

## Artificial Neural Networks Approach to Time Series Forecasting for Electricity Consumption in Gaza Strip

Dr. Samir Khaled Safi  
Associate Professor of Statistics  
Department of Statistics  
The Islamic University of Gaza

**Abstract:** This paper introduces two robust forecasting models for efficient forecasting, Artificial Neural Networks (ANNs) approach and Autoregressive Integrated Moving Average (ARIMA) models. ANNs approach to univariate time series forecasting and relevant theoretical results are briefly discussed. To choose the best training algorithm for the ANN model, several experimental simulations with different training algorithms are made. We compare ANNs approach with ARIMA model on real data for electricity consumption in Gaza Strip.

The main finding is that, comparison of performance between the two proposed models reveals that ANNs outperform and preferable in selecting the most appropriate forecasting model over the ARIMA model.

**Keywords:** Forecasting, Box-Jenkins methodology, Neural Networks, Multilayer Perceptrons.

**ملخص:** في هذا البحث تم استعراض اثنين من نماذج التنبؤ القوية، الشبكات العصبية الاصطناعية (ANNs) ونماذج الانحدار الذاتي-التكاملي-المتوسط المتحرك (ARIMA). تم مناقشة استخدام طريقة الشبكات العصبية الصناعية للتنبؤ في السلاسل الزمنية وكذلك عرض مبسط لبعض النتائج النظرية ذات الصلة. تم استخدام العديد من عمليات المحاكاة التجريبية وذلك من أجل اختيار أفضل خوارزمية لنموذج الشبكات العصبية الصناعية. تم مقارنة نتائج استخدام الشبكات العصبية الصناعية مع ARIMA وذلك بتطبيقها على بيانات لاستهلاك الكهرباء في قطاع غزة. في الفترة 2000-2011. النتيجة الرئيسية للبحث هي أن استخدام نماذج الشبكات العصبية الصناعية أفضل في التنبؤ من نموذج ARIMA.

### 1. Introduction

Neural networks are the preferred tool for many predictive data mining applications because of their flexibility, power, accuracy and ease of use. Electricity consumption forecasting is an important issue for energy service companies. Having reliable electricity consumption forecasting information will make better financial decision. The electricity consumption influence factors, such as load, weather, market forces, and bidding strategy are undulating and undetermined, so the consumption forecasting with high precision is more difficult, see for example Pousinho, H., et al. (2012) and

Unsihuay, V., et al. (2010). Therefore, it has becoming the commonly and difficulty problem to forecast electricity consumption in competitive markets all over the world.

In 1976, Box–Jenkins used statistical models to forecast the financial market, Box, G. & Jenkins, G. (1976). However, the statistical methods assume that data are linearly related and therefore is not true in real life applications. The newly introduced method, the artificial neural network (ANN) has emerged to be popular as it does not make such assumptions. The ANN which is inherently a nonlinear network and does not make such assumptions therefore is well suited for prediction purpose.

Mabel, M. and Fernández, E. (2008), showed that with the development of artificial technique, some artificial intelligent prediction methods have been discussed, including ANNs. To attain better performance, most proposed models are combinations of several kinds of the upper methods, see for example Barbounis, T. and Theocharis, J. (2007).

In this study, ARIMA and the ANN have been conducted for electricity consumption forecasting. The time series models such as ARIMA model is used to find the potential forecasting model. During the calculation process of time series modeling, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF) and the Extended Autocorrelation Function (EACF) criterion will be adopted.

The purpose of this work is to find a simple and reliable forecasting model for the electricity consumption in Gaza Strip. This paper is organized as follows: Section 2 presents overview and literature of ANN; Section 3 illustrates some basic concepts and definitions; Sections 4 and 5 display two forecasting cases fitting ARIMA and ANN models for electricity consumption data; and Section 6 concludes some important results of this work.

**Data Source:** We use a data set of electricity consumption from Palestinian Energy Authority-Gaza branch. The dataset contains the monthly consumption of electricity in Gaza Strip during the period January 2000 through December 2011.

## **2. Overview and Literature of ANN**

The ANN has been used in signal processing due to its nonlinear capacity and robust performance. The structure of the ANN is very important for its performance. Cadenas, E. and Rivera, W. (2009) showed that three-layer network is enough to fit any non-stationary signal. In ANN theory, the training data format can affect the performance of network directly.

ANNs constitute one of the most powerful tools for pattern classification due to their nonlinear and non-parametric adaptive-learning properties.

## **Artificial Neural Networks Approach to Time Series**

Many studies have been conducted that have compared ANNs with other traditional classification techniques, since the default prediction accuracies of ANNs are better than those using classic linear discriminant analysis and logistic regression techniques, see for example Lee, T. and Chen, I., (2005) and Lee, T., et al. (2002).

The Multilayer Perceptron (MLP) produces a predictive model for one or more dependent variables based on the values of the predictor variables.

Blanco, A., et al. (2013) introduced several non-parametric credit scoring models based on the MLP approach and benchmarks their performance against other models which employ the traditional linear discriminant analysis, quadratic discriminant analysis, and logistic regression techniques. Based on a sample of almost 5500 borrowers from a Peruvian microfinance institution, the results reveal that neural network models outperform the other three classic techniques both in terms of area under the receiver-operating characteristic curve (AUC) and as misclassification costs.

ANN usually uses Back Propagation (BP) as its training algorithm. To improve the performance of the neural network with BP, more training algorithms have been reported in recent years, including Quick Back Propagation (QBP), Resilient Back Propagation (RBP), Broyden – Fletcher – Goldfarb - Shanno Quasi-Newton Back Propagation (BFGS). Liu, H., et al. (2012) showed that BGFS algorithm gives the best performance. Hence, BGFS algorithm is chosen as the training algorithm of the ANN model.

Majhi, B. et al., (2012) introduced two robust forecasting models for efficient prediction of different exchange rates for future months ahead. These models employ Wilcoxon artificial neural network (WANN) and Wilcoxon functional link artificial neural network (WFLANN). Comparison of performance between the two proposed models reveals that both provide almost identical performance but the later involved low computational complexity and hence is preferable over the WANN model.

Many hybrid models have been suggested using the ANN for exchange rate forecasting. Khashei, M. and Bijari, M. (2011) proposed a novel hybridization of artificial neural networks and ARIMA model in order to overcome limitation of ANNs and has been demonstrated it to be a more accurate model than the traditional ones. This model has the unique advantages of ARIMA models in linear modeling to identify and magnify the existing linear structure in the data, and then a neural network is used in order to determine a model to capture the underlying data generating process and predict, using preprocessed data.

### 3. Preliminaries

This section introduces some basic definitions and concepts.

