2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 24-25 December, Dhaka

Forecasting Electricity Consumption using ARIMA Model

Farjana Mahia†, Arpita Rani Dey*, Md Abdul Masud‡, and Mohammad Sultan Mahmud‡
Faculty of Computer Science and Engineering
Patuakhali Science and Technology University, Patuakhali, Bangladesh
farjanamahia061@gmail.com†, arpita13@cse.pstu.ac.bd*, masud@pstu.ac.bd‡ and sultan@szu.edu.cn ‡

***Abstract*— Autoregressive integrated moving average, ARIMA, is a popular technique, which is used to fit time series data for prediction and forecasting. This paper proposes ARIMA models with different sets of parameters for forecasting electricity consumption. The three ARIMA models, which are quite good and robust to develop a reliable model, are investigated to forecast electricity consumption for providing the required level of performance. The best fitted model, effective and reliable approach, and network structure are determined according to the prediction performance. For this purpose, we use synthetic dataset and electricity consumption data in industries at Guangdong province in China. The experimental results show that the ARIMA(1,1,1) has high precision, stable predictions and suitable for predicting electricity consumption. The forecasting results are essential to manage the required electricity demand in various kind of industries and other sectors.**

Keywords—Auto Correlation Function, Akaike Information Criterion, Partial Auto Correlation Function, ARIMA

I. INTRODUCTION

Electricity is a fundamental necessary factor in our daily life. The energy source becomes a core component for social and economic development and the central source of its usage of a country. Electric power storage is quite impractical and the demand of it can change dramatically in space and time related to different sectors. The forecasting of electricity consumption is an essential issue for utility owners, power system operators, energy planners and system managers. The methods for prediction are chosen by considering different factors including size of the time series, prediction interval, and prediction period [1]. During the last several decade various methods are being used for consumption of electricity to predict the future consumption accurately. The time series data has four components: trend (long term direction), seasonal (systematic, calendar related movements) effect, cyclical and irregular (unsystematic, short term fluctuations) effect [2]. ARIMA is a core forecasting technique to predict the future electric power production which meets the future energy demand. The prediction figure helps to determine the budget and how much electricity should be produced in various sectors including agricultural, transportation, residential, commercial. Forecasting is used to predict the future information by considering previous and present data and analyzed the trends of them. ARIMA models establish a paramount class of models that can be applied to many real applications. It is derived from autoregressive moving average, ARMA. Forecasting electricity consumption using different ARIMA models on real dataset and comparing them to determine the best model gives highly accurate and stable prediction. Electricity forecasting is a challenging task, it can't be predicted 100% accurately. Because forecasting depends on some factors which varies on different sectors, areas, industries etc. Considering electricity consumption of any sector, there are so many attributes that can be chosen for detecting and predicting the consumption of any area. ARIMA model is more accurate than traditional forecasting techniques. It is one kind of statistical model to analyze and forecast time series data. Specially, ARIMA model is also applied to detect patterns and analyze the trends on electricity consumption in household (daily, weekly, monthly and quarterly) [3].

In this paper, we apply ARIMA model to forecast electricity consumption. The electricity consumption raw data from different manufacturing factories at Guangzhou in China 2012 were collected for prediction [4]. The ultimate goal is to predict highly accurate results by estimating reliable ARIMA model.

II. RELATED WORKS

Increasing electricity demand is a key issue nowadays. Prediction of electricity consumption is one of the vital elements for minimizing the waste of electricity. Various types of approaches of prediction have been introduced to predict the consumption of electricity. In this section we briefly explain the existing various prediction procedure. The early methods of electricity consumption forecasting techniques include exponential smoothing models, moving average, autoregressive models etc. The forecasting methods are of three categories: grey prediction models, statistical analysis models and non-linear intelligent models. Non-linear models consist of Support Vector Machine (SVM), Markov Chain and Artificial Neural Network [5].

In drawing precise prediction, GM(1,1) solutions are statistically comprehensive but in volatility order of applications no satisfactory result is provided [6].

978-1-7281-6099-3/19/$31.00 ©2019 IEEE

The data in ANN training may generate output even with lost information. The performance level depends on incomplete data importance. Apposite network structure is obtained through trial, experience and error but no rule is specified to determine the structure of ANN. Artificial neural networks needs modifications before using it on time series data [7].

