International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

TELECOMMUNICATION DATA FORECASTING BASED ON ARIMA MODEL

Anjuman Akbar Mulani [1], Prof. Sachin Muraraka[2], Prof. K. Sujatha[3]

1Student ME E&TC,Shree Ramchandra College of Engineering, Lonikand,Pune,Maharashtra

2,3Assistant Professor,Shree Ramchandra College of Engineering, Lonikand,Pune,Maharashtra

ABSTRACT

Forecasting of telecommunication data find difficult according to International Telecommunication Union (ITU) recommendation due to uncertainty involved and the continuous growth of data in telecommunication markets as it helps them in planning and determining their networks. So, there is a need of good forecasting model to predict the future. In this paper, ARIMA is utilized for forecasting telecommunication data. This model adaptively uses auto regression, moving average or combined together if required. The three steps involved in the ARIMA model is identification, estimation and forecasting. The adaptive ARIMA model is then applied to M3-Competition Data to do forecasting of telecommunication data. The performance of the model is found out using the evaluation metrics such as Root Mean Square Error, Sum otwf Squared Regression, Maximum Absolute Error, Mean Absolute Deviation, and Mean Absolute Percentage Error. The results proved that the ARIMA models provide 7.6% improvement than the neural network method in forecasting performance.

Keywords: ARIMA Model, Forecasting, ITU Recommendations, Telecommunication, Time Series M3 Competition Data.

I. INTRODUCTION

Forecasting predicts what will occur in the future. Forecasting is becoming important because of the high turbulence in telecommunications market because of rapid technological development and liberalization. Telecommunications industries are growing very fast and the ability to determine the future trends is an important task. Business peoples try to forecast with high accuracy, but it is difficult in today's fast-paced business world [12]. The boundaries of the telecommunication industries are increasing very fast and there is tremendous competition amongst them. Telecommunications companies are in need to implement new technologies to meet the requirement of providing best services. Many industries depend on data monitoring to enhance their business by analyzing the market trends and marketing values for their products [8, 4]. However, there is a problem in bring the gap between the known data and probable value in the future due to the lack of reliable input data for forecasting and adequate model.

The forecastability of a time series data is an important factor for the forecasting practitioners. The forecasters can determine the outcome of any event with the prior knowledge about the event. While forecasting telecommunication data, many errors may occur and it motivates to find a better forecasting approach which

International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

minimizes the forecasting errors. If there is a problem in accuracy of forecasting data, then reducing the forecasting error will be ineffective. So, it is necessary to implement new techniques to minimize the effects of poor forecasts [2]. The forecastability of a time series is dependent on the data regularity. If the time series is less regular, then achieving good level of forecasting accuracy is very difficult [3, 9-11].

In this paper, ARIMA model for forecasting telecommunication data is proposed. M3-Competitive data is used here and the performance is measured in terms of Sum of Maximum Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) Squared Regression (SSR), Root Mean Square Error (RMSE) and Mean Absolute Deviation (MAD). The paper is structured as follows. Section 2 provides the literature review. Section 3 describes the M3-Competition telecommunication data. Section 4 describes about the ARIMA based forecast model. Section 5 evaluates performance of forecasting. Results of forecasting are given in section 6 and section 7 concludes the paper.

II. LITERATURE REVIEW

A forecasting system collects and processes data for thousands of items and develops forecasts using forecasting methods such as ARIMA model. It has an interactive management user interface which maintains a database of demands and has report file-writing capabilities. Three planning horizons for forecasting exist. The short-term forecast covers a period below three months. The medium-term forecast usually covers a period of three months to two years. And, the long-term forecast covers a period of more than two years. Generally, the short-term forecast is used for the routine operations and planning of a company. The long-term forecast is used more for strategic planning [22].

Forecasting techniques can be categorized into quantitative techniques and qualitative techniques. The qualitative models include subjective or intuitive models and the several quantitative techniques are available for forecasting are Random walk (simple forecasting method), exponential smoothing methods such as simple exponential smoothing (SES), robust trend, Holt and Holt-winters etc. In the random walk model, variable values follow a random step for each time interval. It assumes the importance of current observation only and the previous observations provide no information [15].

SES is a suitable method for forecasting seasonal data. But if there is a trend it's not suitable [16]. An extended form of simple exponential smoothing method is HOLT method and it allows forecasting data with trends. There can be improved to deal with both trend & seasonal variations. By considering the seasonal variations Holt-winter method overcomes Holt method [18]. Method suitable for forecasting univariate time series data in the presence of outliers is robust trend. When using this method, incorrect labeling of samples as outliers may occur.