#### The Multilayer Perceptron (MLP)

MLP networks are constructed of multiple layers of computational units. Each neuron in one layer is directly connected to the neurons of the subsequent hidden layer. MLP utilizes a supervised learning technique called back propagation (BP) for training the network, which is the most popular being used. Each MLP is composed of a minimum of three layers consisting of an input layer, one or more hidden layers and an output layer. The input layer distributes the inputs to subsequent layers. Input nodes have linear activation functions and no thresholds. Each hidden unit node and each output node have thresholds associated with them in addition to the weights. The hidden unit nodes have nonlinear activation functions and the outputs have linear activation functions (See for example, Walter, H. and Michael, T., 2005, and Nazzal, J., et al. , 2008). MLPs using a BP algorithm are the standard algorithm for any supervised learning pattern recognition process.

It has been shown most problems it would be enough to have only one layer of hidden neurons, Hornik, K., et al. (1989).

The mathematical representation of the function applied by the hidden neurons in order to obtain an output value  $b_{pj}$ , when faced with the presentation of an output vector  $X_p : x_{p1}, \mathbf{K}, x_{pi}, \mathbf{K}, x_{pN}$ , is defined by:

$$b_{pj} = f_L \left( \mathbf{q}_j + \sum_{i=1}^N w_{ij} \cdot x_{pi} \right) \quad (3.1)$$

where  $f_L$  is the activation function of hidden neurons  $j$ ,  $w_{ij}$  is the weight of the connection between input neuron  $i$  and hidden neuron  $j$  and  $x_{pi}$  is the input signal received by input neuron  $i$  for pattern  $p$ .

Once the output of the output neurons is concerned, it is obtained using

$$\hat{y}_{pk} = f_M \left( \mathbf{q}_k + \sum_{j=1}^L v_{jk} \cdot b_{pj} \right) \quad (3.2)$$

where  $\hat{y}_{pk}$  is the output signal provided by output neuron  $k$  for pattern  $p$ ,  $f_M$  is the activation function of output neurons  $M$ ,  $\mathbf{q}_k$  is the threshold of output neuron  $k$  and  $v_{kj}$  is the weight of the connection between hidden neuron  $j$  and output neuron  $k$ , Moreno, J., et al. (2011).

## Artificial Neural Networks Approach to Time Series

MLP allow a neural network to perform arbitrary mappings. A 2-hidden layer neural network is shown in Figure 3.1. The aim is to map an input vector  $x$  into an output  $y(x)$ .

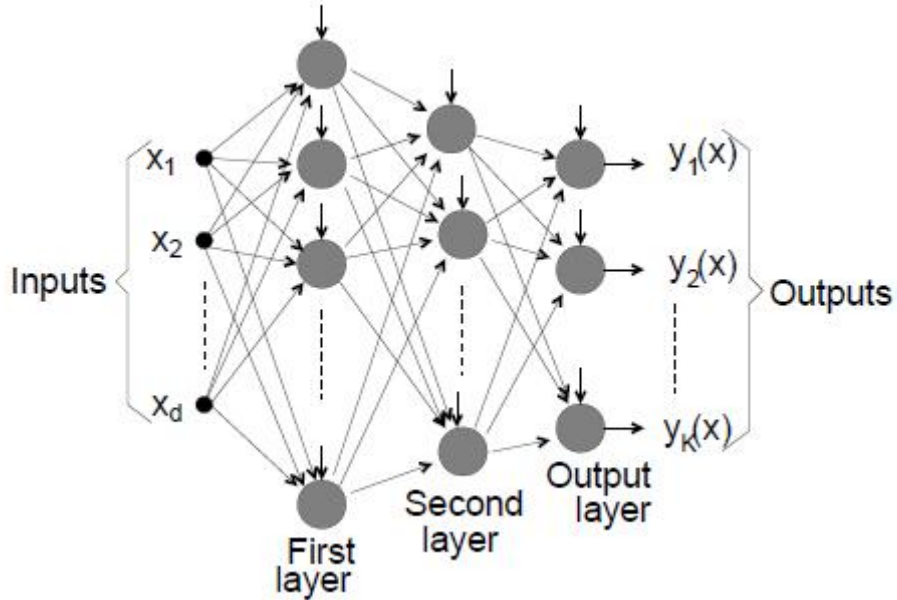


Figure 3.1: A 2-Hidden Layer Neural Network

The overall performance of the MLP is measured by the mean square error (MSE) expressed by :

$$MSE = \frac{\sum_{p=1}^{N_p} \sum_{i=1}^M [t_p(i) - y_p(i)]^2}{N}, \quad (3.3)$$

where,  $N_p$  is a set of training patterns  $(x_p, t_p)$  where  $P$  represents the pattern number.

$X_p$  corresponds to the  $N$ -dimensional input vector of the  $p^{\text{th}}$  training pattern and  $Y_p$  corresponds to the  $M$ -dimensional output vector from the trained network for the  $p^{\text{th}}$  pattern.

Note  $\sum_{i=1}^M [t_p(i) - y_p(i)]^2$  Corresponds to the error for the  $p^{\text{th}}$  pattern and  $t_p$  is the desired output for the  $p^{\text{th}}$  pattern (Nazzal, J., et al. 2008).

### ARIMA Models

A time series  $\{Y_t\}$  is said to follow an autoregressive-integrated moving average model (ARIMA) if the  $d^{\text{th}}$  difference  $W_t = \nabla^d Y_t$  is a stationary ARMA process. If  $\{W_t\}$  follows an ARMA(p,q) model, we say that  $\{Y_t\}$  is an ARIMA (p,d,q) process. An ARIMA (p,d,q) time series can be represented in a shorter form using the notation of lag operator.

The lag operator  $B$ , is defined as  $BY_t = Y_{t-1}$ , the operator which gives the previous value of the series.

**Definition:** The general ARIMA(p,d,q) process is given by (Box, G., et al. 1994)

$$\phi(B) \nabla^d Y_t = \theta(B) \varepsilon_t, \quad (3.4)$$

where  $d \geq 1$  is the degree of differencing,  $\nabla = 1 - B$  is the differencing operator,  $\phi(B)$  and  $\theta(B)$  are polynomials of degree p and q in B,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.5)$$

and

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (3.6)$$

Stationarity requires the roots of  $\phi(B)$  to lie outside the unit circle, and invertibility places the same condition on the roots of  $\theta(B)$ .

### Mean Squared Error

Many measures of forecast accuracy have been developed in the past, and several authors have been made recommendations about what should be used comparing the accuracy of forecast methods applied to univariate time series data. For example, Hyndman, R. and Koehler, A. (2005) introduced the Mean Square Error (MSE) as a measure of dispersion between the actual and the predicted value.

