Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load

Subrina Noureen, Sharif Atique, Vishwajit Roy, Stephen Bayne
Dept. of Electrical and Computer Engineering, Texas Tech University
Lubbock, TX, 79401

***Abstract*—This paper has presented the use of Auto Regressive Integrated Moving Average (ARIMA) method for forecasting of seasonal time series data. The dataset that has been used for modeling and forecasting is a small-scale agricultural load. ARIMA method can be applied only when the time series data is stationary. As seasonal variations make a time series non-stationary, this paper also presents analyses on testing stationarity and transforming non-stationarity into stationarity. Lastly, model has been developed with a specific selection of orders for autoregressive terms, moving average terms, differencing and seasonality and the forecasting performance has been tested and compared with the actual value. The results are encouraging, however there is scope of further research in refining the idea**

***Keywords*—Load Forecasting, ARIMA model, Seasonal ARIMA, Agricultural Load.**

I. Introduction

Global energy demand is speculated to increase by 40% by 2040, and fossil fuels would only provide two-thirds of the overall demand [1]. The increase in carbon emission from the fossil fuel based power generation plants, estimated to be roughly 25%, will worsen the climate even further [2]. The solution to save the world from this disaster is to deploy renewable energy based energy generation plants, which is already gaining more penetration. However, renewable energy brings with it an additional challenge of inherent stochasticity and that's where load forecasting comes into play with even greater significance. An efficient and accurate load forecasting technique can help immensely in energy management and its significance is only going to increase because of increased penetration of inherently volatile distributed renewable energy resources in the power grid.

The amount of energy generated in the power grid should try to match the demand for economic operation. Moreover, large scale electricity storage is not feasible. Therefore, accurate load forecasting is key to ensuring the power grid reliability [3].And its significance is only going to increase because of increased penetration of inherently volatile distributed renewable energy resources in the power grid [4]. Short Term Load Forecasting (STLF) is important in today's deregulated market as an efficient way of load switching and energy buying strategy. Moreover, STLF helps on dynamic control to minimize congestion in the grid. Any good STLF method should satisfy the following two criteria: accuracy and speed. The variables affecting the load demand should also be considered for forecasting [5]. In power grid, the forecast is a representation of the complex relationship between the typical components like weather, time of day, season and the base demand [6].

In spite of drawbacks like complexity of usage, time series methods have gained popularity for STLF recently. Today's data-driven modern systems can provide the huge amount of data needed for time series methods to execute the prediction [7]. According to [8]–[10], time series based stochastic methods have some advantages over some other techniques, e.g. state space model, linear regression, exponential smoothing and artificial neural netowork (ANN). Time series models are easy to understand and also the properties are easier to evaluate. Moreover, historical load demands can be represented as time series data.

In this paper, a particular time series method, ARIMA, has has been proposed by the authors. ARIMA model is based only on the behavior of observed data and completely ignores the independent variable for forecasting [11]. ARIMA is usually applied in stationary processes. However, ARIMA can be used in non-stationary processes too. However, a differencing operation or other necessary transformations need to be carried out to alter the non-stationarity of the process. ARIMA model is characterized by three parameters: p,d and q, where they represent the order of the autoregressive terms, differencing terms and moving average terms, respectively.

The rest of this paper is organized in the following way: the basic ARIMA models for seasonal approach is discussed in section II, ARIMA model is being applied in data set of agricultural load, is analyzed in section III. The dataset is simulated using seasonal ARIMA models and presented in section IV. Finally, section V discusses the relevant future works and conclusion.

II. System Model

Theoretically, the most general models to be used for forecasting stationary time series are ARIMA models. By using operations like differencing, deflating or logging, a stationary time series can be obtained from a non-stationary one [12]. In general terms, If the statistical properties like mean and standard deviation do not have spatial and temporal variation in a time series, then it is deemed to be weakly stationary [13]. A stationary time series is flat with no particular trend, has constant variance and autocorrelation (correlations between a series and another verison of itself separated by time) over time [14]. The power spectrum of a stationary series also doesn't have any temporal variation. A stationary time series, which is one realization of a stochastic process, can then be designed as an additive signal and noise model. The signal can be a sinusoidal oscillation, slow or fast reversion of mean, or rapid sign changes. An ARIMA model acts as a filter that extracts the signal and separates it from the noise. The extracted signal can then be used to measure future forecasted values.

For stationary time series, the ARIMA model is a regression-type linear equation, where the predicted value (dependent variable) consists of lag terms of that projected value and lag terms of the residuals. So the general format of the ARIMA forecasting equation is as follows [12]: Forecasted value of $\mathbf{Y}$ = a weighted sum of one or more autoregressive terms of $\mathbf{Y}$ and/or a weighted sum of forecast errors (along with associated lag terms) and/or a constraint.

