MANTECH Publications
International Journal of Radio, Space and Aerospace Engineering
Volume 1 Issue 1

ARIMA Model for Forecasting Telecommunications Traffic

Nilesh Subhash Nalawade[1], Anjuman Akbar Mulani[2]
[1,2]Assistant Professor
Department of E& TC,
[1]STES's Sinhgad Institute of Technology, Lonavala, Pune, Maharashtra 410401, India
[2]Shriramchandra College of Engineering, Lonikand, Pune, Maharashtra 410305, India
E-mail: nnalawade.sit@sinhgad.edu[1]

Abstract

According to International Telecommunication Union (ITU), forecasting of telecommunication data is found to be difficult 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 a good forecasting model to predict the future. In this paper, Autoregressive Integrated Moving Average model is utilized for forecasting telecommunication data. This model adaptively uses auto regression, moving average or combined together if required. The major 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 Sum of Squared Regression, Root Mean Square Error, Mean Absolute Deviation, Mean Absolute Percentage Error and Maximum Absolute Error. The results proved that the ARIMA models provide 7.6% improvement than the neural network method in forecasting performance.

Keywords: *Telecommunication Forecasting, ITU Recommendations, ARIMA model.*

INTRODUCTION

Nowadays, Forecasting predicts what will occur in the future. Forecasting is becoming important because of the high turbulence in telecommunications market, which is the result of rapid technological development and liberalisation. 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 the competition among the industries is also growing as well. Telecommunications companies are in need to develop new strategies to face challenges in providing the best communication services. Many industries depend on data monitoring to improve their business by analyzing the future trends and marketing value for their products [1, 4]. However, there is a problem in bridging 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 forecasting ability 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 minimizes the forecasting errors. If a time series data is not forecasted accurately, then reducing the forecasting error will be ineffective. So, it is necessary to find new ways to minimize the effects of poor forecasts [2]. The forecastability of a time series depends on the regularity of the data. 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 Squared Regression (SSR), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), Maximum Absolute Error (MAE). 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 forecasting performance. Forecast results are given in section 6 and section 7 concludes the paper.

LITERATURE REVIEW

A forecasting system collects and processes data for thousands of items and develops forecasts using forecasting methods. 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 of less than 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 daily operation and plans of a company. The long-term forecast is used more for strategic planning [22].

Forecasting techniques can be categorized into qualitative and quantitative techniques. The qualitative models include subjective or intuitive models and the several quantitative techniques that are available for forecasting are Random walk (simple forecasting method), exponential smoothing methods such as simple exponential smoothing (SES), Holt, Holt-winters, Robust trend etc. In the random walk model, the value of the variable follows a random step for each time interval. It assumes that the current observation is only important and the previous observations provide no information [15]. Simple exponential smoothing is a suitable method for forecasting seasonal data. But it is not a suitable method when there is a trend [16]. Holt method is an extended form of simple exponential smoothing method and it allows forecasting data with trends. It can be improved to deal with both trend and seasonal variations. Holt-winter method overcomes Holt method by considering the seasonal variations [18]. Robust trend methods are suitable for forecasting univariate time series data in the presence of outliers. When using this method, incorrect labeling of samples as outliers may occur.

Mahsin used Box-Jenkins methodology to create seasonal ARIMA model for monthly rainfall information taken for Dhaka station, Bangladesh, for the amount from 1981-2010. In their paper, ARIMA (0, 0, 1) (0, 1, 1)12 model was found adequate and also the model is employed for forecasting the monthly rain [20]. Seyed et al. used time series methodology to model weather parameter in Islamic Republic of Iran at Abadeh Station 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 monthly average temperature for Abadeh station. They modelled weather parameter using random methods (ARIMA Model) [21].

M3-COMPETITION DATA

As the forecasting availability is growing up, there is a need for analyzing the adequacy of forecasting methods. Makridakis and Hibon developed the M-Competition which is considered to be one of the important researches in this field [5, 13]. M Competitions or M-Competitions is a term used for a series of competitions organized by teams led by forecasting researcher Spyros Makridakis and intended to evaluate and compare the accuracy of different 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 different time intervals between the successive observations are gives by yearly, quarterly, monthly and of unknown periodicity('others').