**Definition:** The MSE is given by:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2, \quad (3.7)$$

where  $Y_i$  is the actual value of the  $i^{\text{th}}$  iteration and  $\hat{Y}_i$  is the predicted value of the same  $i^{\text{th}}$  iteration. MSE is one of the most commonly used measures of forecast accuracy.

## Artificial Neural Networks Approach to Time Series

### AIC Criterion

Akaike's (1973) information criterion (AIC) plays a major role for selecting the best order of the ARIMA(p,d,q) model when we have several models that all adequately represent a given set of time series.

**Definition:** Suppose  $\{Y_t\}$  is a Gaussian autoregressive ARMA(p,q) process with coefficient vector  $\Theta = (f, q)$ . For a zero-mean causal invertible ARMA(p,q) process, the AIC is given by

$$AIC(\Theta) = -2 \ln L_x(\Theta, S_x(\Theta)/n) + 2k, \quad (3.8)$$

where  $L_x(\Theta, S_x(\Theta)/n)$  is the likelihood function,  $n$  is the sample size, and  $k$  is the total number of parameters, i.e.  $k = p + q + 1$ .

For fitting autoregressive models, Jones, R. (1975) and Shibata, R. (1976) suggested that AIC has a tendency to overestimate  $p$ . The AIC is a biased estimator, Hurvich and Tsai (1989) showed that the bias can be approximately eliminated by adding another nonstochastic penalty term to the AIC, resulting in the corrected AIC, denoted by AICc and defined by the formula

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{n-k-2} \quad (3.9)$$

### BIC Criterion

Schwarz's Bayesian information criterion (1978), known as (BIC) is another criterion that attempts to correct the overfitting nature of the AIC. For a zero-mean causal invertible ARMA(p,q) process, the BIC is given by:

$$BIC(\Theta) = -2 \ln L_x(\Theta, S_x(\Theta)/n) + k \log(n) \quad (3.10)$$

As a rule of thumb, we would expect as small value as possible for all of these criteria to select the most appropriate autoregressive model.

### KPSS test

The most commonly used stationarity test, the KPSS test, is due to Kwiatkowski, Phillips, Schmidt and Skin (1992). They derived their test by starting with the model

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 t + \theta_t + u_t \\ \theta_t &= \theta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2), \end{aligned} \quad (3.11)$$

where  $u_t$  is stationary time series and is said to be integrated of order zero,  $I(0)$  and may be heteroskedastic. The null hypothesis that  $Y_t$  is  $I(0)$  is formulated as  $H_0 : \sigma_\varepsilon^2 = 0$ , which implies that  $\theta_t$  is a constant. This test also implies a unit moving average root in the ARMA representation of  $\nabla Y_t$ .

**Definition:** The KPSS test statistic is the Lagrange Multiplier (LM) or score statistic for testing  $H_0 : \sigma_\varepsilon^2 = 0$  versus  $H_a : \sigma_\varepsilon^2 > 0$  and is given by (Kozhan, R., 2010)

$$KPSS = T^{-2} \sum_{t=1}^T \hat{S}_t^2 / \hat{\beta}^2, \quad (3.12)$$

where  $\hat{S}_t^2 = \sum_{j=1}^t \hat{u}_j$ ,  $\hat{u}_t$  is the residual of a regression  $Y_t$  on  $t$  and  $\hat{\beta}^2$  is a consistent estimate of the long-run variance of  $u_t$  using  $\hat{u}_t$ .

**Ljung-Box portmanteau test**

Portmanteau test firstly has been studied by Box, G. and Pierce, D. (1970). Ljung, G. and Box, G. (1978) proposed a modified version of that test.

**Definition:** Ljung-Box  $Q_{LB}^%$  portmanteau test is

$$Q_{LB}^%(\hat{r}) = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n-k}, \quad (3.13)$$

where  $\hat{r}_k$  is the sample autocorrelation of order  $k$  of the residual and  $n$  is the sample size, and  $m$  is the number of lag. Notice that  $(n+2)/(n-k) > 1$  for  $k \geq 1$ .

**The Autocorrelation Function (ACF)**

**Definition:** For a covariance stationary time series  $\{Y_t\}$  the autocorrelation function  $r_k$  is given by

$$r_k = Corr(Y_t, Y_{t-k}) \text{ for } k = 1, 2, 3, \mathbf{K} \quad (3.14)$$

ACF is a good indicator of the order of the MA(q) model since it cuts off after lag q (i.e.  $r_k = 0$  for  $k > q$ ). On the other hand the ACF tails off for AR(p) model.



## Artificial Neural Networks Approach to Time Series

### The Partial Autocorrelation Function (PACF)

**Definition:** If  $\{Y_t\}$  is normally distributed time series, then the PACF at lag  $k$  is given by

$$f_{kk} = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}) \quad (3.15)$$

PACF is a good indicator of the order of the AR(p) model since it cuts off after lag  $p$  (i.e.  $f_{kk} = 0$  for  $k > p$ ). On the other hand the PACF tails off for MA(q) model.

### The Extended Autocorrelation Function (EACF)

For a mixed ARMA model, ACF and PACF have infinitely many nonzero values, making it difficult to identify mixed models from the sample ACF and PACF. The extended autocorrelation function (EACF) (Tsay, R. and Tiao, G., 1984) is a graphical tool is used to identify the ARMA orders.

**Definition:** (Cryer, J. and Chan, K., 2008) Let

$$W_{t,k,j} = Y_t - \phi_1 Y_{t-1} - \dots - \phi_k Y_{t-k} \quad (3.16)$$

be the autoregressive residuals defined with the AR coefficients estimated iteratively assuming the AR order is  $k$  and the MA order is  $j$ . The sample autocorrelations of  $W_{t,k,j}$  are referred to as the EACFs. Tsay, R. and Tiao, G. (1984) suggested summarizing the information in the sample EACF by a table with the element in the  $k$ th row and  $j$ th column equal to the symbol X if the lag  $j + 1$  sample correlation of  $W_{t,k,j}$  is significantly different from 0. In such a table, an ARMA( $p,q$ ) process will have a theoretical pattern of a triangle of zeroes, with the upper left-hand vertex corresponding to the ARMA orders.

## 4. Fitting ARIMA Model for Electricity Consumption Data

Consider the monthly consumption of electricity (in millions of kilowatt-hours, MKWH) in Gaza Strip, from January 2000 through December 2011. R-statistical software is used for fitting ARIMA model for the time series.

Figure 4.1 displays the time series plot. The series displays considerable fluctuations over time, especially since 2004, and a stationary model does not seem to be reasonable. The higher values display considerably more variation than the lower values. Note all Figures are shown in the Appendix. The sample ACF for the data is displayed in Figure 4.2. All values shown are “significantly far from zero,” and the only pattern is perhaps a linear decrease with increasing lag. This means that we are dealing with a nonstationary time series.

