SARIMA and ANN approaches in day-ahead power consumption forecasting

Igor Jovanović, Andrija Petrušić, Miona Andrejević Stošović and Dragan Mančić

Abstract - In this paper, we will present two models, a seasonal autoregressive integrated moving average model (SARIMA) and an artificial neural network (ANN) model in order to forecast hourly electricity consumption of energy in industry for a day-ahead. We will start with a brief analysis of the global electricity market with special reference to the Serbian market. Next, the daily electricity consumption amounts between August 1st and December 19th 2019. will be analyzed using statistical tools. According to the obtained results, we will compare the two models.

Keywords – ANN, SARIMA, energy consumption, forecasting, seasonality.

I. Introduction

Liberalization of electric power sector in the last two decades led to very dynamic wholesale markets and deregulated retail markets. At wholesale market level, traders are focused on price dynamics trends, while on retail level for supplying electricity to final consumers the consumption forecast can make all the difference for fostering cost-efficiency [1].

Furthermore, forecasting energy consumption on short-term, mid-term and long-term level is essential when defining strategies for production planning and expansion of infrastructure capacities for electric power systems.

A good approach to consumption prediction of individual consumers can have a far-reaching effect on the aggregated level. Strategic decision-making tools are more reliable when based on proper models for determining patterns of individual consumptions for various types of industries.

It can be observed on the example of electric power suppliers as balance responsible parties at liberalized energy markets how important is having an adequate mathematical model for forecasting individual consumption. The balance responsibility refers to the penalty-based financial mechanism developed to force market participants to properly anticipate their energy needs on short-term level.

Electric power supplier is obliged to inform in a day-ahead manner transmission system operator (TSO) how much of electric power needs to be provided on hourly level for all his electricity consumers. This is important, because TSO needs to maintain the balance in real-time between production and consumption for entire energy system, based on the data provided from all balancing parties active on the market. There is a certain tolerance that TSO will allow to balancing parties, but at the end if anticipated hour consumptions are not matching actual consumed energy and the tolerance for imbalance goes above established limits, the balancing party will be penalized through settlement prices defined for each hour retroactively based on pre-established methodology issued by Regulatory body.

In liberalized power markets, balance responsibility and imbalance settlement are two closely related elements that constitute the essential part of a balancing market, which is an institutional arrangement establishing market-based balancing [2].

In the end, market participant proportionally to the imbalance they created in the system for each hour pays generated imbalance costs. If Electric power supplier lacks adequate forecasting tool and gives poor judgment on how much electricity is needed in each hour for all its consumers that are in his balancing group, unnecessary expense occurs for each estimation that at the end is above the allowed imbalance limits for measured consumption.

If consumption of the balancing group is observed only on an aggregated level, many studies have shown that there is a correlation with outside temperature, but balancing group dynamics is mostly out of focus. Therefore, the smaller balancing group is (fewer metering points and less volume) and if bigger, the fluctuation of client is on yearly level, there is more need for Electric power supplier to estimate balancing group needs through forecasting models of individual energy consumptions for each consumer or metering point.

The importance of electricity consumption forecasting on one hand, and its complexity on the other hand, have motivated many researchers in this area. In the literature, there are numerous studies on electricity consumption and demand estimation. In these studies, some of commonly used methods are stationary time series models [3], regression models [4], [5], [6] econometric models [7]. However, most of time series models are linear predictors, while electricity consumption is inherently a nonlinear function. So, the behavior of electricity consumption series may not be completely captured by the time series techniques. To solve this problem, some other research papers have proposed Artificial Neural Networks (ANN) and genetic algorithms for electricity consumption

Igor Jovanović, Andrija Petrušić, Miona Andrejević Stošović and Dragan Mančić are with the University of Niš, Faculty of Electronic Engineering, E-mail: igor.jovanovic@elfak.ni.ac.rs, andrija.petrusic@gmail.com, miona.andrejevic@elfak.ni.ac.rs, dragan.mancic@elfak.ni.ac.rs.

forecasting [8], [9].

The main purpose of this study using measured hourly-based consumption data is to develop an adequate forecasting model for predicting electricity consumption for individual consumers that can be used as a proper tool by Electric Power Supplier to avoid unnecessary imbalance costs to properly estimate energy needs on an hourly level.