Minimum number of observations is required for each type of data to ensure that enough data can develop an appropriate 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 - Classification of 3003 Time Series Used In 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

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.
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 future value of a variable is a linear function of past observations and random errors and is expressed as,

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

Where, is the actual value and ε_t is the random error at time t, and $(i = 1, 2, \ldots, p)$ and $(j = 0, 1, 2, \ldots, q)$ are model parameters. Integers, p and q are the order of the model. The random errors, ε are assumed to be independent and identically distributed with a zero mean a constant variance of ^{2} [7].

The steps in ARIMA model can be summarized as follows (shown in Fig 1).

(i) Model Identification:

ARIMA model has the assumption that the time series is stationary. So, data transformation is done to produce a stationary time series. For stationary time series, the mean and autocorrelation structure are constant over time. So, differentiation and power transformation are needed to change the time series to be stationary. To identify the appropriate model form, autocorrelation and partial autocorrelation are calculated from the data and it is compared to the theoretical autocorrelation and partial autocorrelation for the various ARIMA models. Steps (ii) and (iii) will determine whether the model is appropriate [8].

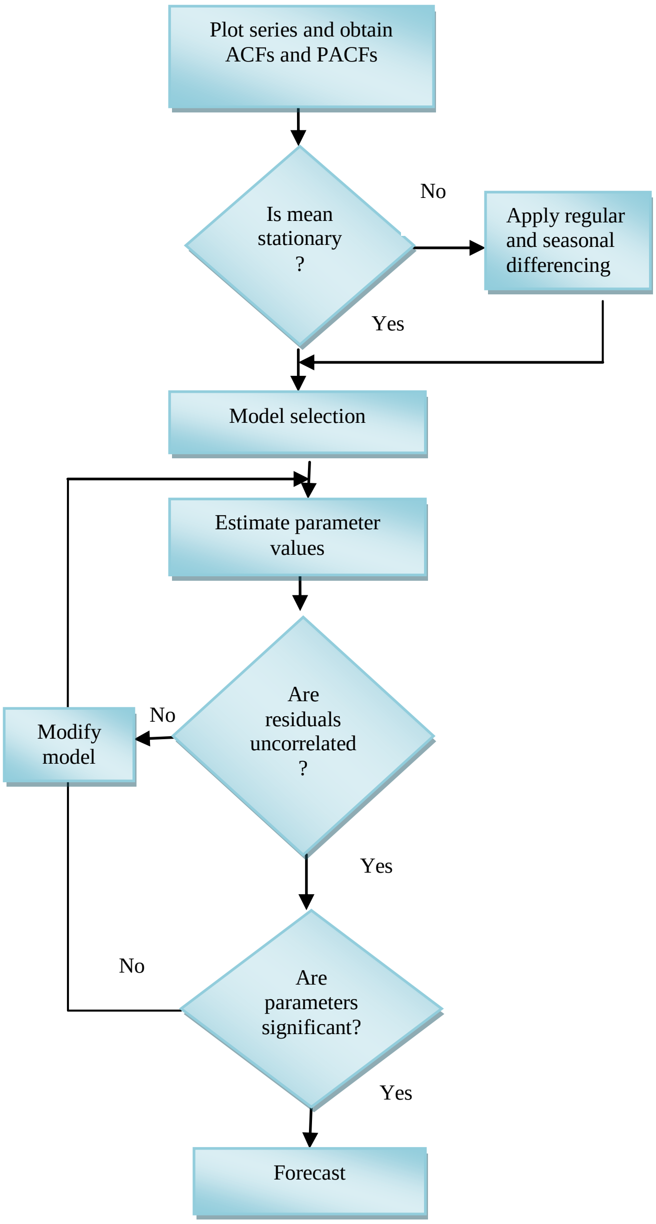

Fig 1: Steps of the ARIMA model

(ii) Parameter Estimation:

Parameters in ARIMA model is estimated using the nonlinear least square procedure.

(iii) Diagnostic Checking:

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

EVALUATION METRICS

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

(i) SSR: Sum of Squared Regression

$$SSR = \sum_{l=1}^{L} y - \bar{y} \tag{2}$$

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

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

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

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

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

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

(iv) MAPE: Mean Absolute Percentage Error (MAPE) averages the absolute percentage errors in the predictions of multiple UUTd at the same prediction horizon. Instead of the mean, median can be used to campute Median absolute percentage error (MdAPE) in a similar fashion.

