic electric grid, *Journal of Zhejiang University SCIENCE A*, Vol. 9, No. 12, 2008, pp. 1724-1730.

[13] S.Sp. Pappas, L. Ekonomou, P. Karampelas, D.C. Karamousantas, S.K. Katsikas, G.E. Chatzarakis, P. D. Skafidas, Electricity demand load forecasting of the Hellenic power system using an ARMA model, *Electric Power Systems Research*, Vol. 80, No. 3, 2010, pp. 256-266.

[14] C. Yuan, S. Liu, Z. Fang, Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model, *Energy*, Vol. 100, 2010. pp. 384-390.

[15] S-J. Huang, K-R. Shih, Short-term load forecasting via ARMA model identification including non-Gaussian process considerations, *IEEE Trans on Power Systems*, Vol. 18, No. 2, 2003, pp. 673-679.

[16] A. Tarsitano, I.L. Amerise, Short-term load forecasting using a two-stage sarimax model, *Energy*, Vol. 133, 2017, pp. 108-114.

[17] N. Singh, S.R. Mohanty, R.D. Shukla, Short term electricity price forecast based on environmentally adapted generalized neuron, *Energy*, Vol. 125, 2017, pp. 127-139.

[18] A. Badri, Z. Ameli, A.M. Birjandi, Application of artificial neural networks and fuzzy logic methods for short term load forecasting, *Energy Procedia*, Vol. 14, 2012. pp. 1883-1888.

[19] X. Zhang, J. Wang, K. Zhang, Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm, *Electric Power Systems Research*, Vol. 146, 2017, pp. 270-285.

[20] Y. Yaslan, B. Bican, Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting, *Measurement*, Vol. 103, 2017, pp. 52-61.

[21] S. Pappas, Application and comparison of evolutionary techniques for forecasting the Hellenic grid electricity load, *International Journal of Power and Energy Research*, Vol. 1, No. 3, 2017, pp. 139-149.

[22] L. Ekonomou, C.A. Christodoulou, V. Mladenov, A short-term load forecasting method using artificial neural networks and wavelet analysis, *International Journal of Power Systems*, Vol. 1, 2016, pp. 64-68.

[23] L. Ekonomou, Greek long-term energy consumption prediction using artificial neural networks, *Energy*, Vol. 35, 2010, pp. 512-517.

[24] Authority of Hellenic Statistics, http://www.statistics.gr/

[25] M. Besseca, J. Fouquaub, The non-linear link between electricity consumption and temperature in Europe: Athreshold panel approach, *Energy Economics*, Vol. 30, No. 5, 2008, pp. 2705-2721.

[26] A.D. Amato, M. Ruth, P. Kirshen, J. Horwitz, Regional energy demand response to climate change: methodology and application to the

commonwealth of Massachusetts, *Climate Change*, Vol. 71, 2005, pp. 175-201.

[27] Greek National Meteorological Service, http://www.hnms.gr.

[28] Official Government Gazzete, Issue B, Number 2534, 17th August 2016.

[29] D.G. Lainiotis, Optimal adaptive estimation: Structure and parameter adaptation, *IEEE Trans on Automatic Control*, Vol. AC-16, 1971 pp. 160-170.

[30] D.G. Lainiotis, Partitioning: A unifying framework for adaptive systems I: Estimation, *Proc. of the IEEE,* Vol. 64, No. 8, 1976, pp. 1126-1143.

[31] D.G. Lainiotis, Partitioning: A unifying framework for adaptive systems II: Control, *Proc. of the IEEE*, Vol. 64, No. 8, 1976, pp. 1182-1198.

[32] S.Sp. Pappas, L. Ekonomou, P. Karampelas, S. K. Katsikas, P. Liatsis, Modeling of the grounding resistance variation using ARMA models, *Simulation Modelling Practice and Theory*, vol. 16, no. 5, 2008, pp. 560-570.

[33] S.Sp. Pappas, V.C. Moussas, S.K. Katsikas, Application of the multi-model partitioning theory for simultaneous order and parameter estimation of multivariate ARMA models, *International Journal of Modelling, Identification and Control (IJMIC)*, Vol. 4, No. 3, 2008, pp. 242-249.

[34] S.Sp. Pappas, V.C. Moussas, S.K. Katsikas, Adaptive MV ARMA identification under the presence of noise, *Proceedings of the European Computing Conference, Lecture Notes in Electrical Engineering 27*, 2005, pp. 787-797.