In addition, software implementation of the KPSS test for level stationarity applied to the original consumption leads to a test statistic of 3.9841 and a  $p$ -value of 0.01. With stationarity as the null hypothesis, this provides strong evidence supporting the nonstationarity and the appropriateness of taking a difference of the original series.

The differences of the electricity values are displayed in Figure 4.3. The differenced series looks much more stationary when compared with the original time series shown in Figure 4.1. On the basis of this plot, we might well consider a stationary model as appropriate.

KPSS test is applied to the differenced series leads to a test statistic of 0.0156 and a  $p$ -value of 0.10. That is, we do not reject the null hypothesis of Stationarity.

The sample ACF and PACF are shown in Figures 4.4 and 4.5, respectively. It is quite difficult to identify the AR, MA, or mixed model from these figures.

The sample EACF computed on the first differences of the electricity consumption series is shown in Table 4.1. In this table, an ARMA( $p,q$ ) process will have a theoretical pattern of a triangle of zeroes, with the upper left-hand vertex corresponding to the ARMA orders.

**Table 4.1:** EACF for Difference of Electricity Consumption Series

		MA													
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
AR	0	0	0	x	X	0	X	0	0	0	0	0	0	0	0
	1	X	0	0	X	0*	0	0	0	0	0	0	0	0	0
	2	0	0	0	X	0	0	0	0	0	0	0	0	0	0
	3	x	x	x	X	0	0	0	0	0	0	0	0	0	0
	4	x	x	x	0	0	0	0	0	0	0	0	0	0	0
	5	x	x	x	x	0	X	0	0	0	0	0	0	0	0
	6	0	x	x	0	0	0	0	0	0	0	0	0	0	0
	7	x	0	x	x	x	0	0	0	0	0	0	0	0	0

Table 4.1 displays the schematic pattern for an ARMA(1,4) model. The upper left-hand vertex of the triangle of zeros is marked with the symbol 0\* and is located in the  $p = 1$  row and  $q = 4$  column—an indication of an ARMA(1,4) model. The model for the original electricity consumption series would then be a nonstationary ARIMA(1,1,4) model.

Different combinations of ARIMA models with  $p + q \leq 5$  and their corresponding criteria are shown in Table 4.2. These choices confirm our

## Artificial Neural Networks Approach to Time Series

suggestion-ARIMA (1,1,4)- based on the smallest values of AIC, AICc, BIC and RMSE among the other ARIMA choices.

**Table 4.2:** Different combinations of ARIMA models

Model Order	AIC	AICc	BIC	RMSE
(1,1,0)	836.96	837.15	845.58	5.769887
(2,1,0)	838.91	839.22	850.41	5.768732
(3,1,0)	833.89	834.37	848.27	5.612885
(4,1,0)	828.56	829.24	845.81	5.453366
(5,1,0)	830.12	831.03	850.25	5.443784
(0,1,1)	836.96	837.15	845.58	5.769885
(0,1,2)	838.83	839.15	850.33	5.766890
(0,1,3)	815.38	815.86	829.76	5.148814
(0,1,4)	813.08	813.76	830.33	5.072234
(0,1,5)	814.14	815.05	834.27	5.050224
(1,1,1)	820.98	821.30	832.48	5.322629
(1,1,2)	820.36	820.84	834.74	5.265777
(1,1,3)	813.90	814.57	831.15	5.088764
<b>(1,1,4)</b>	<b>811.59</b>	<b>812.50</b>	<b>831.72</b>	<b>4.980446</b>
(2,1,1)	818.87	819.35	833.24	5.232006
(2,1,2)	823.35	824.03	840.61	5.287416
(2,1,3)	815.78	816.69	835.91	5.085641
(3,1,1)	816.08	816.76	833.33	5.128288
(3,1,2)	818.02	818.93	838.14	5.127499
(4,1,1)	826.38	827.30	846.51	5.363132

We use maximum likelihood estimation and show the results obtained from the R statistical software in Table 4.3. Here we see that  $\hat{f} = -0.5743$ ,  $\hat{q}_1 = 0.4091$ ,  $\hat{q}_2 = -0.3326$ ,  $\hat{q}_3 = -0.5791$ , and  $\hat{q}_4 = -0.4974$ . We also see that the estimated noise variance is  $\hat{\sigma}_e^2 = 24.99$ . Noting the P-values, the estimates of all autoregressive and moving average coefficients are significantly different from zero statistically, as is the intercept term.

**Table 4.3:** Maximum Likelihood Estimates from R Software: Electricity Consumption Series

Coefficients:	AR(1)	MA(1)	MA(2)	MA(3)	MA(4)	Intercept*
	-0.5743	0.4091	-0.3326	-0.5791	-0.4974	0.4235
SE	0.1822	0.1767	0.0817	0.1068	0.0814	0.0283
T	-3.1528	2.3151	-4.0703	-5.4205	-6.1074	14.9573
P-value	0.0020	0.0222	0.0008	< 0.0001	< 0.0001	< 0.0001

sigma^2 estimated as 24.99: log likelihood = -398.8 AIC = 811.59 AICc = 812.5  
BIC = 831.72

\* The intercept here is the estimate of the process mean  $m$  not of  $q_0$ .

The estimated model would be written

$$W_t - 0.424 = -0.574(W_{t-1} - 0.424) + e_t - 0.409e_{t-1} + 0.333e_{t-2} + 0.579e_{t-3} + 0.497e_{t-4}, \quad (4.1)$$

where  $W_t = Y_t - Y_{t-1}$ , and the intercept of ARIMA is  $q_0 = m(1-f)$ , then  $q_0 = 0.4235(1+0.5743) = 0.6667$ . Therefore, the estimated model is

$$Y_t = 0.667 + 0.426Y_{t-1} + 0.574Y_{t-2} + e_t - 0.409e_{t-1} + 0.333e_{t-2} + 0.579e_{t-3} + 0.497e_{t-4} \quad (4.2)$$

Figure 4.6 displays the time series plot of the standardized residuals from the ARIMA(1,1,4) model estimated for the electricity consumption time series. The model was fitted using maximum likelihood estimation. There is only one residual with magnitude larger than 1.

A quantile-quantile plots are an effective tool for assessing normality. Here we apply them to the residuals of the fitted model. A quantile-quantile plot of the residuals from the ARIMA(1,1,4) model estimated for the electricity consumption series is shown in Figure 4.7. The points seem to follow the straight line fairly closely. This graph would not lead us to reject normality of the error terms in this model. In addition, the Kolmogorov-Smirnov of composite normality test applied to the residuals produces a test statistic of  $ks = 0.0546$ , which corresponds to a  $p$ -value of 0.50, and we would not reject normality based on this test.