Mahsin used Box-Jenkins methodology for creating seasonal ARIMA model for monthly basis rainfall information taken for Dhaka station, Bangladesh, for amount during 1981-2010. In their paper, ARIMA (0, 0, 1) $(0, 1, 1)^{12}$ model was found adequate and also the model is used for forecasting the monthly rain [20]. Seyed *et al*. used time series methodology to model weather parameter in Islamic, Abadeh Station in Iran and counseled ARIMA$(0,0,1)(1,1,1)^{12}$ because the best appropriate monthly rainfall information and ARIMA$(2,1,0)(2,1,0)^{12}$ for average temperature on monthly basis for Abadeh station. They modelled weather parameter using random methods (ARIMA Model) [21].

International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

III. M3-COMPETITION DATA

As the forecastability is growing up, there is a great importance for checking the adequacy of forecasting methods. Makridakis and Hibon developed the M-Competition which is one of the important researches in this field [5, 13]. M Competitions or M-Competitions is used for a series of competitions organized by teams led by forecasting researcher Spyros Makridakis and intended to find out and compare the accuracy of various forecasting methods. Data used here are obtained from the Institute of Forecasters. The last edition (M3) is referred to year 2000 and it includes 3003 time series classified in to various types such as micro (828), industry (519), macro (731), finance (308), demographic (413) and other (204) and the various time durations between the successive observations are gives by yearly, quarterly, monthly and of unknown periodicity('others').

Minimum number of observations for each type of data is required to ensure that enough data can develop an absolute forecasting model. This minimum was set as 14 observations for yearly series, 16 for quarterly, 48 for monthly and 60 for 'other' series. Table 1 show the classification of the 3003 series according to the two major grouping discussed above. The table shows the number of time series based on both time interval and domain.

Table i: The classification of the 3003 time series used in the m3-competition

Time interval between successive observations	Types of time series data						
	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57		756
Monthly	474	334	312	145	111	52	1428
Other	4			29		141	174
Total	828	519	731	308	413	204	3003

IV. ARIMA-BASED FORECAST MODEL

Box and Jenkins introduced the Autoregressive integrated moving average (ARIMA) model. It is used as one of the popular method for forecasting time series data [6, 14]. Auto regression and Moving average can be used separately and combined together if required. This method is considered as sophisticated method and required usually large dataset of past data. Major steps involved in this process are identification, estimation and forecasting.

ARIMA model takes into account the past data and decomposes it into (i) an Autoregressive process (AR), where there is a memory of past events; (ii) an integrated (I) process, which makes the data stationary and ergodic making it easier to forecast; (iii) a moving average (MA) of the forecast errors, such that the forecast will be accurate for longer historical data. ARIMA models therefore have three model parameters, one for the AR(p) process, one for the I(d) process and one for the MA(q) process, all combined into the ARIMA (p,d,q) model.

International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

ARIMA model is most commonly used in time-series analysis and multivariate regressions. Because, the error residuals are correlated with their own lagged values and this serial correlation violates the standard assumption that disturbances are not correlated with other disturbances. The ARIMA model is used for time series forecasting where the variables future value is a linear function of past values or observations and random errors and is expressed as,

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (1)$$

Where, actual value y_t and random error ε_t at time t, and ϕ_i (i = 1, 2, . . . , p) and θ_j (j= 0, 1, 2, . . . , q) are the model parameters. Integers, p and q are the order of the model. The random errors, ε_t are assumed to be independent and equally distributed with a zero mean a constant variance of σ^2 [7].

The steps in ARIMA model can be summarized as follows (shown in Fig 1). Model Identification: ARIMA model assumes that the time series is stationary. So, data transformation is done to produce a stationary time series. For the stationary time series over the time the mean and autocorrelation structure are constant. So, differentiation and power transformation are required to change the time series to stationary. To identify the appropriate model form, partial autocorrelation and autocorrelation are calculated from the data and it is compared to the theoretical autocorrelation and partial autocorrelation for different ARIMA models. Steps (ii) and (iii) will determine whether the model is appropriate [1].