The ARIMA model is purely autoregressive if contains only the lagged terms of Y. Autoregression is a specific variation of regression and can be fitted with any standard software of regression. A first order autoregressive model ($AR(1)$)'s independent variable is Y_{t-1}, which is one time period lagged version of Y. When the model includes the lagged terms of the errors, it deviates away from being a linear regression model. Although the predictions are still linear functions of past values, however the prediction doesn't have a linear relationship with the coefficients. That's why the coefficients of an ARIMA model that contains lagged errors need to be optimized by non-linear techniques like hill-climbing.

Auto-Regressive Integrated Moving Average - that's what stands for ARIMA. In the forecasting equation, the lag terms of the dependent variable and forecasting errors are called autoregressive terms and moving average terms, respectively. An integrated version of stationary series comes into play when one or more differencing operations are performed. Exponential smoothing, random walk, autoregressive models can be modeled as ARIMA processes too.

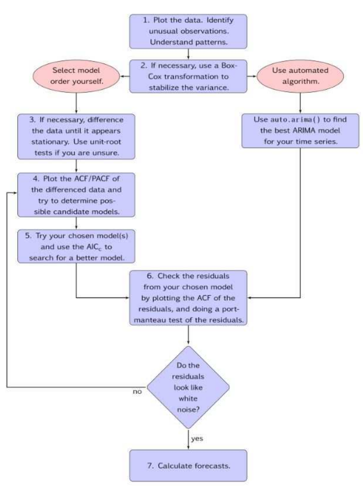

Fig. 1: Forecasting framework for ARIMA model

A. Seasonal ARIMA

Seasonality is the presence of regular pattern in the dataset where the pattern is periodic. The seasonality in a time series is represented by parameter **S**, the span of the seasonality. For example, if yearly sales data of some equipment tend to have high and low values during winter and summer respectively, then $S = 12$. Just like non-seasonal time series, seasonal data can also be modeled and forecasted as ARIMA process.

A seasonal ARIMA is a multiplicative model that incorporates both the variations, seasonal and non-seasonal. The generalized seasonal ARIMA model is expressed in the following way [15]:

$$ARIMA(p, d, q) * (P, D, Q)_S, \tag{1}$$

where

- p = order of AR terms (non-seasonal)
- P = order of AR terms (seasonal)
- q = order of MA terms (non-seasonal)
- Q = order of MA terms (seasonal)
- d = order of differencing (non-seasonal)
- D = order of differencing (seasonal)
- S = span of pattern in seasonality

A seasonal ARIMA model in expressed in a more mathematically detailed way by the following equation 2:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - \Phi_1 B^S - \cdots - \Phi_P B^{PS})(1 - B)^d (1-B^S)^D y_t = (1+\theta_1 B+\cdots+\theta_q B^q)(1+\Theta_1 B^S+\cdots+\Theta_Q B^{QS})\epsilon_t, \tag{2}$$

where B represents the backshift operator which is defined by the following operation:

$$B^m y_t = y_{t-m} \tag{3}$$

1) Model Formulation: The seasonal ARIMA model can be formulated using the following steps:

1) Data plotting and identifying patterns of seasonality and trend.
2) Performing differencing operation if either seasonality or trend is present. For only seasonality, differncing of lag S needs to be done. When only trend is present, a first order differencing needs to be performed. Finally, when both seasonality and trend are present, a first order differencing should be done after seasonal differencing if presence of trend is still there.
3) Analysis of the differenced time series data using ACF and PACF. ACF and PACF give us the order of MA terms and AR terms respectively, for both seasonal and non-seasonal data
4) Estimation of the seasonal model.
5) Analysis of residual to check whether the model is a good fit.

III. Analysis

The ACF and PACF plots of a time series can demonstrate the stationarity of a time series. If the ACF of a time series decreases very fast or the PACF of a time series has a sharp cutoff after the first lag [16], then the time series is deemed to be stationarity. Otherwise, the time series is non-stationary and differencing operation needs to be performed to ensure stationarity. After the necessary ACF and PACF plotting, Augmented Dickey Fuller (ADF) test is performed for hypothesis testing to confirm stationarity.