To check on the independence of the error terms in the model, we consider the sample autocorrelation function of the residuals. Figure 4.8 displays the sample ACF of the residuals from the ARIMA(1,1,4) model of the electricity consumption data. The dashed horizontal lines plotted are based on the large lag standard error of  $\pm 2/\sqrt{n} = \pm 0.174$ . The graph does not show statistically significant evidence of nonzero autocorrelation in the residuals. In other words, there is no evidence of autocorrelation in the residuals of this model. These residual autocorrelations look excellent.

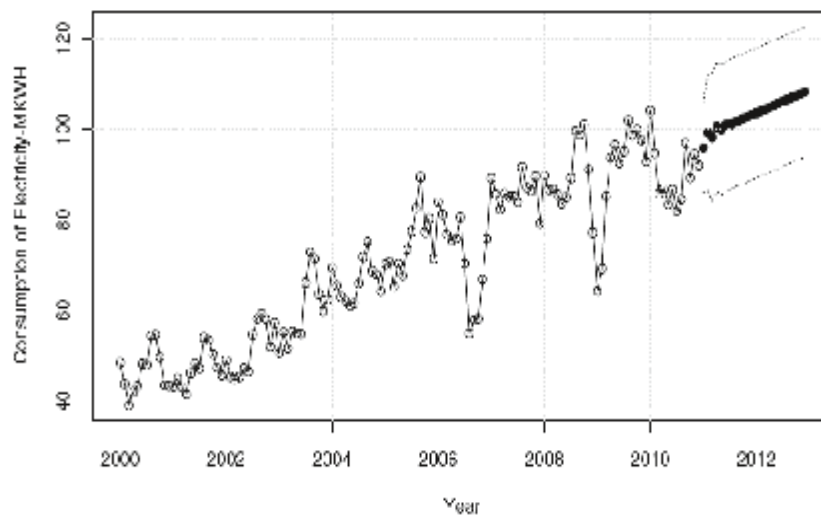
In addition to looking at residual correlations at individual lags, it is useful to have a test that takes into account their magnitudes as a group. Figure 4.9 shows the  $p$ -values for the Ljung-Box test statistic for a whole range of values of  $K$  from 6 to 20. The horizontal dashed line at 5% helps judge the size of the  $p$ -values. The Ljung-Box test statistic with  $K = 7$  is equal to 2.996. This is referred to a chi-square distribution with two degrees of freedom. This leads to a  $p$ -value of 0.2236, so we have no evidence to reject

## Artificial Neural Networks Approach to Time Series

the null hypothesis that the error terms are uncorrelated. The suggested model looks to fit the modeling time series very well.

Therefore the estimated ARIMA(1,1,4) model seems to be capturing the dependence structure of the difference of electricity consumption time series quite well. Figure 4.10 shows the data and forecasting results of ARIMA (1,1,4) models for Electricity consumption (MKWH) in 2012.

**Figure 4.10:** Data and Forecasting results of ARIMA (1,1,4) models for Electricity consumption (MKWH) in 2012



The runs test may also be used to assess dependence in error terms via the residuals. Applying the test to the residuals from the ARIMA(1,1,4) model for the electricity consumption series, we obtain expected runs of 66.86364 versus observed runs of 74. The corresponding  $p$ -value is 0.245, so we do not have statistically significant evidence against independence of the error terms in this model. In addition, the minimum Root Mean Squares Error (RMSE) for ARIMA (1,1,4) model equals 4.9804.

### 5. Fitting ANN Model for Electricity Consumption Data

Applying ANN, the percentage of observations for training, which must have the same number of observations, 132, as we have in ARIMA for training is determined, so we have increased in a series of 12 observations. Thus, we have an input consists of 144 observations, 90% for training, and 10% for comparison in the prediction. The layers may be described as: *Input* layer: accepts the data vector or pattern; *Hidden* layers: one or more layers. *Output* layer: takes the output from the final hidden layer to produce the target values.

In choosing the number of layers the following considerations are made. Multi-layer networks are harder to train than single layer networks. A two layer network (one hidden) can model any decision boundary. Two layer networks are most commonly used in pattern recognition.

The number of output units is determined by the number of output classes. The number of inputs is determined by the number of input dimensions. The network will not model complex decision boundaries for few hidden units and it will have poor generalization for too many number of hidden units

We started with one hidden layer and end with fifteen layers. The performance of the algorithm is influence with choosing different learning rates. The algorithm may could become unstable for high learning rate and might take longer time to converge.

*R*-software is used for fitting ANN model for the time series. Some commands and functions with input and output variables have been used. The *R* library '*neuralnet*' is used to train and build the neural network. The *nnet* function is used to fit neural networks. The arguments are: *size* which determines the number of units in the hidden layer, and *maxit* determines the maximum number of iterations. The objects are: *fitted.values* is used for the fitted values for the training data and *residuals* is used to show the residuals for the training data (Venables, W. N. and Ripley, B. D. ,2002).

RMSE is used as stopping criteria in the network. Smaller values of RMSE indicate higher accuracy in forecasting. The Neural network result shows that the minimum RMSE equals 0.0768 for considering the model with fifteen units in the hidden layer, two lags and the learning rate equals to 0.01.

Table 5.1 shows the actual and forecasting results for Electricity consumption (MKWH) in 2011 based on ANN and ARIMA (1,1,4) models. It is quite obvious that the ANN forecasts mimic the actual values of the electricity consumption. Table 5.2 and shows the forecasting results for Electricity consumption (MKWH) in 2012 based on ANN and ARIMA (1,1,4) models.

## Artificial Neural Networks Approach to Time Series

**Table 5.1:** Actual and Forecasting results of ANN and ARIMA (1,1,4) models for Electricity consumption (MKWH) in 2011

Year (2011)	Actual data	Forecast	
		ANN	ARIMA
Jan	96.375285	96.375300	95.939790
Feb	104.044598	104.044600	99.279110
Mar	92.962289	92.962300	98.211320
Apr	99.571429	99.571400	100.520520
May	96.067993	96.068000	99.861080
Jun	101.550216	101.550200	100.906510
Jul	104.943501	104.943500	100.972850
Aug	105.816438	105.816400	101.601470
Sep	113.183204	113.183200	101.907180
Oct	107.519680	107.519700	102.398330
Nov	120.037919	120.037900	102.782980
Dec	91.942274	91.942300	103.228800

**Table 5.2:** Forecasting results of ANN and ARIMA (1,1,4) models for Electricity consumption (MKWH) in 2012

Year (2012)	Forecast	
	ANN	ARIMA
Jan	103.0393	103.6395
Feb	105.8420	104.0704
Mar	96.60480	104.4896
Apr	99.73830	104.9156
May	101.6009	105.3377
Jun	97.95320	105.7620
Jul	98.71340	106.1850
Aug	99.75960	106.6088
Sep	98.27490	107.0321
Oct	98.34590	107.4557
Nov	98.88840	107.8792
Dec	98.28100	108.3027

The RMSE for ARIMA and ANN equal 4.9804 and 0.0768, respectively. This result shows that RMSE of ANN is 1.54% of RMSE for ARIMA. In other words, the RMSE of ARIMA model is 64.85 times RMSE of the

ANN model. This means ANN model for forecasting is much more accurate and efficient than the ARIMA forecasting model.