For the purpose of this study, United Green Energy Ltd. [10] consulting company for Energy market has provided data of an individual consumer. Data was strategically chosen so irregular intermittent consumption profile can be used as the input data set. Furthermore, this consumption profile lacks correlation with global factors (e.g. outside temperature), as it is mostly dependable on the business potential company that creates for itself.

Collected hourly data is presented in the form of the time series for the period August 1st –December 19th 2019.

II. Forecasting Models

A. Seasonal Autoregressive Integrated Moving Average Model

Autoregressive Integrated Moving Average Model (ARIMA) is a widely used time series analysis model in statistics. ARIMA(p, d, q) is a kind of short-term prediction model in time series analysis, where p, d and q are non-negative integers that correspond to the order of the autoregressive, integrated and moving average parts of the model, respectively. The ARIMA model is also applicable for non-stationary time series that have some clearly identifiable trends.

The periodicity, which has an impact not only in the electricity market but also in other sectors of the economy, is an important indicator for planning and policy-making [6]. Periodicity of the periodical time series is usually due to seasonal changes. For such time series, a Seasonal ARIMA (SARIMA) is used. Therefore, SARIMA(p, d, q)(P, D, Q)$_m$ is used for time series with seasonality, where P, D and Q are relevant seasonal autoregressive parameter, seasonal integrated parameter and seasonal moving average parameter, respectively, and m is period of time series. The seasonal part of the model is very similar to the non-seasonal part, but it is involved in backshifts of the seasonal period.

Fig. 1 displays the time series plot for electricity consumption data in kWh. The series displays considerable fluctuations over time, and a stationary model does not seem to be reasonable.

To stabilize the variance, data need to be rescaled using the formula:

$$a_i = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)}, \quad (1)$$

where a_i is the rescaled value, y_i represents the original data, and $\min(y_i)$ and $\max(y_i)$ are the minimum and maximum values of the original data set, respectively. After stabilization of variance, the standard deviation is reduced from 22.1195 to 0.1991.

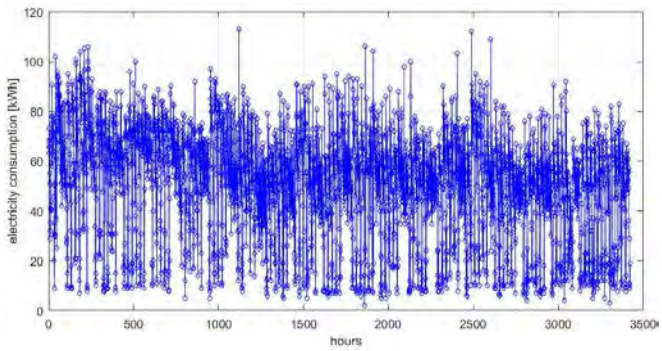

Fig. 1. The hourly electricity consumption from August 1st –December 19th 2019

The sample autocorrelation function (ACF) for the rescaled data is displayed in the Fig. 2. From that figure, one can see that there is a seasonal period of time (24 lags). In addition, software implementation of an augmented Dickey–Fuller (ADF) test for level stationarity applied to the rescaled data leads to a test statistic of −9.3863 for $p<0.01$. Although the test showed that the data was stationarity, a seasonal difference of the rescaled data series must be applied.

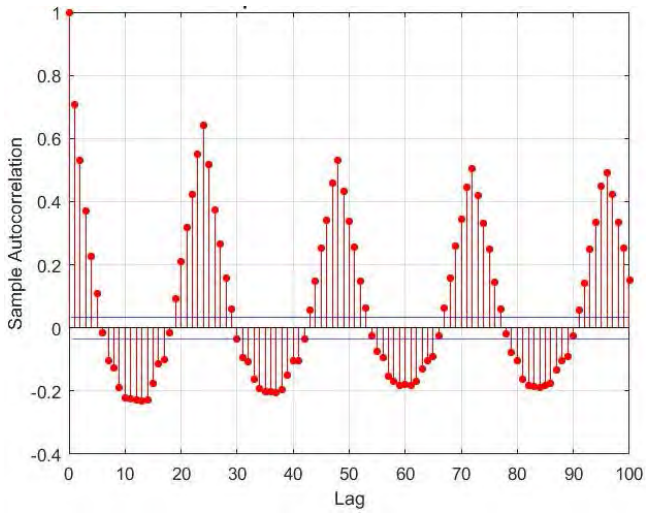

Fig. 2. The sample ACF for the rescaled electricity consumption data

The sample ACF and partial autocorrelation function (PACF) for the seasonal differences of the rescaled electricity consumption data are shown in Fig. 3.