Hierarchical multi-matrices Markov, HMM, model is used when direction of the next observed point is stated rather than forecasting [8].

SVM, Supervised learning method, have been widely used in time series predicting complications though they have not been broadly explored in seasonal time series forecasting. Only the binary classification problems are solved through standard SVM formulation and output variables are limited to take only binary values [9].

III. METHODOLOGY

ARIMA is a model which is commonly used to forecast and predict future information on time series data. There are different settings of ARIMA model which are used as complementary methods for non-stationary data analysis. In this paper, we use three ARIMA models with different sets of parameters to forecast electricity data. We define ARIMA model with different parameters (*p*, *d*, *q*) where *p*, *d*, *q* represents the number of autoregressive terms, the number of non-seasonal differences, and the number of lagged error values in prediction respectively.

The forecasting of electricity consumption consists of the following steps:

1. Visualize the time series data: It is important to visualize the electricity consumption data to understand the trends, seasonality or random behavior for developing time series model.

2. Test stationary property by Augmented Dickey Fuller Test: The ARIMA model, an ARIMA(*p*, *d*, *q*), works on stationary data. Therefore, after visualizing the electricity consumption data, the stationary property is tested with Augmented Dickey Fuller Test, ADF. The *ADF* test is an advanced model tests where the null hypothesis that a unit root is present in an autoregressive model. The existence of unit roots leads unwanted results in time series analysis, which can cause inaccurate forecasting. The ADF is able to test stationary property and handle more complex statistics than the traditional Dickey-Fuller test.

3. Stationarize the time series data: Dataset should be stationarized if the time series is not stationary. Three methods which are widely used to convert a time series stationary: detrending, seasonality and differencing. Detrending is performed by using regression analysis on a time related trend and identified the residuals. Seasonality makes a component linear or nonlinear which changes and repeats on time related data. Differencing technique, which is generally used for data transforming and stationarizing. We use differencing function to stationarize the electricity consumption data. Let the consecutive consumption values are denoted with t and (t-1) time unit. This function is expressed as

$$x(t) - x(t\text{-}1) = \text{ARMA}\ (p, q) \qquad (1)$$

The difference from equation 1 is called as the Integration part in AR(I)MA. The three parameters are obtained: *p*: AR, *d*: I and *q*: MA.

4. Visualize stationary time series with ACF/PACF: We plot ACF/PACF before estimating ARIMA parameters. Auto Correlation Function (ACF) shows lagged correlation, which is the correlation between two series over time. It helps to visualize the processed series, which returns one lag, compute the correlation, again returns one lag, again compute the correlation and so on. If the dataset is strongly seasonal, peaks coincide with the seasonality period. Plotting ACF may assist to guide the selection of moving average lags. This is a popular approach to visualize the trend of time series data. A regression of time series, partial autocorrelation function (PACF), against its past lags helps to find out a likely order for the AR term. According to a standard linear regression, the term can be treated as the contribution of a change in that particular lag while holding others constant. As stated in the rule of thumb, the ACF confirms trend and infers possible values of the moving average parameters, and the PACF is for the auto regressive part.

5. Estimate parameters for ARIMA model: Parameters are needed to be estimated for developing ARIMA models. The *p*, *d* and *q* values define the order of ARIMA model. ARIMA(*p*, *d*, *q*) model integrates AR(*p*), MA(*q*) models where ACF cuts off after lag '*p*', PACF cuts off after lag '*q*' and '*d*' shows how many times the difference of time series is needed.

The AR model depends on the lagged values of the data. We define that AR(*p*) is an autoregressive model with *p* lags, particular lagged values of y_t are predictor variables.

