I J C T A, 9(26) 2016, pp. 417-424

Genetic ARIMA (GARIMA): A Fuzzy based ARIMA Model for Time Series Forecasting

B. Arputhamary[1], Z. Asha farhath[2] and L. Arockiam[3]

ABSTRACT

The model of prediction and forecasting has existed from time traditional. Only most recently scientific methods have been involved in the process in time series models. Such forecasting when done through the modern techniques like ARIMA naturally improves the accuracy which is the primary focus of this study. In this work, a fuzzy based time series model Genetic ARIMA is proposed to compare with the ARIMA model and the results show that the proposed model (Genetic ARIMA) performs better than the existing ARIMA Model. This elaborates on the choices to be made for the appropriate dataset and also found that error rates are considerably less when compared with the earlier models.

Keywords: ARIMA Models, Forecasting, Fuzzy, Prediction, Time Series Models.

1. INTRODUCTION

Time-series methods make forecasts or predictions based on the historical patterns found in the recorded data in the past. Such time-series methods utilizes the time as an independent variable while the factors are in the *y* plane. For any time series, such *y* measurements are taken at successive points (intervals) or over successive periods of time may be equal or non-equal. Such as the regarding measurements may be taken every hour, day, week, month, year or any other regular or irregular interval of time.

Therefore the first step in using such time-series approaches is to gather the historical data of the object in focus. This historical data is representative of the conditions expected in the future. Time series models are sufficient to forecasting if demand it has shown a consistent pattern in the past that is expected to occur in the future. Such as, new homebuilders in US may see variation in sales from month to month, but analysis of past years data may expose that the sales of new homes are increased gradually over the period of time. In this case trend is increased in new home sales. Time series models are categorized into four components; they are cyclical component, trend component, seasonal component and irregular component. Trend is important characteristics of time series models. Even if times series may display trend, there might be data points lying above or below trend line. Any recurring sequence of points are above and below the trend line that last for more than a year is considered to represent the cyclical component of the time series that is these observations in the time series deviate from the trend due to fluctuations. The component of the time series data that captures the changeability in the data due to seasonal fluctuation is called the seasonal component. The seasonal component is similar to the cyclical component in that they both refer to some regular fluctuation in a time series. Seasonal components are capturing the regular pattern of variability in the time series with in one-year periods. Seasonal commodities are best example for seasonal components. Random variation in

[1] Research Scholar, Department of Computer Science, Mother Teresa Women's University, Kodaikanal, India. *E-mail: arputhambaskaran@rediffmail.com*

[2] Research Scholar, Department of Computer Science, Bishop Heber College, Tiruchirappalli, India. *E-mail: asha.zakir@yahoo.com*

[3] Associate Professor, Department of Computer Applications, St. Joseph's College, Tiruchirappalli, India. *E-mail: larockiam@gmail.com*

the times series are represented by the irregular component. The irregular component of the time series cannot be predicted in advance. The random variation in the time series is caused by short term, unexpected and non recurring factors that affect the time series.

The focus is here to propose popular time series model Genetic ARIMA based on fuzzy for the same dataset and then measures the parameters like prediction accuracy, time consumed, and overheads. The Genetic model's prediction is more accurate than the normal ARIMA model. In that the limitation of the requirements of huge historical data in a short span of time is overcome by combining the genetic model which utilizes the fuzzy regression model in combination with an ARIMA model to work with little historical data. Combination of these two models the practical limitation of requiring huge historical data is overcome and also the overheads are significantly less naturally with increased speed and accuracy.

2. RELATED WORKS

Johnston and Harrison [1] found forecast variances for the simple and Holt exponential smoothing methods for state space models with multiple sources of errors. W. Jacobs *et al* [2] artificial neural network models and compare the results to the individual models, exemplify the combined forecast for the production planning. Stylianos I. Vagropoulos *et al* [3] in compares four practical methods for electricity generation forecasting of grid-connected Photovoltaic (PV) plants, that is Seasonal Autoregressive Integrated Moving Average (SARIMA) model, SARIMAX model (SARIMA modeling with exogenous factor),modified SARIMA model, as a result of an a posteriori modification of the SARIMA model, and ANN-based modeling.