On the basis of this plot, one can well consider a stationary model as appropriate. After first differencing, the

standard deviation is reduced to 0.1711, and ADF test for level stationarity applied to data leads to a test statistic of −87.5573 for $p<0.01$. Taking the second seasonal differences also results in stationarity (i.e., the trend is removed), but leads to an over differences series with a standard deviation which moved to a level higher than the series after first differencing (σ^2=0.4876).

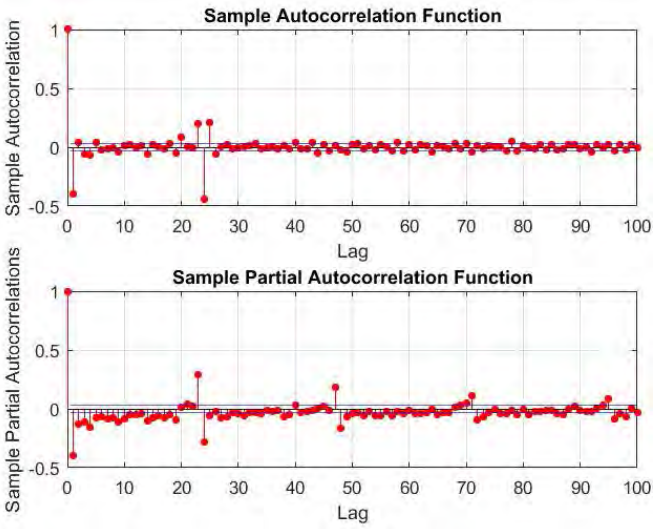

Fig. 3. Sample ACF and PCF for seasonal differences of the rescaled electricity consumption data

After the order of seasonal I(D) terms has been identified (here, 1), in SARIMA(p,d,q)(P,D,Q)$_m$, the next step is to determine whether the pattern of autocorrelation can be better explained by (S)AR terms, (S)MA terms, or a combination of both.

ACF indicating that there are two significant seasonal and non-seasonal spikes and PACF will die out over time. In this case, theoretically, we should use SMA(2) and MA(2) after differencing the data. The orders p and P are selected from a reasonable range of non-negative values to create several SARIMA models.

In order to maximize the probability of obtaining the data that we have observed the best SARIMA(1,0,2)(2,1,2)$_{24}$ model (Table 1) was selected based on the lowest Akaike's Information Criterion value (AIC=-3580). The lower value of Schwarz Bayesian Information Criterion (BIC=-3537.1) and lower root mean square error were desired. The low RMSE indicates that the dependent series is closest to the model predicted levels and thus, the predictive model is useful for forecasting purposes.

Through a combination of regression modelling and one seasonal differencing proposed model successfully extracted all the autocorrelation from the data in order to achieve more efficient forecasts.

The relative success of statistical models in reproducing the measured time-series can also be measured in terms of residuals of error. The frequency distributions of the residuals of the SARIMA models for all data are presented in Figure 4. The points of a quantile-quantile plot of the residuals 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 Ljung-Box Q-test for residual autocorrelation applied to the residuals, produces a test statistic of 66.8844, which corresponds to a p-value <0.01, 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 ACF of the residuals in Figure 5. In other words, the residuals follow a linear trend. Thus, the residuals are normally distributed. In general, the model shows good forecasting accuracy and can be used to predict future values.

TABLE I
PARAMETERS VALUE OF THE SARIMA(1,0,2)(2,1,2)$_{24}$ MODEL

Parameter	Value
Constant	0.401641
AR{1}	0.032015
SAR{1}	0.00657
SAR{2}	0.005856
MA{1}	0.000825
MA{2}	0.000824
SMA{1}	0.033497
SMA{2}	0.02343

Cross-validation prediction checks of proposed SARIMA model can predict and verify the value for the observed period, and the forecast future trends.