ADF [17] is a widely used method for checking the stationarity of a time series [18], [19]. ADF is also called unit root test. Presence and absence of a unit root in the characteristic equation indicate non-stationarity and stationarity, respectively. The model for the ADF test is as follows:

$$\partial Y_r = \mu + \beta t + \rho Y_{t-1} + \partial_1 Y_{t-1} + \cdots + \partial_p Y_{t-p} + e_t. \quad (4)$$

Here, μ, p and β represent a constant value, autoregressive order and trend, respectively. Moreover, e_t represents a sequence of zero mean and unit variance independent normal random variables. So, the hypotheses for ADF are presented in the following way [18]:

$$NullHypothesis : H_0 :\mid \rho \mid = 0(Non - stationarity)$$
$$AlternativeHypothesis : H_1 :\mid \rho \mid < 0(Stationarity)$$

The p-value that we obtain after running ADF test decides whether to reject or accept the null hypothesis. For a 95% confidence level, if $p \geq 0.05$, null hypothesis is true. For $p < 0.05$, the value is significant enough to reject the null hypothesis and the time series is stationary.

A. Agricultural Load

The data for agricultural load [20] for the year $2015-2017$ that has been used to test our model is given by Figure 2 below:

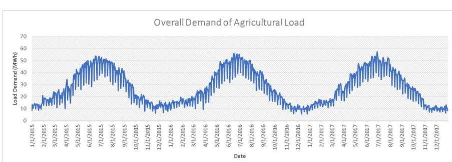

Fig. 2: Agricultural load profile

The overall demand, rolling mean and standard deviation of the industrial data have been plotted in Figure 3. In the case of agricultural data, both the standard deviation and rolling mean are varying that makes the non-stationarity of the agricultural time series data quite obvious. Finally the ACF and PACF plots of the agricultural time series would further validate the non-stationarity.

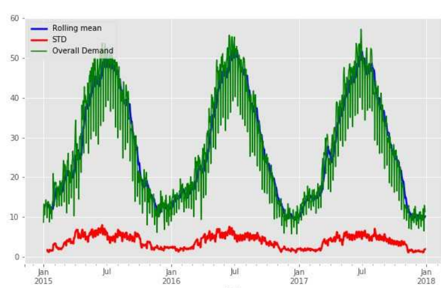

Fig. 3: Rolling mean, overall demand and standard deviation of the agricultural load

The ACF and PACF of the agricultural time series data have been plotted in Figures 4 and 5, respectively. It is easily understood from these figures that the agricultural time series data is not stationary.

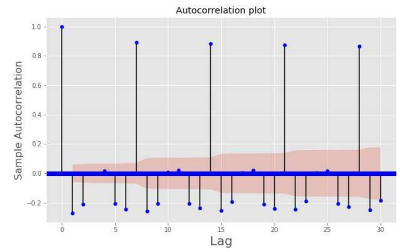

Fig. 4: ACF plot for the agricultural load

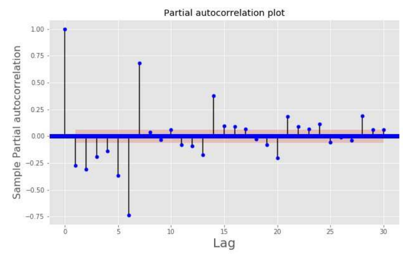

Fig. 5: PACF plot for the agricultural load

So, ADF is performed after applying the first order differencing and a p-value of $4.1485 * 10^{-5}$ indicates the stationarity and thus the null hypothesis is rejected.

IV. Simulation

A. Agricultural Load

The industrial load, along with its trend, seasonality and residuals are plotted below in Figure 6. ARIMA $(0, 1, 0, 1, 1, 1)_{12}$ model has been used for the agricultural time series. The equation of the model is as follows:

$$(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})y_t = (1 + \Theta_1 B^{12})\epsilon_t. \quad (5)$$

The trend and seasonality are clear from the figure. However, the residuals need further introspection,which is done afterwards.

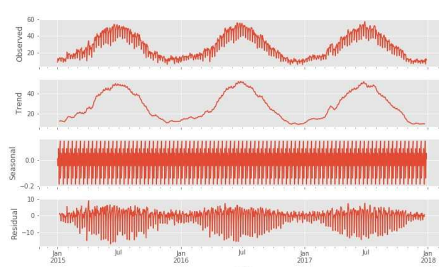

Fig. 6: Trend, seasonality and residual plot for agricultural load

The residuals and its histogram, q-q plot and correlogram are plotted in Figure 7. The residuals doesn't exhibit any seasonality and they are mostly uncorrelated. The density plot drawn from the histogram resemble the standard normal distribution (N(0,1)) to a small extent. The q-q plot also somewhat follows the linear trend of the standard normal distribution. Even though it is not perfect, however our model is a good starting point for the forecasting study.