Rodrigo N. Calheiros *et al* [4] presented the realization of a cloud workload prediction module for SaaS provider based on the autoregressive integrated moving average (ARIMA) model. Ling Wang *et al* [5] proposed a simple and effective prediction method for metro wheel wear based on the time series modeling ARIMA (p, d, q) model in "Wear prediction of metro wheels based on the ARIMA model".Takaomi HIRATA *et al* [6] a novel prediction method which composes not only a kind of DBN with RBM and MLP but also ARIMA to study of nonlinear phenomenon.

Guoqiang Liu [7] proposed a hybrid model which consists of two methods, Singular Spectrum Analysis (SSA) and Auto Regressive Integrated Moving Average (ARIMA) for forecasting medium and long-term software failure time. Sornpon Wichaidit *et al* [8] predicted short-term stock prices of SET50 of Stock Exchange of Thailand using CARIMA (Cross Correlation Autoregressive Integrated Moving Average) and the results of CARIMA model yield better price trends.Vaccaro *et al* [9] Local Learning-ARIMA adaptive hybrid architecture for hourly electricity price forecasting in proposed hybrid architecture for electricity price forecasting. The proposed architecture combines the advantages of the easy-to-use and relatively easy-to-tune Autoregressive Integrated Moving Average (ARIMA) models and the approximation power of local learning techniques. Shanzhi Li *et al* [10] proposed hybrid model which combine an Autoregressive Integrated Moving Average (ARIMA) model and a Radial Basis Function Neural Network (RBFNN) model where the result of fuzzy-neural network combination method reduce both mean square and mean absolute errors.

3. PROPOSED MODEL

The proposed model of genetic fuzzy ARIMA is implemented on the basis of calculating the forecast values by taking the number of steps to be predicted. In this section the Genetic ARIMA (GA) is proposed to consider the least amount of data's will be predicted and produce the future accuracy values. It takes fuzzy logic and it should be predicted the lower bound and upper bound linear values. In genetic arima to reduce the time taken, overheads, RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error) is compared to the Arima. Further the RMSE and MAPE values for the airline passenger dataset are computed for both the models and displayed proving the latter model to accurately. The output graph shows the plot moving average or moving variance and it will visible, if it varies with time. We will take the average/ variance of the last year that is last 12 months. This is one of the statistical tests for checking stationary.

The membership functions of the dataset that represents prediction or target parameters which is at the center of the number where a high and low are captured while an average is taken. The weight of the predicted value depends on the relation of time lag and the present observation. Fitting the ARIMA (p, d, q) by using the available information of observations, *i.e.*, input data is the optimum solution of the parameters and residuals. The number of constraint functions is the same as the number of observations. Next delete the data around the model's upper bound and lower bound when the Genetic ARIMA (fuzzy) model has outliers with a wide spread. Finally, delete the data around the model's upper and lower bound when the fuzzy ARIMA model has outliers with a wide spread. After forecast the values have been generated, the error rates for mean absolute percentage error, root mean square value are calculated from the forecasted values, which are tabulated and graphed. Genetic ARIMA is compared to ARIMA it requires the accurate value of the forecasted data. It also calculated the error rates are better than ARIMA model.

A. The Arima Model does the Following

- In ARIMA (p, d, q) p is the number of autoregressive terms.
- d is the number of non-seasonal differences needed for stationary.
- q is the number of lagged forecast errors in the prediction equation.

To identify the appropriate ARIMA model for Y, begin by determining the order of differencing (d) needing to stationeries the series and remove the gross features of seasonality, perhaps in conjunction with a variance-stabilize transformation such as logging or deflating. If stop this point and predict the differenced series are constant, have only fitted a random walk or random trend model. However, the stationarized series may still have auto correlated errors, suggesting that some number of AR terms ($p \geq 1$) and some number MA terms ($q \geq 1$) are also needed in the forecasting equation.