The AR(*p*) model is defined by the equation:

$$y_t = \delta + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-1} + \epsilon t \qquad \textbf{(2)}$$

Where

- $y_{t-1}, y_{t-2} \ldots y_{t-p}$ are the past series values (lags)
- ϵt is white noise (i.e. randomness)

  and δ is defined by the following equation:

  $$\delta = ((1 - \textstyle\sum_{i=1}^{p} \emptyset i)\ \mu \qquad \textbf{(2.1)}$$

  where μ is the process mean

A moving average model depends on the errors (residuals) of the previous forecasts. It uses past prediction errors in a regression-like model and is common to have negative sign for the parameters MA(*q*) is a moving average model defined by the equation:

$$y_t = c + \epsilon_t + \theta\, 1\, \epsilon_t - 1 + \theta\, 2\, \epsilon_t - 2 + \cdots + \theta\, q\, \epsilon_t - q \qquad \textbf{(3)}$$

Where

- '*q*' is the moving-average trend parameter
- $\epsilon t-1,\ \epsilon t-2 \ldots \epsilon t-q$ are the error at previous time periods.
- ϵt is white noise (i.e. randomness)

An ARMA model describes weakly stochastic stationary time series data for two polynomials. The first and second of these polynomials are for the AR and the MA respectively. This model is stated as the ARMA(p, q) model.

Here,

- *p* denotes the order of the AR polynomial,
- *q* denotes the order of the MA polynomial.

ARMA(p, q) model is defined by the equation:

$$X_t = c + \varepsilon_t + \textstyle\sum_{i=1}^{p} \varphi i\ Xt - i + \sum_{i=1}^{q} \theta i \varepsilon t - i \qquad \textbf{(4)}$$

Where

- φ = the autoregressive model's parameters,
- θ = the moving average model's parameters.

- c = a constant,
- ε = error terms (white noise)

ARMA(p, q) and ARIMA(p, d, q) models have many resemblances such as the AR and MA components are alike, combining a general autoregressive model AR(p) and general moving average model MA(q). AR(p) uses previous values of the dependent variable to predict future information. On the other hand, MA(q) uses the series mean and previous errors to complete predictions.

$$\Delta y_t = a_i \Delta y_{t-i} + b_i \epsilon_{t-i} \quad \textbf{(5)}$$

6. Calculate AIC value: The Akaike Information Criterion is broadly used to measure a statistical model. We compute AIC to estimate the goodness of fit of a model. The model with lower AIC is better than other.

7. Select best ARIMA model: Visualization of ARIMA model is most effective way to compare and determine the best model. In case of multiple models with almost similar or slightly different AIC values ARIMA models plotting reduce the confusion in selecting the best model.

By comparing the AIC values and visualization of ARIMA models based on forecasting performance, the best ARIMA model is obtained.

8. Forecast time series data with the best model: The best ARIMA model with estimated parameters is used to forecast the future behavior of time series data.

The forecasting process with best ARIMA model is presented in Fig. 1.

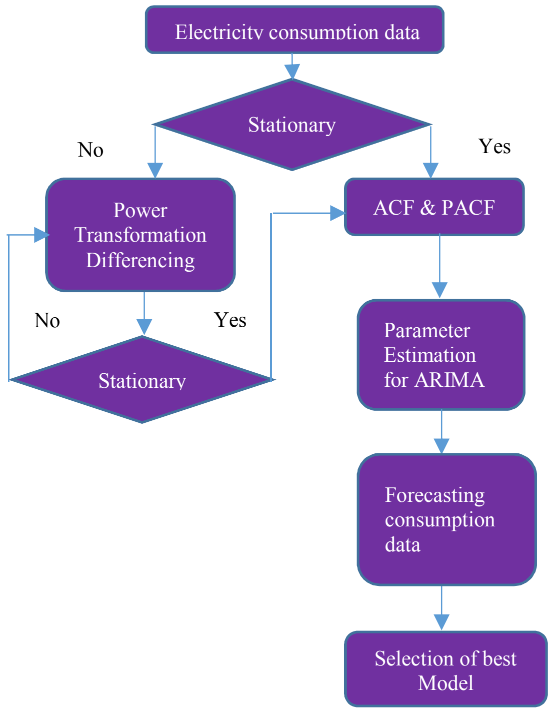

Fig. 1. Model selection process by forecasting.

IV. EXPERIMENTS AND RESULT DISCUSSION

In this segment, experimental results of ARIMA(1,1,2), ARIMA(1,1,7) and ARIMA(1,1,1) are presented for both synthetic and real-world datasets. The datasets demonstrate the performance of these models to forecast electricity consumption. We also show the ACF/PACF plotting, AIC values, coefficient values and ARIMA model plotting for both types of data. Finally, the experimental results are analyzed and discussed.