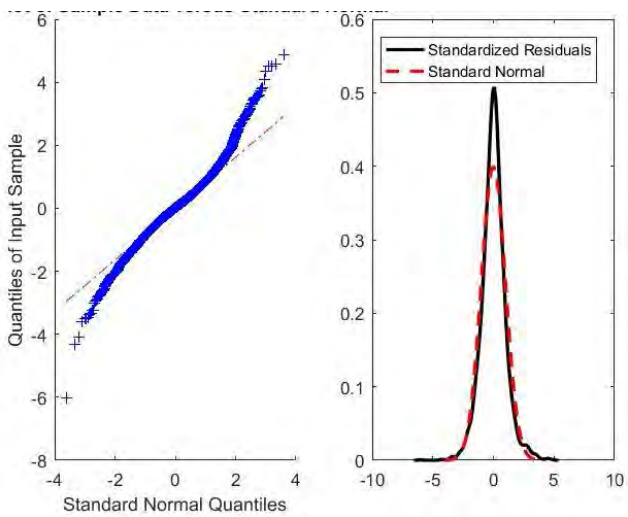

Fig. 4. Plots of the fitted SARIMA(1,0,2)(2,1,2)$_{24}$ model ACF and PCF for seasonal differences of the rescaled electricity consumption data

The model predicts the rescaled electricity consumption

data (testing set, or observed data) for the next 139 days, based on the 37 hours of past data (presample data) with RMSE=19.2581, while minimum presample data for the selected model is 27 hours (Fig. 6). From Fig. 6 it can be concluded that overall model performance is satisfactory for all the electricity consumption data, although the observed peaks are reproduced in some cases with higher amplitude.

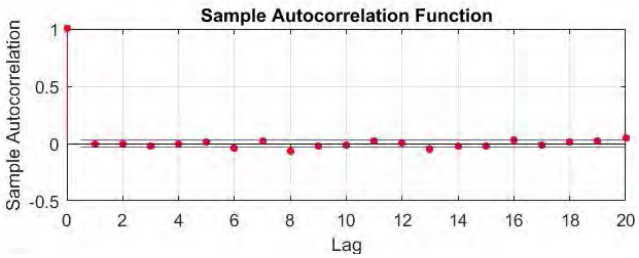

Fig. 5. Sample ACF for the SARIMA residuals

Verified value for the observed period can serve as confirmation of the adequacy of the proposed model, while forecasted future trend for electricity consumption data can be an indication of a future state. Fig. 7 shows a comparison between testing set and training set. For the sake of clarity, Fig. 7 shows values for the presample (black line) and observed (red dotted line) data only for the last four days (the data are already shown in Fig. 6).

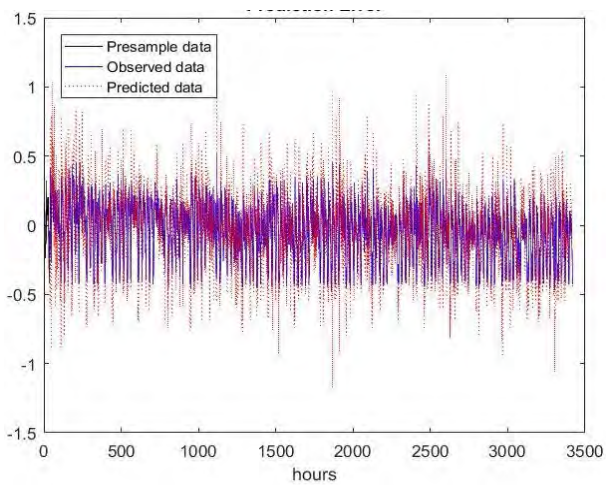

Fig. 6. Diagnostics proposed SARIMA model with rescaled data

Fig. 8 shows forecasted values (red line) for the day-ahead period. Evaluating forecast accuracy is accomplished by examining the residuals for any systematic pattern of misspecification. Forecasts should ideally be located within the 95% confidence limits, and formal statistics can be calculated from the model residuals in order to evaluate its adequacy. In other words, according to Fig. 8, the real values appear within the 95% confidence intervals (between red dotted lines).