B. Genetic Fuzzy Arima Algorithm

Step 1: Input N Dataset

Step 2: Find Intervals for i periods

Step 3: Cleanse

If interval invalid

Remove data point d*p*

Else if interval valid and no data value

Add data d*i* based on historical data

Step 4: Input to Fuzzy *F* for *i* points

$$\Sigma(x) = 1/(1 + e\wedge - x)$$

$$\Sigma(x) + \Sigma(-x) = 1$$

$$(\Sigma(x) + \Sigma(-x)) * (\Sigma(y) + \Sigma(-y)) * (\Sigma(z) + \Sigma(-z)) = 1$$

Where x is for data points across y time period i

Step 5: Fuzzy set $(..., x\langle x_n, f_n(x), x = x_{n+1}, y_{n+1}, ...)$ interpolates the interval $[x_n, x_{n+1}]$ with a linear n connecting the points $(x_n, f_n(x_n))$ and (x_{n+1}, y_{n+1})

Step 6: Take average values and forecast.

Step 7: Find the error rates,

$$MAPE = \frac{1}{n}\sum \left| Actual - Forecast \right|$$

Step 8:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(x_{1,t} - x_{2,t})^2}{n}}$$

Step 9: Forecast with Actual Error Percentages.

C. Algorithm Steps

- First the input dataset is loaded and it is fit into the ARIMA model where the noise is identified and cleared.
- Next the incomplete dataset is filled up using fuzzy logic.
- This leads to dataset values of intervals with upper and lower bound values.
- The upper bound and the actual values averages are got.
- Next the upper wide area is cleared, similarly for the lower wide area.
- This area is deleted and the middle value is got and the regression model is then applied with ARIMA model by taking the average values.
- This yields the forecast of the period required.
- Next calculate the error percent using MAPE and RMSE.