A. Datasets: In this experiment, we used artificially generated, synthetic, and real-world application, electricity consumption, datasets. We generated 250 random variates as artificial consumption values with Gaussian distribution, which is considered as synthetic dataset. A sample data from the synthetic dataset is presented in Table I.

The real dataset, electricity consumption data, contains the power consumption values of 21330 manufacturing factories at Guangdong province in China [1]. The electricity consumption values were taken every fifteen minutes from smart meters.

We used 96 electricity consumption records as load profile data where each load profile contains 500 consumption values as instances in January 2012. A sample of load profile data is presented in Table II.

TABLE I. A SAMPLE OF SYNTHETIC DATASETS

SL NO.	V1
1.	20.000000
2.	20.594485
3.	19.446299
4.	18.950653
5.	18.577397
6.	18.584372
7.	18.633751
8.	17.915348
9.	16.738053
10.	15.514705

TABLE II. A SAMPLE OF REAL DATASETS

SL No.	V2	V5
1.	2012-01-01	19.09
2.	2012-01-01	21.74
3.	2012-01-01	21.93
4.	2012-01-01	24.86
5.	2012-01-01	22.07
6.	2012-01-01	26.68
7.	2012-01-01	19.96
8.	2012-01-01	21.27
9.	2012-01-01	53.48
10.	2012-01-01	55.68

B. Experiment Setting: Three sets of experiments were conducted on both synthetic and real-world dataset. One was to forecast with ARIMA(1,1,1) model and the other two was ARIMA(1,1,2) and ARIMA(1,1,7) model. R and Rstudio are used [13], [14], to construct the model [15], [17], [19]. We compared the forecasting plots found by these three models. ARIMA modeling needs stationary datasets.

We plotted this synthetic time series and the real dataset in R to see if the dataset was already stationary.

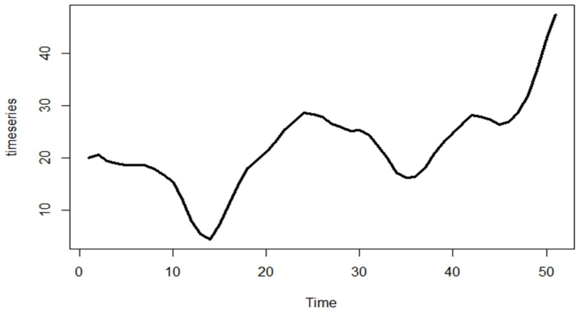

Fig. 2. A time series presentation of synthetic data.

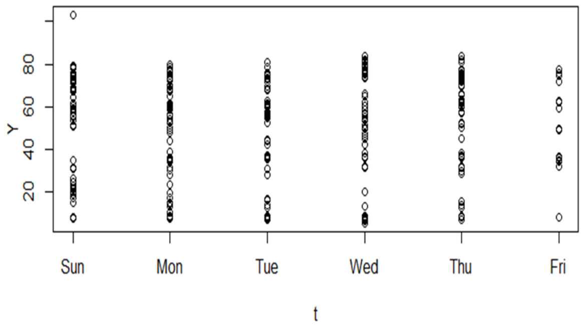

Fig. 3. A time series presentation of electricity consumption data

From the figs. 2 and 3 we can see that synthetic and the electricity consumption data are not stationary enough.
So, we differenced the datasets to make them stationary to apply ARIMA model. Then both of these two datasets became quite stationary. We get the value of d = 1.