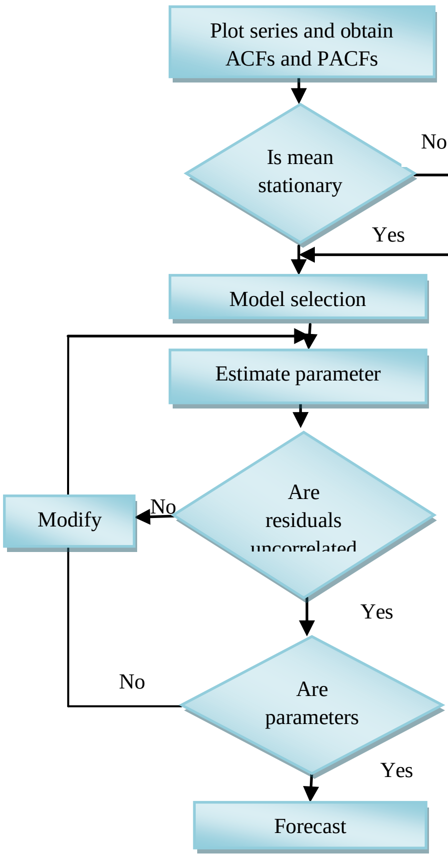


Fig1. Steps of the ARIMA model

International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

(ii) Parameter Estimation: The nonlinear least square procedure is used for parameter estimation in ARIMA Model.

(iii) Diagnostic Checking: Several diagnostic statistics and plots such as Histogram, normal time sequence plot & probability plot are used to check the fitness of the model estimated in step (i). Model adequacy can be tested by Chi-square test. Once a satisfactory model is obtained, the selected model will be used for forecasting.

V. EVALUATION METRICS

Forecast accuracy is measured using Sum of Squared Regression (SSR), Mean Absolute Deviation (MAD), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Maximum Absolute Error (MAE).

(i) SSR: Sum of Squared Regression

$$SSR = \sum_{l=1}^{L} y - \overline{y} \qquad (2)$$

where, y = original data; $\overline{y}$ = predicted data; L = number of samples.

(ii) MSE: (MSE) averages the squared prediction error at the same prediction horizon. A derivation of MSE is Root mean Square Error (RMSE).

$$MSE(i) = \frac{1}{L}\sum_{l=1}^{L} \Delta^l(i)^2 \qquad (3)$$

(iii) MAD: Mean Absolute Deviation (MAD) is the resistant estimator of the dispersion of the original error from the prediction error.

$$AD(i) = \frac{1}{n}\sum_{l=1}^{L} \left|\Delta^l(i)\right| - m \qquad (4)$$

where, $M = \underset{l}{median}\left(\Delta^l(i)\right)$ and median is the $\frac{n+1}{2}th$ order statistic.

(iv) MAPE: MAPE averages the APE in the predictions of multiple UUTd at the same prediction horizon.Rather than the mean, Median absolute percentage error (MdAPE) can be computed by median in a similar fashion.

$$MAPE(i) = \frac{1}{L}\sum_{l=1}^{L} \left|\frac{100\Delta^l(i)}{r_\bullet^l(i)}\right| \qquad (5)$$

v) MAE :The absolute prediction error is averaged by MAE for multiple UUTs at the same prediction horizon. If we use median rather than mean gives absolute error (MdAE).

$$MAE(i) = \frac{1}{L}\sum_{l=1}^{L} \left|\Delta^l(i)\right|$$

International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

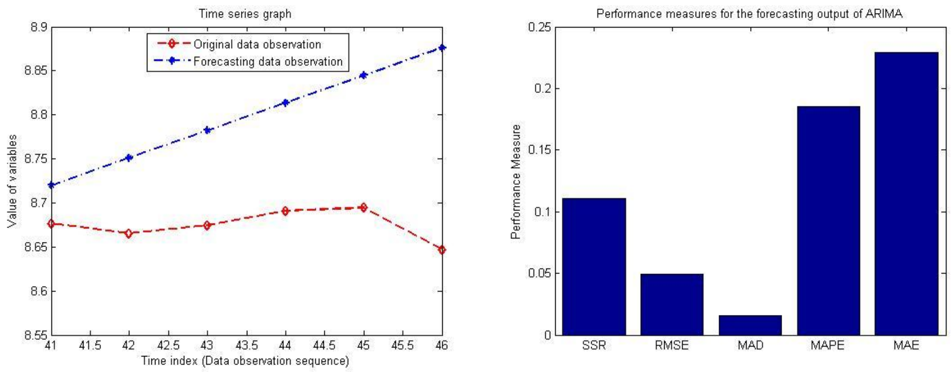
VI. RESULTS AND DISCUSSION

This section presents the performance evaluation of the ARIMA model in M3C data. Matlab 8.2.0.701 (R2013b) with a system configuration of 2GB RAM Intel processor and 32 bit OS is used for ARIMA model implementation. In order to calculate the forecastability of the ARIMA model, it is applied to the yearly, quarterly, monthly and 'other' data.