### **6. Conclusion**

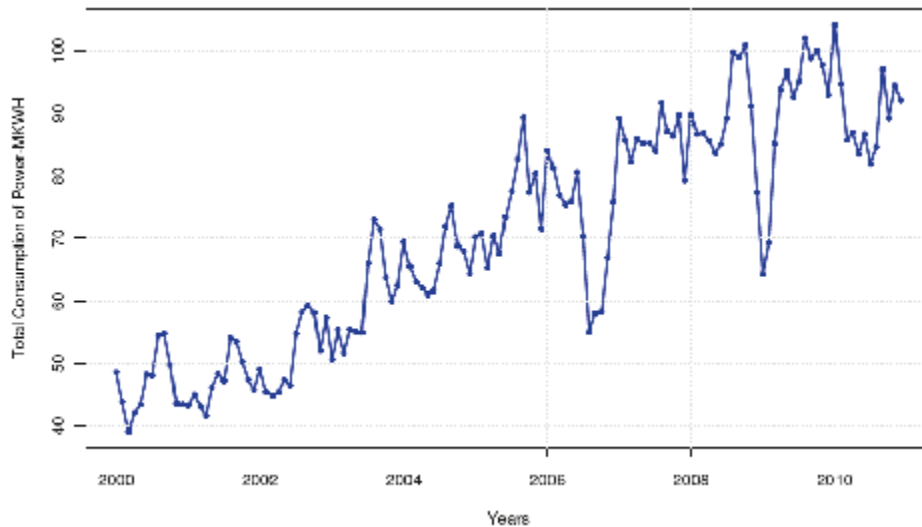
This paper has proposed two efficient approaches forecasting models. In the first model multilayer neural network is trained by minimizing RMSE and the second model consists of using ARIMA model on real data for electricity consumption in Gaza Strip. The results of both models reveal that ANNs outperform and offer consistent prediction performance compared to ARIMA model and hence preferable as a robust prediction model for electricity consumption.

### **Acknowledgements**

This study was supported by the Scientific Research Deanship at the Islamic University of Gaza-Palestine. We are grateful for the referees for their valuable comments and suggestions on earlier draft of this paper.

### **Appendix**

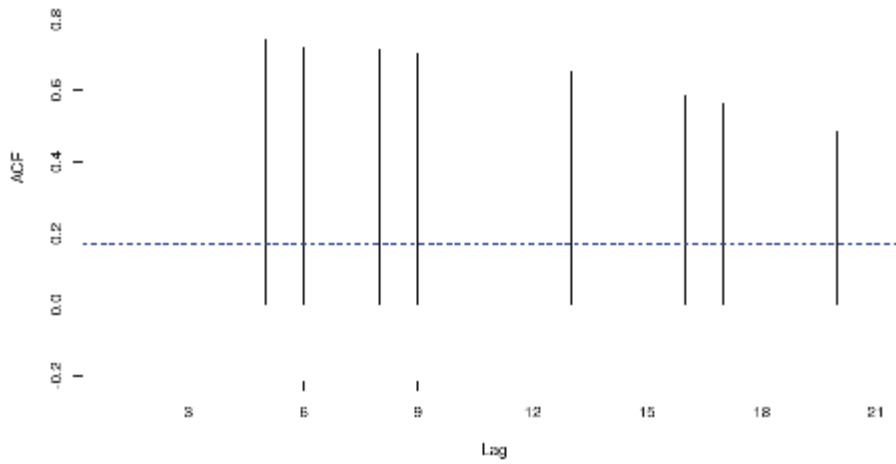
**Figure 4.1:** Monthly Consumption of Electricity (MKWH): January 2000–December 2011



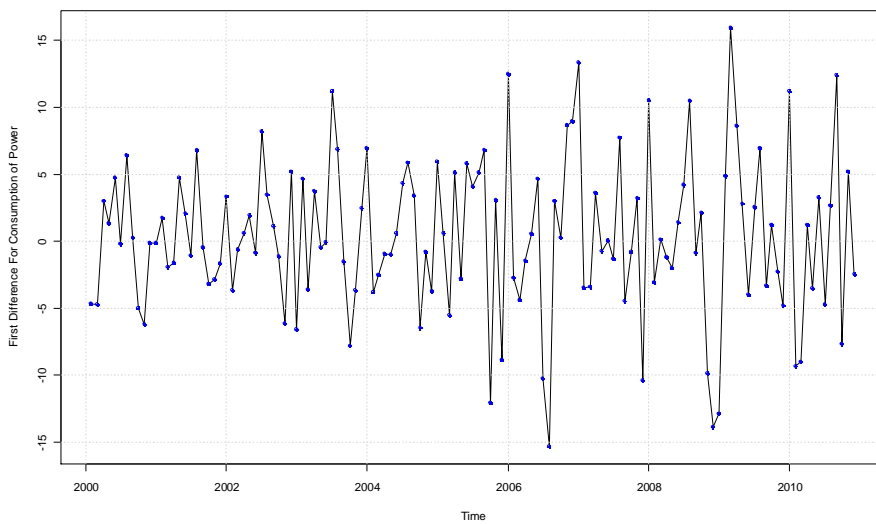
**Figure 4.2:** Sample ACF for the Electricity Consumption Time Series