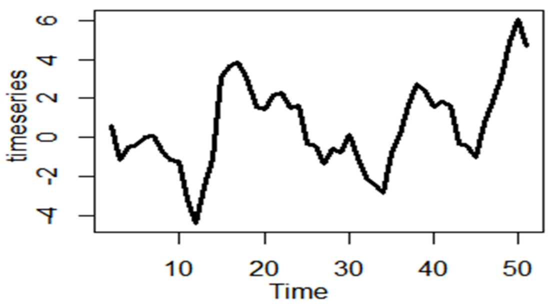

Fig. 4. A time series presentation of differenced synthetic data

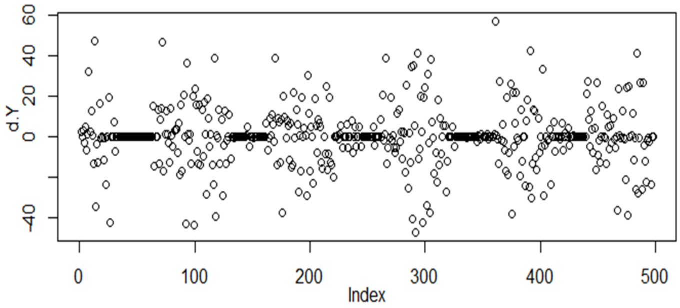

Fig. 5. A time series presentation of differenced electricity consumption data

ARIMA(p, d, q) model integrates AR(p), MA(q) models where ACF and PACF cuts off after lag 'p' and 'q' respectively. ACF and PACF plotting is needed to estimate parameters p and q of the ARIMA model. In figs. 6 and 7, we see that the significance thresholds are represented by horizontal blue dashed lines and the vertical lines, which exceed the horizontal lines, are considered significant.

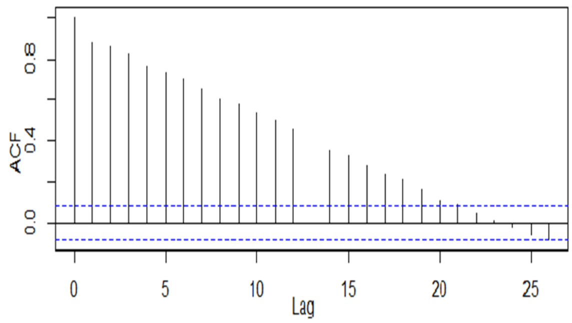

Fig. 6. ACF of synthetic data

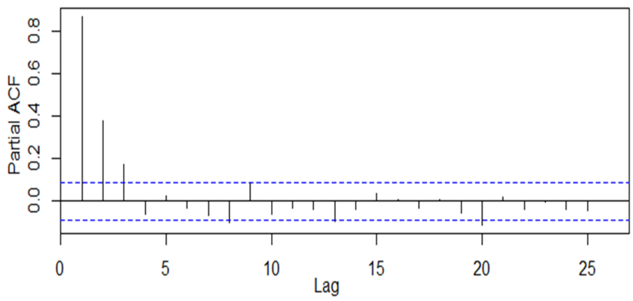

Fig. 7. PACF of synthetic data

C. Experimental Results discussion and Analysis: First, we present the forecasting performance of the three models for both synthetic and real dataset of electricity consumption in the plots. Then we present the comparison of the three ARIMA models for those datasets.
The AIC value is shown as it quantifies the goodness of fit and the simplicity of the model into a single statistic. The model with the lower AIC is generally considered as "better" forecasting model. We obtain near to similar AIC values with ARIMA(1,1,2), ARIMA(1,1,7), and ARIMA(1,1,1) on synthetic dataset. Similarly, we also obtain AIC values, 3944.35, 3946.5, 3945.14 respectively with ARIMA(1,1,7), ARIMA(1,1,1), and ARIMA(1,1,2) on electricity consumption data.
Therefore, we consider the visualization approach for selecting the best model. Visualization is used to forecast the

stationary data, which is easy and acceptable process. We develop ARIMA(1,1,7), ARIMA(1,1,2) and ARIMA(1,1,1) models, which are used to predict future behavior of synthetic and real time electricity consumption data.