D. Configuration Setup for Genetic Arima

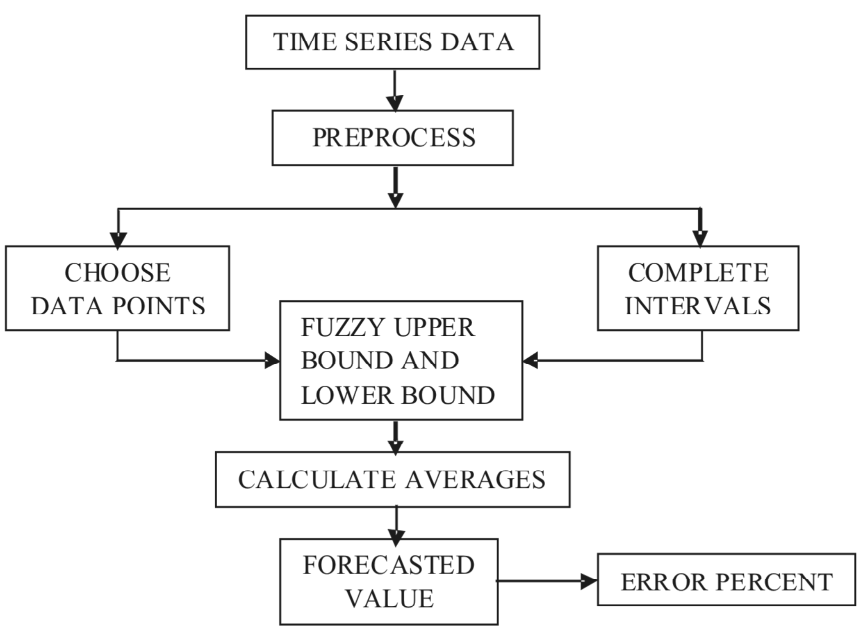

Figure1: Proposed methodology of Genetic ARIMA

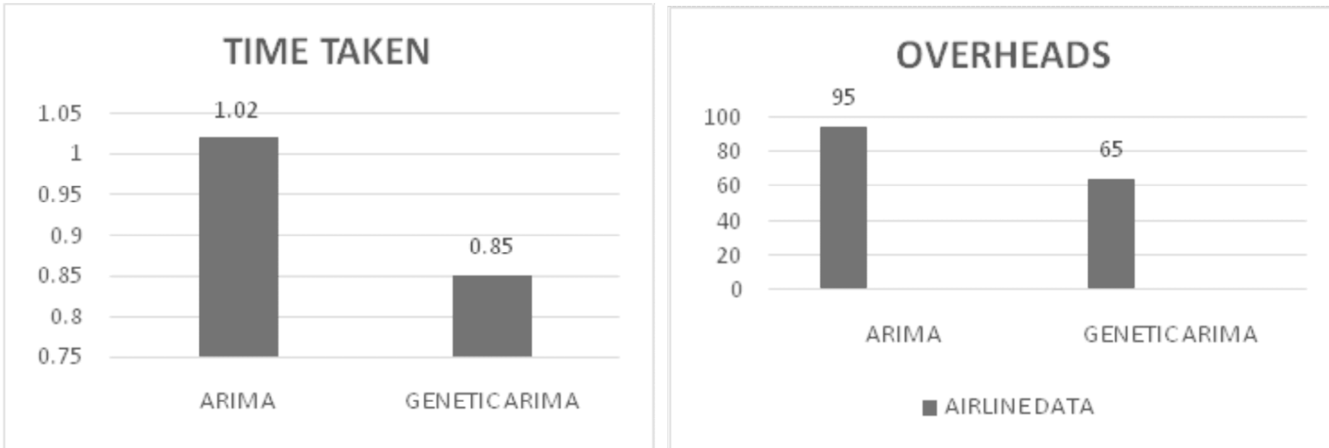

Figure 2: Graphs shows the performance of the Genetic ARIMA is better than ARIMA.