## Artificial Neural Networks Approach to Time Series



**Figure 4.3:** The Difference Series of the Monthly Electricity Consumption



**Figure 4.4:** Sample ACF for Difference of Electricity Consumption Series

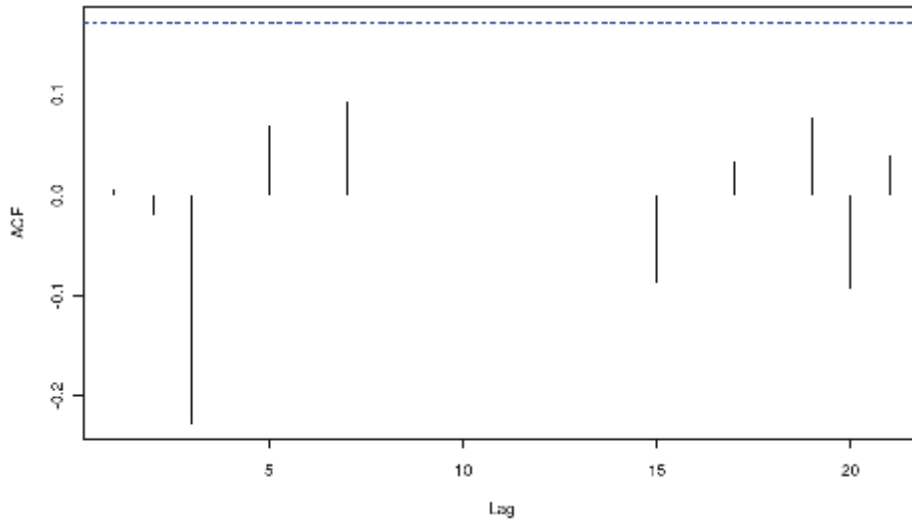
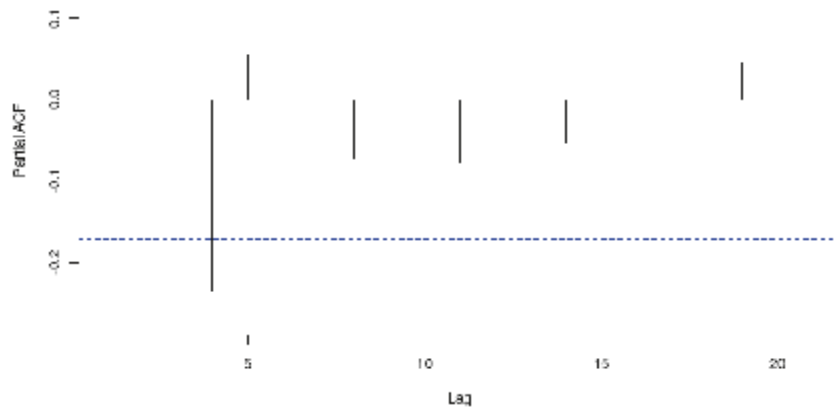
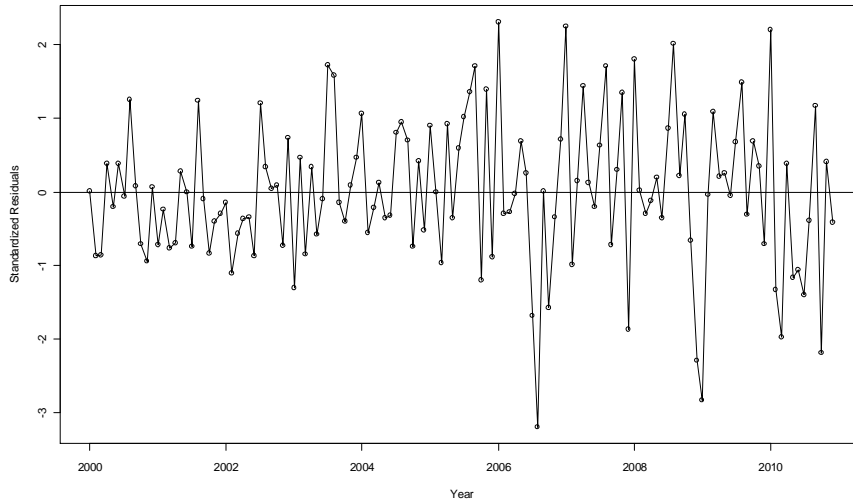


Figure 4.5: Sample PACF for Difference of Electricity Consumption Series

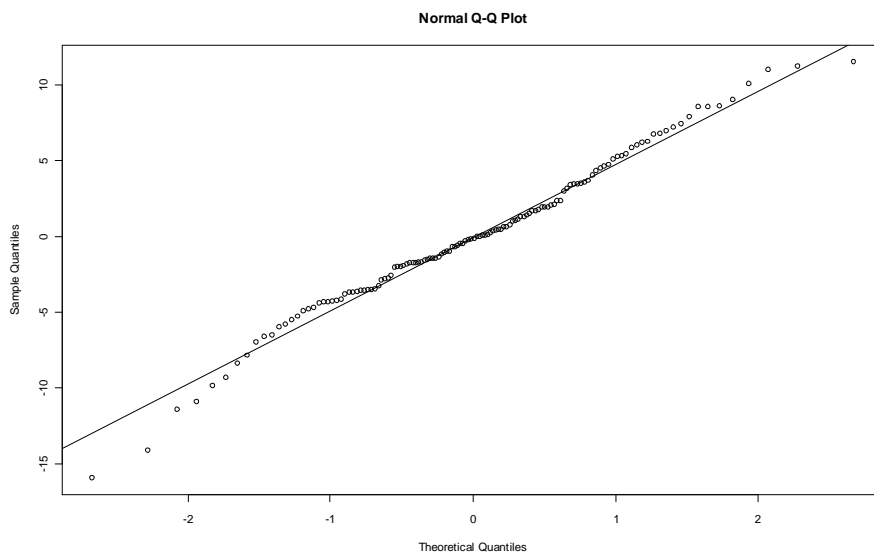


## Artificial Neural Networks Approach to Time Series

**Figure 4.6:** Standardized Residuals of the Fitted Model from Electricity Consumption ARIMA (1,1,4) Model



**Figure 4.7:** Quantile-Quantile Plot of the Residuals of the Fitted Model from Electricity Consumption ARIMA (1,1,4) Model