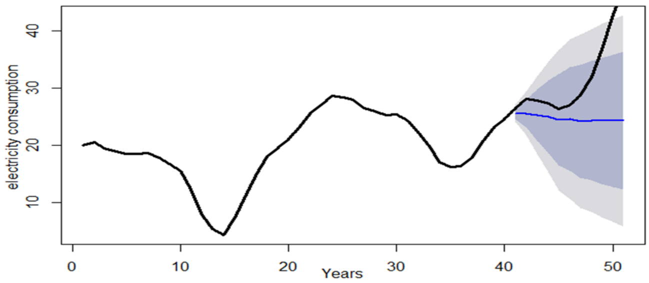

Fig. 8. Forecasting with ARIMA(1,1,7) on synthetic data

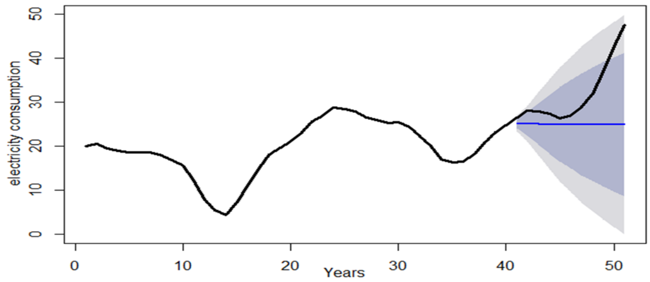

Fig. 9. Forecasting with ARIMA(1,1,2) on synthetic data

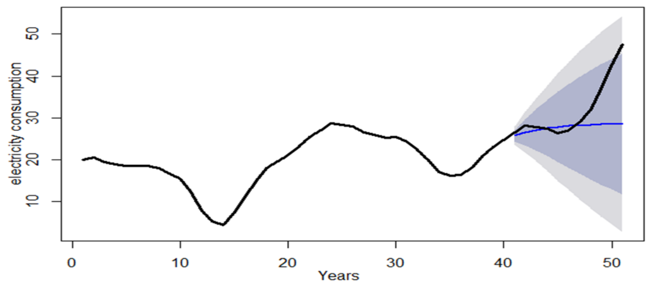

Fig. 10. Forecasting with ARIMA(1,1,1) on synthetic data

From figs. 8, 9, 10 we observe that the ARIMA(1,1,1) model forecasting aligns with the true values (blue line) very well and performs better forecasting than others on synthetic dataset.

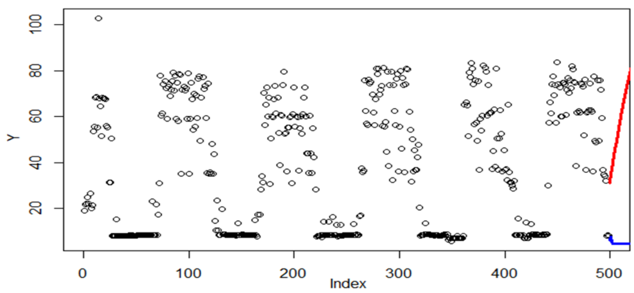

Fig. 11. Forecasting with ARIMA(1,1,7) on electricity consumption data

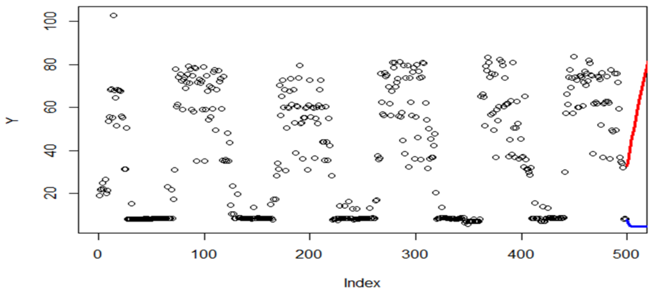

Fig. 12. Forecasting with ARIMA(1,1,2) on electricity consumption data

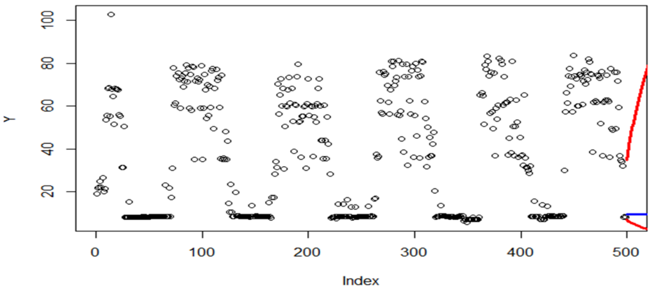

Fig. 13. Forecasting with ARIMA(1,1,1) on electricity consumption data

Similarly, from figs. 11, 12, 13 we see that the ARIMA(1,1,1) model performs more accurate forecasting electricity consumption than other models on real world electricity consumption datasets.

We notice that the forecasting curve generated by ARIMA(1,1,1) model is close to the curve of original data on both synthetic and real datasets. Therefore, the forecasting of electricity consumption provided by ARIMA (1,1,1) model is more accurate than others.