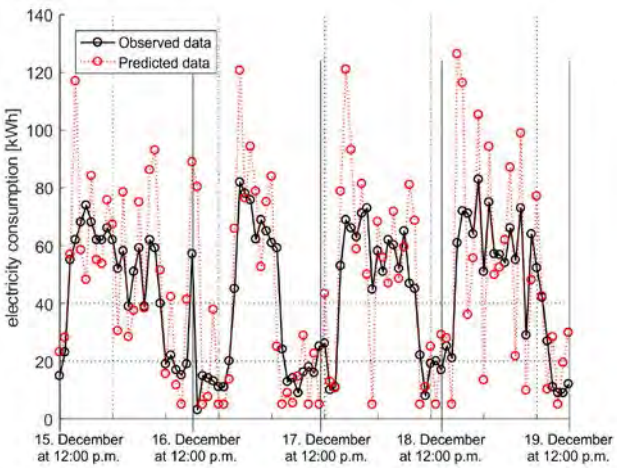

Fig. 7. Diagnostics proposed SARIMA model with original power consumptions data

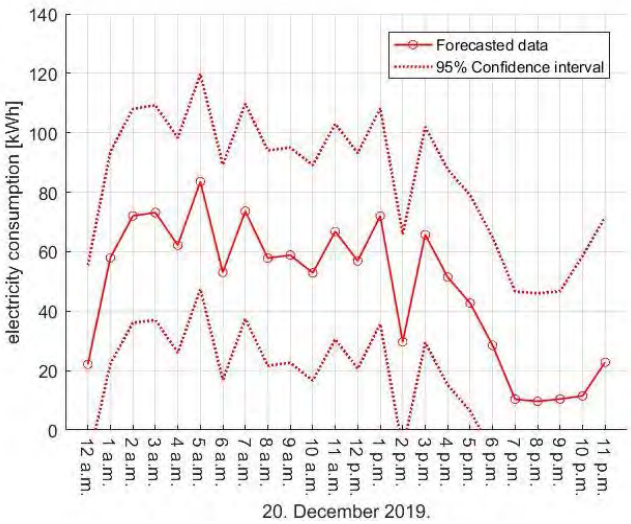

Fig. 8. Forecasted power consumption data using SARIMA model for the day-ahead period

B. ANN approach

Artificial neural networks have been used for short-time forecasting in electronics [11], [12]. Models were also developed for prediction of electricity consumption on a monthly and weekly level.

In this paper we will present a model based on a recurrent neural network. In fact, we used data from 24 previous measurements (measurement is done every hour), so we use consumption from a day before in order to predict consumption in the next hour. So, the artificial neural network has 24 inputs, describing what was happening for the previous 24 hours. For the network training we used the next consumption value as a network output. The obtained network has one hidden layer with 40 hidden neurons. After the training was completed, a

training error was $1.23*10^{-17}$, what is negligible. That means that when data used in the training process are used as an excitation to the obtained ANN, no error is occurred.

In the Fig. 9 we present forecasted data for the next day, i.e. for next 24 hours. Forecasted data is obtained when we excited ANN with values for next 24 hours. We also should stress that part of the data used for prediction are also predicted in earlier iterations.

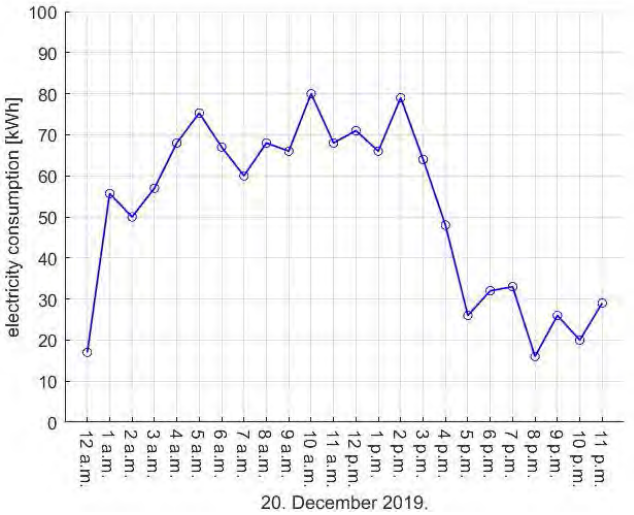

Fig. 9. Forecasted power consumption data using ANN model for the day-ahead period