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

v) MAE : Mean Absolute Error (MAE) averages the absolute prediction error for multiple UUTs at the same prediction horizon. Using median instead of mean gives absolute error (MdAE).

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

RESULTS AND DISCUSSION

This section presents the performance evaluation of the ARIMA model in M3C data. The ARIMA is implemented using Matlab 8.2.0.701 (R2013b) with a system configuration of 2GB RAM Intel processor and 32 bit OS. To determine the forecasting ability 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. Autocorrelation plots are used to check the randomness in a data set. To find the randomness, the autocorrelations for data values at varying time lags are computed. If the autocorrelation is near to zero for all the time-lag separations, then the data is random. If the autocorrelations are not zero, then the data is not random. 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. The time series graph for the comparison 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.

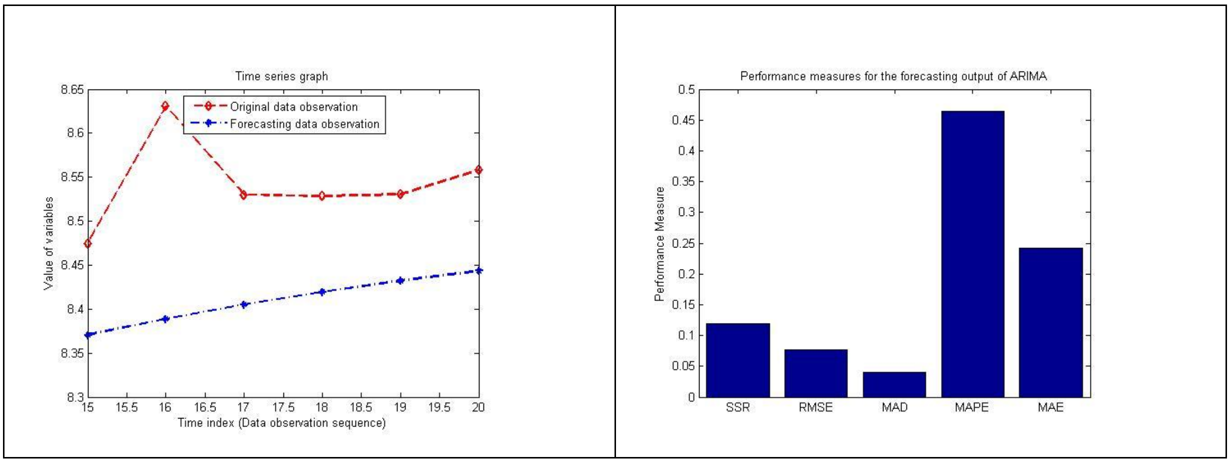

Fig 2: Yearly data

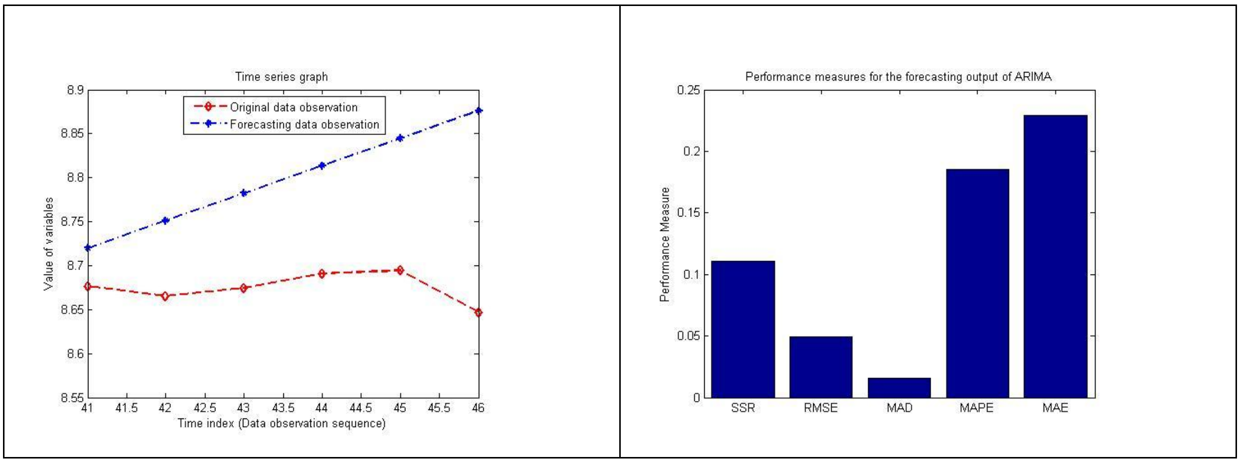

Fig 3: Quarterly data

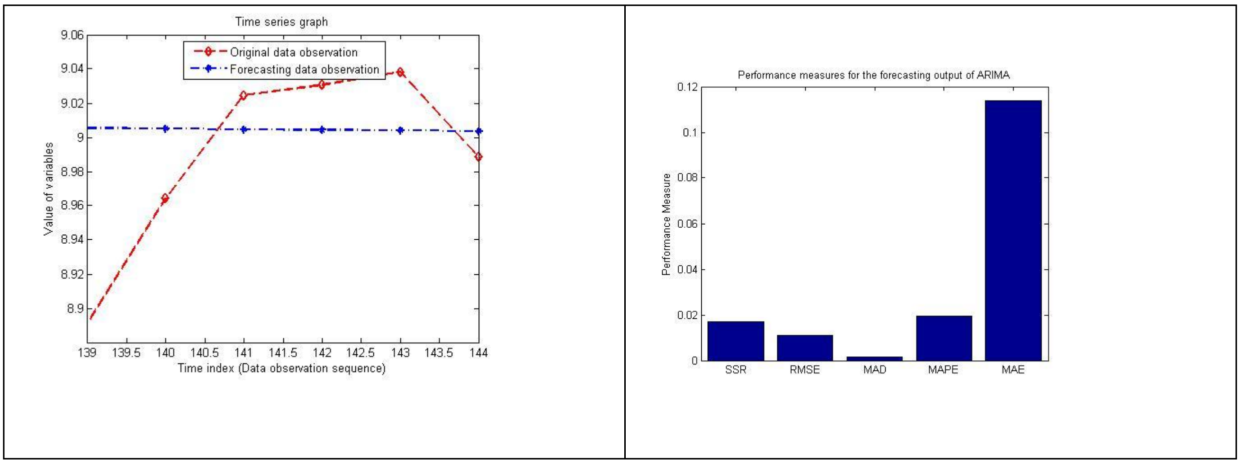

Fig 4: Monthly data

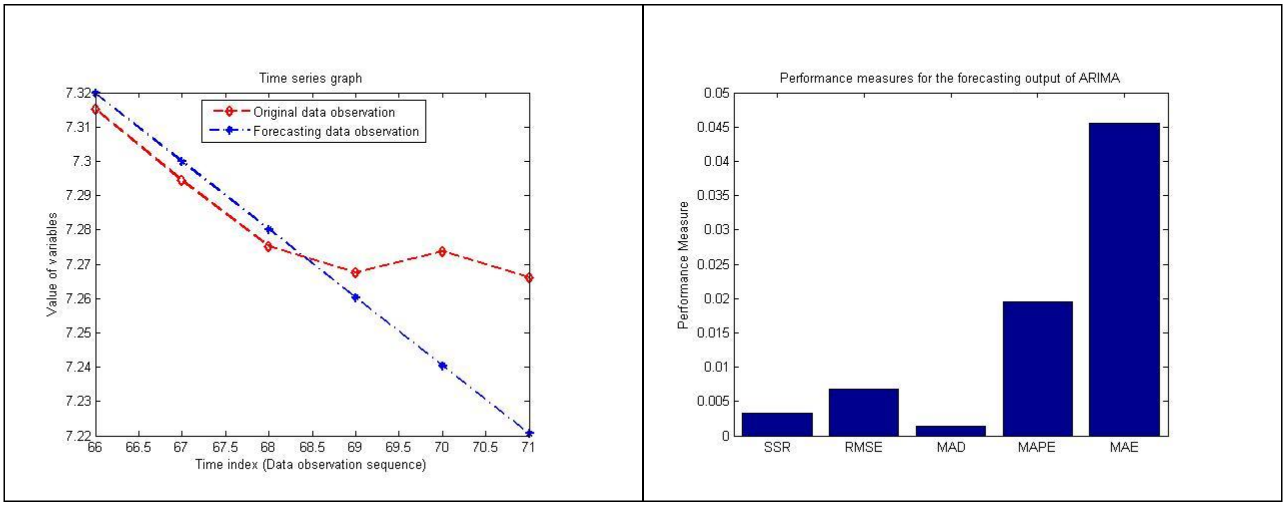

Fig 5: 'Other' data

CONCLUSION

In this study, we have applied the ARIMA model to the M3-competition data set of 3003 time series to determine its forecasting ability. It is observed that the ARIMA model provides better forecasting even when little detail about the data is available. The forecasting accuracy of the ARIMA model is measured using parameters such as 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