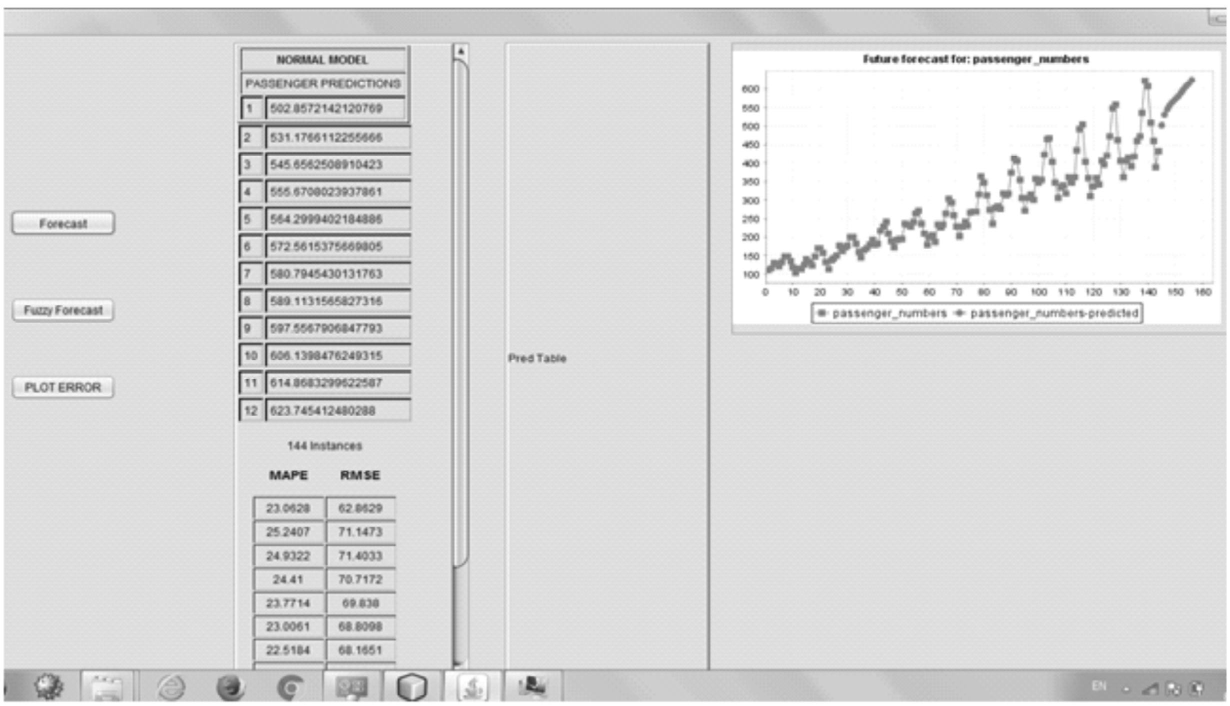

Figure 3: Prediction for ARIMA forecasting model

4. EVALUATION AND DISCUSSION

The predictive ability of the fuzzy ARIMA is rather encouraging and the possible interval of the Genetic ARIMA is narrower than 95% of the confidence interval of normal ARIMA. It has the tendency to increase in the confidence interval. The time taken for the algorithm to complete the forecasting is calculated based on the System. Nano method in milliseconds which are recorded both before and after the process is completed. This millisecond is divided by 1000 to get the seconds taken and then the values are plotted on a graph.

In this figure represent to predict the forecasting values of airline passenger dataset using ARIMA model and calculate the MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Squared Error).

In this figure represent to predict the future accurate forecasting values using fuzzy and it finds the Error rates also.

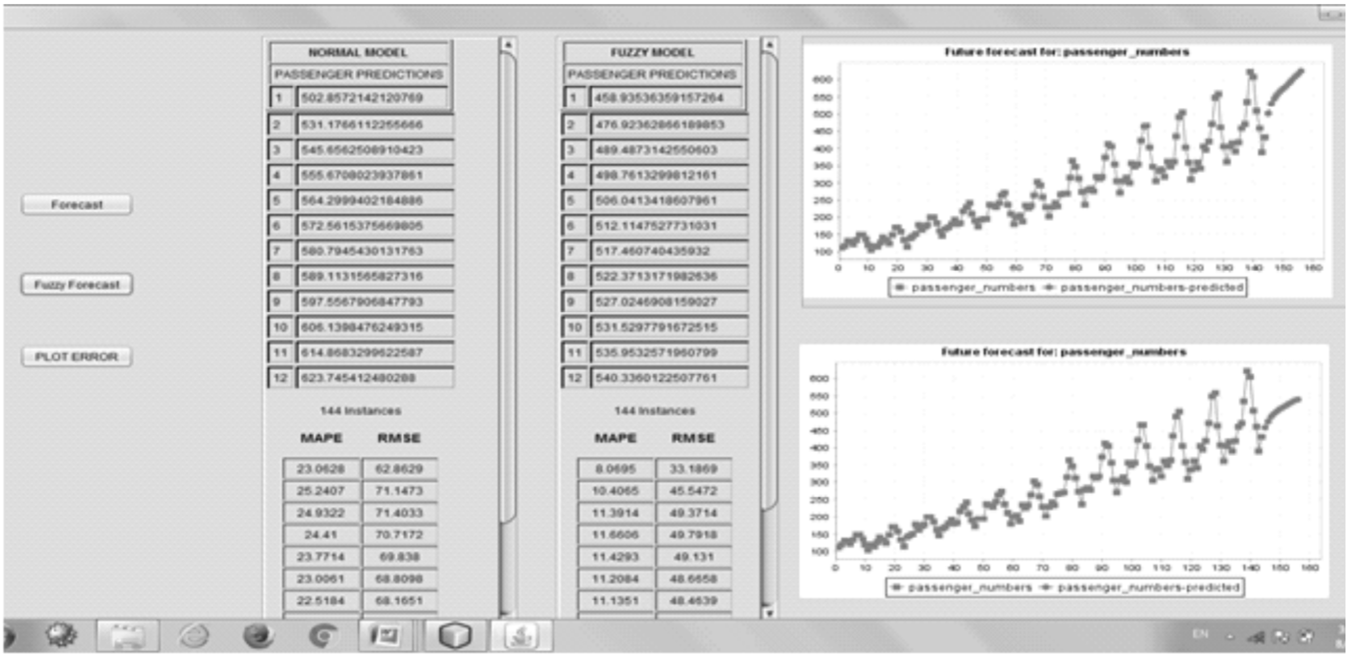

Figure 4: Prediction for Genetic ARIMA forecasting model (fuzzy)