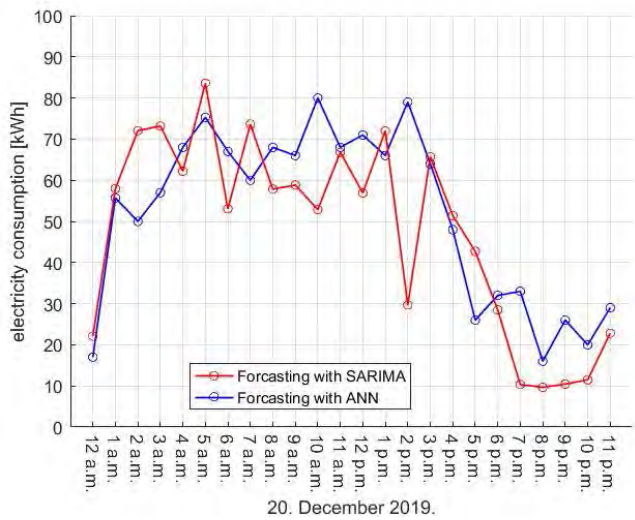

Fig. 10. Comparison of forecasted power consumption data using SARIMA and ANN model for the day-ahead period

III. Conclusion

In this paper, two approaches, statistical and neural network approach, were deployed for obtaining models for time series, i.e. day-ahead power consumption forecasting.

By analyzing and comparing forecasted data obtained by these two approached, we can see from the Fig. 10 that forecasted data have the same trend. We can also notice from the paper that ANN approach is much simpler.

It can be seen from the measured data that the energy consumption is correlated with outside temperature and with few more factors related to the daily operations of the observed company. In our future work, we plan to do further research and develop some hybrid forecasting methods which will take into account most of the above parameters.

Acknowledgement

The research presented in this paper is financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia under the projects TR33035 and TR32004.

References

[1] Pepermans, G., "European energy market liberalization: experiences and challenges", *IJEPS*, Vol. 13, pp. 3–26, 2019.

[2] Van der Veen, R. A. C., Hakvoort, R. A., "Balance responsibility and imbalance settlement in Northern Europe — An evaluation", *6th International Conference on the European Energy Market*, Leuven, pp. 1-6, May 2009.

[3] Fosso, O. B., Gjelsvik, A., Haugstad, A., Mo B., Wangensteen, I., "Generation scheduling in a deregulated system. The Norwegian case", *IEEE Transactions on Power Systems*, Vol. 14, No. 1, pp. 75-81, February 1999.

[4] Bianco, V., Manca, O., Nardini, S., "Electricity consumption forecasting in Italy using linear regression models", *Energy*, Vol. 34, No. 9, pp. 1413-1421, September 2009.

[5] Abdel-Aal, R.E., Al-Garni, A.Z., "Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis", *Energy*, Vol. 22, No. 11, pp. 1059-1069, November 1997.

[6] Akdi, Y., Gölveren, E., Okkaoğlu, Y., "Daily electrical energy consumption: Periodicity, harmonic regression method and forecasting", *Energy*, Vol. 191, January 2020.

[7] Gori, F., Takanen, C., Forecast of energy consumption of industry and household & services in Italy", *International Journal of Heat and Technology*, Vol. 22, No. 2, pp. 115-121, January 2004.

[8] Azadeh, A., Ghaderi, S.F., Sohrabkhani, S., "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors", *Energy Conversion and Management*, Vol. 49, No. 8, pp. 2272-2278, August 2008.

[9] Amber, K.P., Aslam, M.W., Hussain, S.K., "Electricity consumption forecasting models for administration buildings of the UK higher education sector", *Energy and Buildings*, Vol. 90, pp. 127-136,

March 2015.
[10] http://www.uge.rs/
[11] Milojković, J.B., Litovski, V.B., "Short term forecasting in Electronics", *International Journal of Electronics*, Vol. 98, No. 2, pp. 161-172, February 2011.
[12] Milojković, J, Litovski, V.B., "Dynamic Short-Term Forecasting of Electricity Load Using Feed-Forward ANNs", *Int. J. of Engineering Intelligent Systems for Electrical Engineering and Communication*, Vol. 17, No. 1, pp. 38-48, March 2009.