I. Boylan, J.E., "Towards a more precise definition of forecastability", Foresight: the International Journal of Applied Forecasting, pp. 34–40, 2009.

II. Kolassa, S. "How to assess forecastability" Foresight: the International Journal of Applied Forecasting, pp. 41–45, 2009.

III. Tashman, L.J., "Special feature on forecastability: Preview", Foresight:

the International Journal of Applied Forecasting, vol. 23, 2009.

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

V. Makridakis, S. and Hibon, M. ”The M-3 Competition: results, conclusions and implications”, International Journal of Forecasting, 16, pp. 451-476, 2002.

VI. P. F. Pai and C. S. Lin, “A hybrid ARIMA and support vector machines model in stock price forecasting,” Omega, vol. 33, no. 6, pp. 497–505, 2005.

VII. G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing, vol. 50, pp. 159–175, 2003.

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

IX. Armstrong, J., & Collopy, F., “Error measures for generalizing about forecast methods: Empirical comparisons”, International Journal of Forecasting, vol. 8, pp. 69–80, 1992.

X. Fildes, R., “The evaluation of extrapolative forecasting methods”, International Journal of Forecasting, vol. 8, pp. 81–98, 1992.

XI. Fildes, R., Hibon, M., Makridakis, S., & Meade, N., “Generalising about univariate forecasting methods: Further empirical evidence”, International Journal of Forecasting, vol. 14, pp. 339–358, 1998.

XII. Grubesic, T., & Murray, A. “Geographies of imperfection in telecommunication analysis”. Telecommunications Policy, vol. 29, pp. 69–94, 2005.