Table 1
Comparison of ARIMA – Genetic ARIMA Error rates

Error Rate			
RMSE		*MAPE*	
Arima	*Genetic Arima*	*Arima*	*Genetic Arima*
62.8629	33.1869	23.0628	8.0695
71.1473	45.5472	25.2407	10.4065
71.4033	49.3714	24.9322	11.3914
70.7172	49.7918	24.41	11.6606
69.838	49.131	23.7714	11.4293
68.8098	48.6658	23.0061	11.2084
68.1651	48.4639	22.5184	11.1351
67.8049	48.3694	22.2103	11.0099
67.4495	48.3416	21.9021	10.9464
66.9412	48.442	21.4945	10.9726
66.0754	48.5931	20.8209	10.9645
64.7833	48.696	19.8619	10.7951

MAPE (Mean Absolute Percentage Error) = $\frac{1}{n}\sum|\text{Actual} - \text{Forecast}|$. It calculates the Mean Absolute Percentage Error with the actual and forecasted values. RMSE (Root Mean Squared Error) = $\sqrt{\frac{\sum_{t=1}^{n}(x_{1,t} - x_{2,t})^2}{n}}$. It calculates the Root Mean Squared Error with forecasted prediction values, and it gives the accurate values of the Error rate.

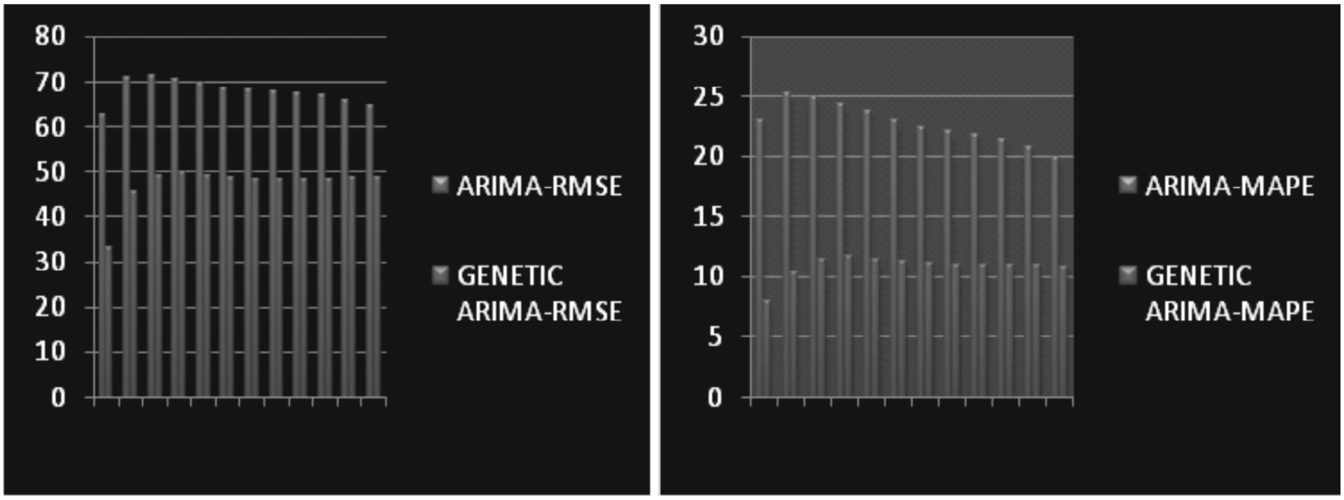

Figure 5: Error rates for the ARIMA and Genetic ARIMA.