[35] V.C. Moussas, S.S. Pappas, Adaptive network anomaly detection using bandwidth utilisation data, *1st International Conference on Experiments/Process/System Modeling/ Simulation/Optimization, (1st IC-EpsMsO),* Athens, 6-9 July, 2005.

[36] V.C. Moussas, S.D. Likothanassis, S.K. Katsikas, A.K. Leros, Adaptive on-line multiple source detection, *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'05)*, Vol. 4, 2005, pp. 1029-1032.

[37] N.V. Nikitakos, A.K. Leros, S.K. Katsikas, Towed array shape estimation using multimodel partitioning filters, *IEEE Journal of Oceanic Engineering*, Vol. 23, No. 4, 1998, pp. 380-384.

[38] S. Katsikas, S. Likothanassis, G. Beligiannis, K. Berketis and D. Fotakis, Genetically determined variable structure multiple model estimation, *IEEE, Signal Processing*, Vol. 49, No. 10, 2001, pp. 2253-2261.

[39] G. Beligiannis, L. Skarlas and S. Likothanassis, A generic applied evolutionary hybrid technique, *IEEE Signal Processing Magazine*, Vol. 21, No. 3, 2004, pp. 28-38.

[40] EUROSTAT, http://ec.europa.eu/eurostat

[41] International Energy Service, https://www.iea.org/

[42] H. Demuth, M. Beale, Neural Network Toolbox: For use with MATLAB. The Math Works; 1994

APPENDIX

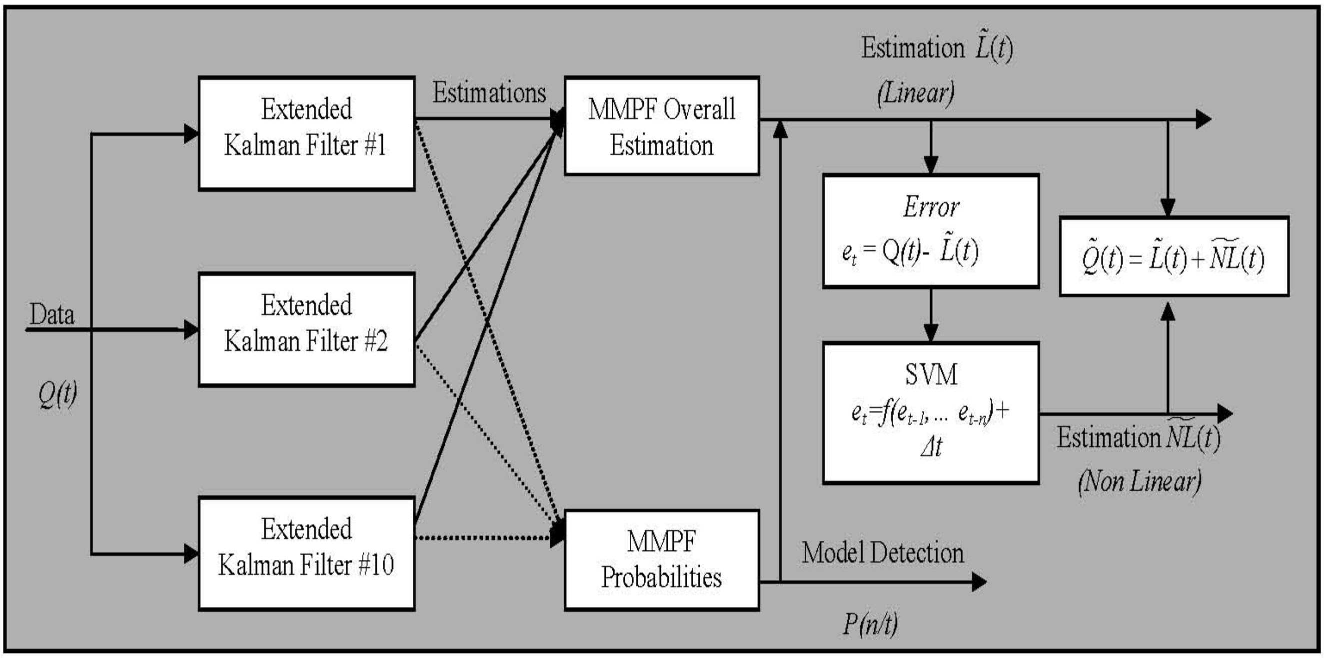


Figure 3: Schematic Representation of the proposed method