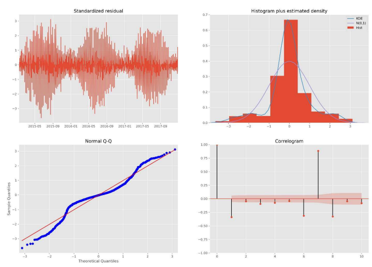

Fig. 7: Agricultural load residual analysis

Finally, the seasonal ARIMA model is utilized to forecast the agricultural loads for the last one year of the three year data. The forecast has been plotted with the actual data in Figure 8. The mean absolute error (MAE) of our forecast is calculated to be 13.23%.

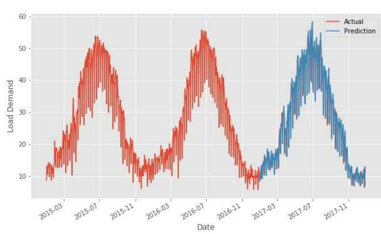

Fig. 8: Forecast for agricultural load

V. Conclusion and Future Works

This paper discusses the use of a widely used time series modeling tool, named ARIMA, to forecast seasonal agricultural loads. Seasonal datasets are non-stationary and ARIMA models can be used only in stationary time series. That's why differencing and other transformations need to be performed before ARIMA can be applied to a time series. This paper includes application of techniques to test stationarity. Finally, forecasting performance has been demonstrated for the agricultural loads using one ARIMA model. The result that has been achieved are competitive, however further research would be conducted to further enhance the performance. Further research could potentially include comparing different models and selecting the optimum ARIMA model by analyzing their respective Akaike and Bayesian information criterios or standardized errors. Moreover, other statistical models like generalized autoregressive conditional heteroskedasticity (GARCH) and autoregressive conditional heteroskedasticity (ARCH) can also be analyzed and compared for enhancing the forecasting performance.

References

[1] V. Dehalwar, A. Kalam, M. L. Kolhe, and A. Zayegh, "Electricity load forecasting for urban area using weather forecast information," in *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, Oct 2016, pp. 355–359.

[2] I. Agency, "World energy outlook 2014 executive summary."

[3] S. Karthika, V. Margaret, and K. Balaraman, "Hybrid short term load forecasting using arima-svm," in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, April 2017, pp. 1–7.

[4] S. Atique, S. Noureen, V. Roy, V. H. Subburaj, S. Bayne, and J. Macfie, "Forecasting of total daily solar energy generation using ARIMA: a case study," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) (IEEE CCWC 2019)*, , USA, Jan. 2019.

[5] E. E. El-Attar, J. Y. Goulermas, and Q. H. Wu, "Forecasting electric daily peak load based on local prediction," in *2009 IEEE Power Energy Society General Meeting*, July 2009, pp. 1–6.

[6] D. Fay and J. V. Ringwood, "On the influence of weather forecast errors in short-term load forecasting models," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1751–1758, Aug 2010.

[7] A. K. Srivastava, A. S. Pandey, and D. Singh, "Short-term load forecasting methods: A review," in *2016 International Conference on Emerging Trends in Electrical Electronics Sustainable Energy Systems (ICETEESES)*, March 2016, pp. 130–138.

[8] C. Hor, S. J. Watson, and S. Majithia, "Daily load forecasting and maximum demand estimation using arima and garch," in *2006 International Conference on Probabilistic Methods Applied to Power Systems*, June 2006, pp. 1–6.

[9] I. Moghram and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1484–1491, Nov 1989.

[10] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, Feb 2001.

[11] A. Qonita, A. G. Pertiwi, and T. Widiyaningtyas, "Prediction of rupiah against us dollar by using arima," in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Sept 2017, pp. 1–5.

[12] "Introduction to arima: nonseasonal models," https://people.duke.edu/ rnau/411arim.htm.

[13] M. Poulos and S. Papavlasopoulos, "Automatic stationary detection of time series using auto-correlation coefficients and lvq neural network," in *IISA 2013*, July 2013, pp. 1–4.

[14] "Stationarity," https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm.

[15] R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, ser. Springer Texts in Statistics. Springer New York, 2010. [Online]. Available: https://books.google.com/books?id=dbS5IQ8P5gYC

[16] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Australia: OTexts, 2018.

[17] D. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the Americal Statistical Association*, vol. 74, no. 366, pp. 427–431, 1979.

[18] S. Halim, I. N. Bisono, Melissa, and C. Thia, "Automatic seasonal auto regressive moving average models and unit root test detection," in *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, Dec 2007, pp. 1129–1133.

[19] J. Wu and C. K. Chan, "The prediction of monthly average solar radiation with tdnn and arima," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2, Dec 2012, pp. 469–474.

[20] "Load profile," https://www.pge.com/nots/rates/006f1c4_class_load_prof.shtml, accessed: 2018-10-16.