5. CONCLUSIONS

The proposed model requires considerably lesser observed historical data than the ARIMA model. In addition to the numerous advantages outlined above the proposed model gives both scenarios *i.e.* the best and the worst etc. Also with a less observations the prediction accuracy is high while the time consumed and overheads are significantly less compared to the traditional ARIMA model, because of the increased confidence provided. The forecast values are also interpolated with the error rates *i.e.* Root Mean Squared Error (RMSE) and Mean Absolute Percentage (MAPE) and it is found that proposed model is having the least error differential rate when compared with earlier models.

REFERENCES

[1] Melard.G, and Pasteels, "Automatic ARIMA Modeling Including Interventions, Using Time Series Expert Software", *International Journal of Forecasting*, **16**, 497–508, 2000.

[2] M Claudio and S. Rocco, "Singular spectrum analysis and forecasting of failure time series", *Reliability Engineering and System Safety*, **114**, 126-136, 2013.

[3] Johnston, F.R., and Boylan, J.E. "Forecasting Intermittent Demand: A Comparative Evaluation of Crostons Method", *International Journal of Forecasting*, **12**, 297–298, 1996.

[4] W. Jacobs, A. M. Souza and R. R. Zanini – "Combination of Box-Jenkins and MLP/RNA Models for Forecasting Combining Forecasting of Box-Jenkins", *IEEE Latin America Transactions*, **14(4)**, 2016.

[5] S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou and A. G. Bakirtzis, "Comparison of SARIMAX, SARIMA, Modified SARIMA and ANN-based Models for short-term PV generation forecasting," *IEEE International Energy Conference (ENERGYCON)*, Leuven 1-6, 2016.

[6] Rodrigo N. Calheiros ; Dept. of Comput. and Inf. Syst., Univ. of Melbourne, Melbourne, VIC, Australia; Enayat Masoumi; Rajiv Ranjan; Rajkumar Buyya, "Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications", *QoS IEEE Transactions on Cloud Computing* **3(4)**, 2015.

[7] L. Wang, W. Zhao, H. Xu, C. Chen, X. Chen and W. Na, "Wear Prediction of Metro Wheels Based on the ARIMA Model", *The 27th Chinese Control and Decision Conference (2015 CCDC)*, Qingdao, 2697-2701, 2015.

[8] Takaomi HIRATA, Takashi KUREMOTO, Masanao OBAYASHI, Shingo MABU, "Time Series Prediction Using DBN and ARIMA", *International Conference on Computer Application Technologies,* IEEE, 2015.

[9] G. Liu, D. Zhang and T. Zhang, "Software Reliability Forecasting: Singular Spectrum Analysis and ARIMA Hybrid Model", *Theoretical Aspects of Software Engineering (TASE), 2015 International Symposium on Nanjing*, 111-118, 2015.

[10] Sornpon Wichaidit; Dept. of Computer Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Thailand; Surin Kittitornkun "Predicting SET50 Stock Prices Using CARIMA", *International Computer Science and Engineering Conference (ICSEC)*, 2015.

[11] A. Vaccaro, T. H. M. EL-Fouly, C. A. Cañizares and K. Bhattacharya, "Local Learning-ARIMA Adaptive Hybrid Architecture for Hourly Electricity Price Forecasting", *PowerTech, 2015 IEEE Eindhoven*, 1-6, 2015.

[12] Chatfield, C. "Neural Network: Forecasting Breakthrough or Passing Fad", *International Journal of Forecasting*, **9**, 1-3, 1993.

[13] Kalid Yunus, Torbjorn Thiringer, and Peiyuan Chen, "ARIMA-Based Frequency-Decomposed Modeling of Wind Speed Time Series", *IEEE Transaction on Power Systems*, **31(4)**, 2546-2556, 2016.

[14] S. Li, H. Wang, Y. Tian, Y. Shen and A. Aitouche, "Wind Speed Forecasting Based on Fuzzy-Neural Network Combination Method", *The 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao*, 4811-4816, 2015.