V. RELATED TO INDUSTRY 4.0

Industry 4.0 represents the 4th revolution and significant transformation in manufacturing section. After the 1st industrial revolution, electricity (the second revolution) made significant change in the transformation and adaptation with machine learning approach in autonomous systems. The volume and diversity of electricity consumption data are presented as big data. Industry sectors want to forecast the electricity consumption for their productivity in future and power providers need to predict the consumption demand of their clients. It will help them to adopt their best in unique cases and executing changes for today and preparing for a future and to improve their management and demand. Therefore, the forecasting of electricity consumption time series data can be contributed to the sustainable technologies for industry 4.0.

VI. CONCLUSION AND FUTURE WORKS

This paper focuses on the forecasting time series data with several settings of ARIMA models. Best model with estimated parameters, is selected based on the prediction performance. This prediction result presents the electricity demand of consumers and offers an opportunity of power providers to manage their electricity power in different industrial sectors.

The current work can be extended by developing robust and reliable model for complex time series datasets.

REFERENCES

[1] Bruce L. Bowerman, Richard T. O' Connell, & Anne B. Koehler, "Forecasting time series, and regression: an applied approach," 4th ed. The United States of America: Thomson Brooks, 2005.

[2] Makridakis, S., S.C. Wheelwright and RJ. Hyndman, "Forecasting: Methods and Applications," 3 ed. Wiley, Inc., New York, 1998.

[3] UCI repository of machine learning database [Online].

[4] MD ABDUL MASUD, JOSHUA ZHEXUE HUANG, MING ZHONG, AND XIANGHUA FU, "Cluster Survival Model of Concept Drift in Load Profile Data," IEEE ACESS. Vol 6. 2018.

[5] Ning Xu, Yaoguo Yande Gong. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricityconsumption in China. Energy 2016.

[6] Song Ding, Keith W. Hipel, Yao-guo Dang. Forecasting China's electricity consumption using a new grey prediction model. Energy 2018.

[7] Aowabin Rahman, Vivek Srikumar, Amanda D. Smith. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy 2017.

[8] Yunyou Huang, Jianfeng Zhan, Chunjie Luo, Lei Wang, Nana Wang, Daoyi Zheng, Fanda Fan, Rui Ren. An electricity consumption model for synthesizing scalable electricity load curves. Energy 2018.

[9] Gamze Oĝcu, Omer F. Demirelb, Selim Zaimc. Forecasting Electricity Consumption with Neural Networks and Support Vector Regression. Procedia 2012.

[10] Antonio J. Conejo,Miguel A. Plazas, Rosa Espinola, and Ana B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," IEEE Trans. Power Syst., vol. 20, no. 2, 2005, pp. 1035-1042.

[11] Box, G.E.P. and G. Jenkins, "Time Series Analysis Forecasting and Control," Holden-Day, San Francisco, CA, 1976.

[12] Samer Saab, Elie Badr and Geoge Nasr, "Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon," Energy vol.26,2001, pp. 1-14.

[13] R-project [Online].

[14] R-studio [Online].

[15] Jonathan D. Cryer & Kung-Sik Chan, "Time series analysis: with applications in R," 2nd ed. New York: Springer, 2008.

[16] Volkan S. Ediger, Sertac Aktar, "ARIMA forecasting of primary energy demand by fuel in Turkey," Energy Policy, vol.35, 2007, pp.1701-1708.

[17] Robert H. Shumway & David S. Stoffer, "Time Series Analysis and Its Applications with r Examples," 3rd ed. New York: Springer, 2011.

[18] Qing Zhu, Yujing Guo, Genfu Feng, "Household energy consumption in China forecasting with BVAR model up to 2015," 2012 Fifth International Joint Conference on Computational Sciences and Optimization, 2012.

[19] Oleg Nenadic, Walter Zucchini, "Statistical Analysis with R – a quick start -," Retrieved November 10, 2012.

[20] Javier contreras, Rosario Espinola, Francisco J. Nogales, and Antonio J. Conejo, "ARIMA models to predict next-day electricity prices. Power Systems," IEEE Transactions on 2003, vol.18, no. 3, pp. 10141020.