The autocorrelation plots are used in the model identification stage of the ARIMA time series models. The randomness in a data set is checked by autocorrelation plots. To find the randomness, autocorrelations of data values at variable time lags are calculated. If we get autocorrelation is near to zero for all the time-lag separations, then the data is random. The data is not random if the autocorrelations are not zero. Partial autocorrelation are used to find the order of the autoregressive model.

The time-series graph illustrates the data points at successive time intervals. The time is measured on the horizontal axis and the variable is measured on the vertical axis. For comparing between the original data observed and the observed forecasting data and respective performance measure for the yearly data, quarterly data, monthly data and 'other' data are shown in Fig 2, 3, 4 and 5 respectively the time series graph is used.

Fig 2. Yearly data



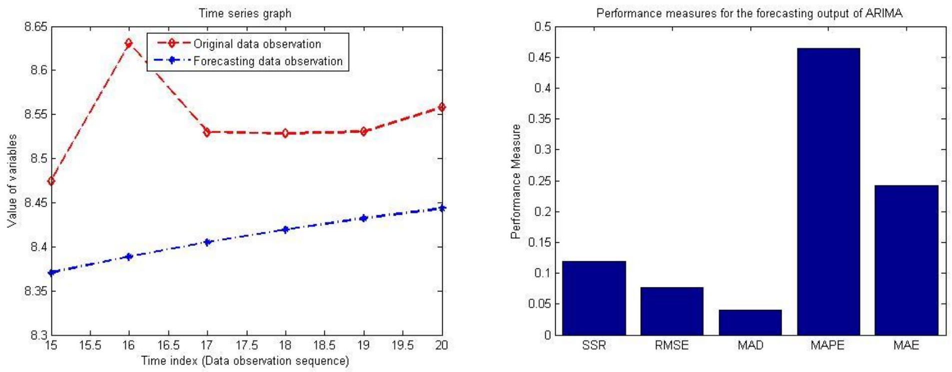


Fig 3. Quarterly data

International Journal of Electrical and Electronics Engineers
Vol. No. 8, Issue No. 02, July-December 2016
ISSN (O) 2321-2055
ISSN (P) 2321-2045