XIII. Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al., “The M-2 competition: A real-time judgmentally based forecasting study", International Journal of Forecasting, vol. 9, pp. 5–23, 1993.

XIV. Parzen, E, "ARARMA models for time series analysis and forecasting", Journal of Forecasting, vol. 1, pp. 67–82, 1982.

XV. Simmons, L. F., "M-Competition — A closer look at Naıve2 and median APE: a note"., International Journal of Forecasting, vol. 4, pp. 457-460, 1986.

XVI. Gardner, E., "Exponential smoothing: The state of the artpart ii", International Journal of Forecasting, vol. 22, pp. 637–666, 1992.

XVII. Kotsialos, A., Papageorgiou, M., Poulimenos, A. "Long-term sales forecasting using Holt-Winters and neural network methods". Journal of Forecasting, vol. 24, pp. 353–368, 2005.

XVIII. Fildes, R. & Petropoulos F. "An evaluation of simple forecasting model selection rules (LUMS Working Paper 2013:2)". Lancaster University: The Department of Management Science.

XIX. Makridakis, S., & Hibon, M, "The M-3 competition: results, conclusion and implications", International Journal of Forecasting, vol. 16, pp. 451-476, 2000.

XX. Mahsin, M, "Modeling rainfall in Dhaka division of Bangladesh using time series analysis", Journal of Mathematical Modelling and Application, vol. 5, pp. 67-73, 2011.

XXI. Shamsnia, S.A., Shahida, N., Liaghat, A., Sarraf, A., and Vahdat, S.F., "Modeling of Weather Parameters Using Stochastic Methods (ARIMA Model)(Case Study: Abadeh Region, Iran)", International Conference on Environment and Industrial Innovation, vol. 12, pp. 282-285, 2011.

XXII. Stephen A. Delurgio and Carl D. Bhame, "Forecasting systems for operations management", 1991.