**Figure 4.8:** Sample ACF of Residuals of the Fitted Model ARIMA(1,1,4) Model

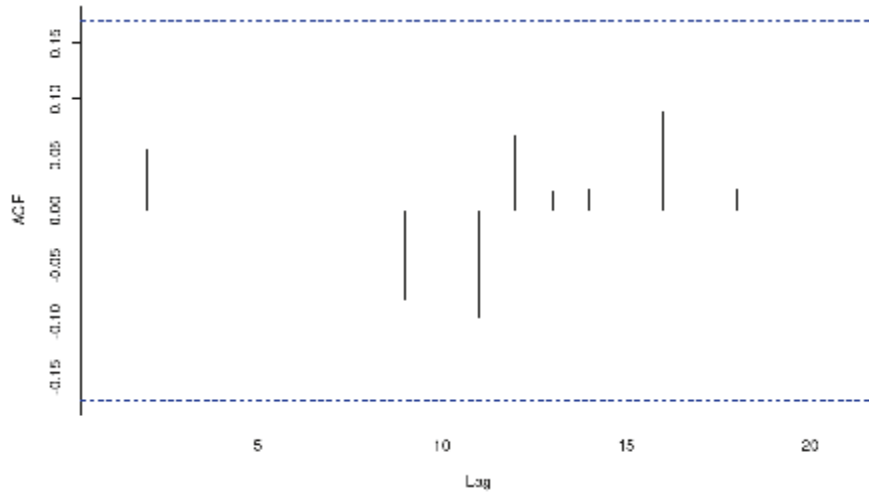
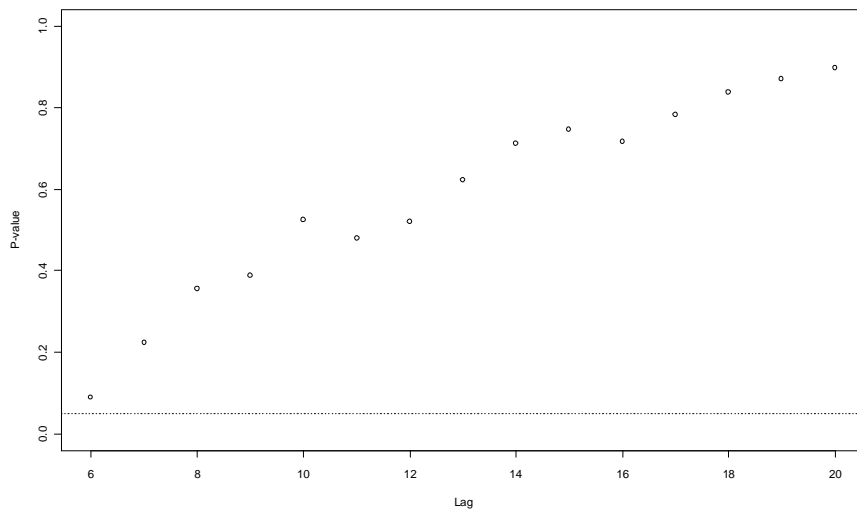


Figure 4.9: P-values for the Ljung-Box Test for the Fitted Model



## References

1. Akaike, H. (1973). "Maximum likelihood identification of Gaussian auto-regressive moving-average models." *Biometrika*, **60**, 255–266.
2. Barbounis, T.G., Theocharis, J.B. (2007). "A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation." *Neuro-computing*, **70** (7/9):1525-42.
3. Blanco, A., Pino-Mejías, R., Lara, J., Rayo, S. (2013). "Credit scoring models for the microfinance industry using neural networks: Evidence from Peru." *Expert Systems With Applications*, **40**(1), 356-364.

## Artificial Neural Networks Approach to Time Series

4. Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
5. Box, G. E. P. and Pierce, D. A. (1970) . “Distribution of residual correlations in autoregressive-integrated moving average time series models.” *Journal of the American Statistical Association*, **65**, 1509-1526.
6. Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994) . *Time series Analysis Forecasting and Control*, 3rd ed. New Jersey: Prentice Hall.
7. Cadenas E. and Rivera W (2009). “Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks.” *Renew Energy*, **34**(1):274-8
8. Cryer, J. D. and Chan, K. S. (2008). *Time series Analysis With Applications in R*, 2nd ed. New York: Springer.
9. Hornik, K., Stinchcombe, M., and White, H. (1989). “Multilayer Feedforward Networks are Universal Approximators.” *Neural Networks*, **2**, 359–366.
10. Hurvich, C. M. and Tsai, C. L. (1989). “Regression and time series model selection in small samples.” *Biometrika*, **76**, 2, 297–307.
11. Hyndman, R. J., and Koehler, A. B. (2005). *Another Look at measures of forecast accuracy*, Monash University, Australia.
12. Jones, R. H. (1975), “Fitting Autoregressive.” *Journal of American Statistical Association*, **70**, 590-592.
13. Khashei, M. and Bijari, M. (2011). “A novel hybridization of artificial neural networks and ARIMA models for time series forecasting.” *Applied Soft Computing*, **11**: 2664 2675.
14. Kozhan, R., (2010). *Financial Econometrics with Eviews*. Ventus Publishing ApS.
15. Kwiatkowski, D., Phillips, P., Schmidt, P. and Skin Y. (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root.” *Journal of Econometrics*, **54**, 159-178.
16. Lee, T. S. and Chen, I. F. (2005). “A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines.” *Expert Systems with Applications*, **28**(4), 743–752.
17. Lee, T. S., Chiu, C. C., Lu, C. J. and Chen, I. F. (2002). “Credit scoring using the hybrid neural discriminant technique.” *Expert Systems with Applications*, **23**(3), 245–254.
18. Liu, H., Chen, C., Tian, H., and Li, Y. (2012). “A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks.” *Renewable Energy*, **48**:545-556.

19. Ljung, G. M. and Box, G. E. P. (1978). "On a measure of lack of fit in time series models." *Biometrika*, **65**, 553-564.
20. Mabel, M. and Fernández, E. (2008). "Analysis of wind power generation and prediction using ANN: a case study." *Renew Energy*, **33**(5):986-92.
21. Majhi, B., Rout, M., Majhi, R., Panda, G., and Fleming, P. (2012 ). "New robust forecasting models for exchange rates prediction." *Expert Systems with Applications*, **39**:12658–12670.
22. Moreno, J., Pol, A. and Gracia, P. (2011 ). "Artificial Neural Networks Applied to Forecasting Time Series." *Psicothema*, **23**(2):322–329.
23. Nazzal, J., El-Emary, I. and Najim, S., 2008, "Multilayer Perceptron Neural Network (MLPs) For Analyzing the Properties of Jordan Oil Shale." *World Applied Sciences Journal* , **5**(5): 546-552, IDOSI Publications.
24. Pousinho H.M.I., Mendes V.M.F., and Catalão J.P.S. (2012). "Short-term electricity prices forecasting in a competitive market by a hybrid PSO–ANFIS approach." *International Journal of Electrical Power & Energy Systems*", **39**(1): 29–35.
25. Schwarz, G. (1978). "Estimating the Dimension of a Model." *Annals of Statistics*, **6**, 461-464.
26. Shibata R. (1976), "Selection of Order of an Autoregressive Model by Akaike's Information Criterion." *Biometrika*, **63**, 117-126.
27. Tsay, R. S. and Tiao, G. (1984). "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA Models." *Journal of the American Statistical Association*", **79**, 385, 84–96.
28. Unsihuay-Vila C, Zambroni de Souza AC, and Marangon-Lima JW. (2010). "Electricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model." *International Journal of Electrical Power & Energy Systems*, **32**(2):108–16.
29. Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer.
30. Walter, H. and Michael T. (2005). "Recent Developments in Multilayer Perceptron Neural Networks." *Proceedings of the 7th Annual Memphis Area Engineering and Science Conference, MAESC*.