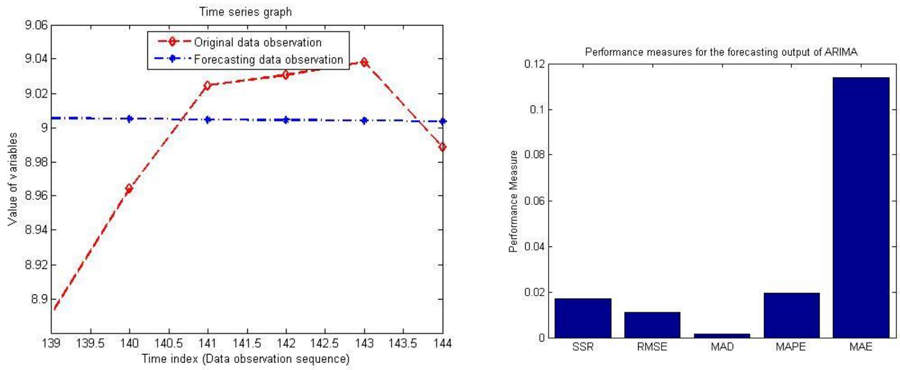


Fig 4. Monthly data

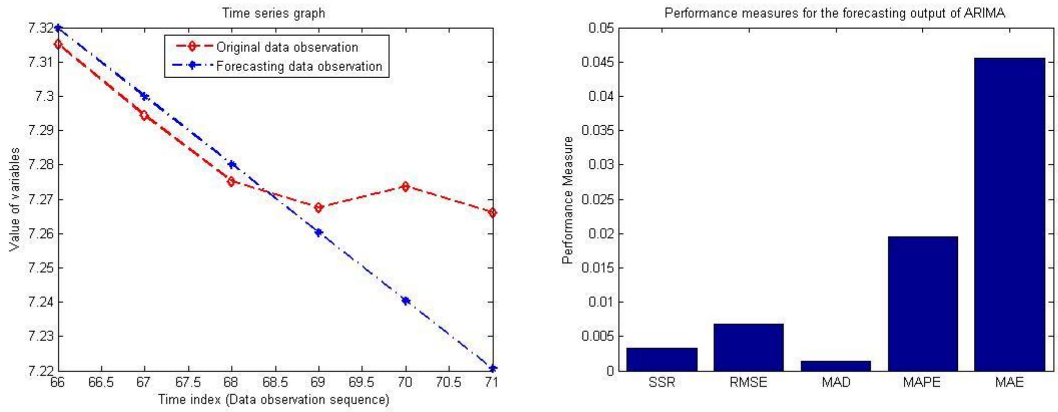


Fig 5. 'Other' data

VII. CONCLUSION

In this study, we have applied the ARIMA model for the M3-competition data set of 3003 time series to determine its forecastability. Better forecasting even when little detail about the data is available is provided by ARIMA model. The forecasting accuracy of the ARIMA model is determined by parameters like SSR, RMSE, MAD, MAPE, and MAE for the yearly, quarterly monthly and 'other' data and it is compared with the results given in [19]. It showed that the performance of ARIMA model is better than the other models. We therefore conclude that the ARIMA model gave 7.6% improvement than the neural network method in telecommunication data.

REFERENCES

[1] J. E. Hanke and D. W. Wichern, Business Forecasting, Prentice Hall, Englewood Cliffs, NJ, USA, 2009.

[2] Kolassa S., How to assess forecastability Foresight, the International Journal of Applied Forecasting, 2009, pp. 41–45.

[3] Tashman L.J., Special feature on forecastability: Preview, Foresight, The International Journal of Applied Forecasting, vol. 23, 2009.

[4] Gary Madden, Joachim Tan, Forecasting telecommunications data with linear models, Telecommunications Policy, vol. 31, pp. 31–44, 2007.

[5] Makridakis S. and Hibon M., The M-3 Competition-Results, implications and conclusions, IJOF 16, pages: 451-476, year 2002.

[6] P. F. Pai and C. S. Lin, A hybrid ARIMA & SVM model for stock prizing forecasting, Omega, volume 33, no-6, pages. 498–506, year 2005.

[7] G. P. Zhang, Time series forecasting using a hybrid ARIMA model & neural network model, Neurocomputing, vol. 50, pp. 159–175, 2003.

[8] Boylan J.E., Forecasting ability definitions in more precise way, Foresight, IJAF, pages. 34 to 40, year 2009.

[9] Collopy F. & Armstrong J, Empirical Comparison of Error Measures for generalizing forecast methods, IJOF, vol. 8, pp. 69–80, 1992.

[10] [10] Fildes R., The Extrapolative Forecasting Methods Evaluation, IJF, volume-8, pages. 81 to 98, year 1992.

[11] Fildes R., Makridakis S., Meade N. & Hibon M., Generalising about Univariate Forecasting Methods: Further Evidence for Empirical, IJOF, volume- 14, pp. 339–358, 1998.

[12] Grubesic T. & Murray A., Geographies for imperfection in telecom analysis, Policies of Telecom. , volume. - 29, pages: 69–94, year 2005.

[13] Makridakis S., Chatfield C., Mills T., Hibon M., Lawrence M., Ord K., The Makridakis & Hibon two competition: a real-time judgmentally based forecasting study, IJOF, vol.-9, pp: 5 to 23, year 1993.

[14] E. Parzen, ARIMA Models about Time Series Analysis and Forecasting, Journal of Forecasting, vol. 1, pages- 67–82, 1982.

[15] Simmons L. F., M-Competition — A closer

[16] look at Naıve2 & median APE: a note, IJOF, volume -4, pp. 457-460, 1986.

[17] Gardner E., Exponential Smoothening: The state of the artpart ii, IJOF, volume-22, pages: 637–666, 1992.

[18] Kotsialos A., Papageorgiou M. & Poulimenos A., Long-term sales forecasting using Neural Network Methods & Holt-Winters, JOF, volume-24, pages: 353–368, year 2005.

[19] Fildes R. & Petropoulos F., An Evaluation of SFM Selection Rules (Working Paper on LUMS 2013:2), The Department of Management Science, University Lancaster.

[20] Hibon M. & Makridakis S., Makridakis & Hibon-three competition results, conclusion and implications, IJOF, volume-16, pp. 451-476, 2000.

[21] Mahsin M, Rainfall modeling in division Dhaka of Bangladesh using time series analysis, JFMMAA, volume - 5, pages: 67-73, 2011.

[22] Vahdat S.F., Liaghat A., Shahida N., Sarraf A., & Shamsnia S.A., Weather Parameters Modeling Using Stochastic Methods (ARIMA Model) (Study case: Iran, Abadeh Region), International Conference on Environment and Industrial Innovation, volume-12, pages: 282-285, year 2011.

[23] Stephen A. Delurgio & Carl D. Bhame, Forecasting systems for Operations Management, 1991.