Predicting Fault In Telecommunication Network: Lessons Learned

Mohd Izhan Mohd Yusoff[1]

[1]*Telekom Research & Development Sdn Bhd, TM Innovation Center, Lingkaran Teknokrat Timur, 63000 Cyberjaya, Selangor Darul Ehsan, Malaysia*

izhan@tmrnd.com.my

Abstract - *Faults in the telecommunication network has attracted a large amount of interest from researchers and practitioners who have introduced, for example, algorithms to extract faulty signatures from noisy historical event data; rules and decision tree data mining classifiers to upgrade fault detection and handling, and cluster head selection algorithms to address the failure under uncertain situations. This article discusses lessons learned from studying and applying statistical methods or techniques to predict faults in the telecommunication networks (which is our main objective or target). Cochrane Orcutt's approach or procedure was identified for having properties that we believe would fulfill or achieve the above objective or target. The challenges we faced were: real data collected from the device show seasonal trends (meaning the device is faulty periodically), and we showed, by using simulated seasonal data, the Cochrane Orcutt approach or procedure failed to get the desired results. We proposed the Cochrane Orcutt adjusted approach or procedure where we showed, by using simulated seasonal data, the said adjusted approach or procedure managed to get the desired results. We also suggest future research recommendations using advanced methods (especially when met with the ideal case or scenario).*

Keywords - *Telecommunication network, Faults, Cochrane-Orcutt adjusted approach or procedure, Multiple regression, Autocorrelation function (ACF) plot, Partial Autocorrelation function (PACF) plot, Correlation matrix,*

I. Introduction

Wang et al. [9] introduced a new class of indexing fault signatures that encode the temporal evolution of events generated by a network fault and topology relationships among the nodes where these events occur. They presented an efficient learning algorithm to extract such faulty signatures from noisy historical event data. With the help of novel space-time indexing structures, they showed how to perform efficient online signature matching. Further, Rozaki [10] presented a monitoring scheme for mobile networks based on rules and decision tree data mining classifiers to upgrade fault detection and handling. Their goal was to have optimization rules that would improve anomaly detection. In addition, a monitoring scheme that relies on Bayesian classifiers was also implemented for fault isolation and localization.

The data mining techniques described in this paper are intended to allow a system to learn the network fault rules. The results of the tests conducted lead to the conclusion that the rules were highly effective to improve network troubleshooting. Umadevi et al. [12] said that Underwater Acoustic Sensor Networks have a delicate design and proposed a model wherein an energy-efficient fault-tolerant technique was integrated along with customized MAC protocol to build an accurate, fast, and reliable environment for reasonably better performance. They initiated a novel cluster head selection algorithm to address failure under uncertain situations. Moreover, Stelling et al. [14] addressed the difficulty in implementing techniques for detecting and correcting faults in distributed computing systems via the introduction of a fault detection service designed to be incorporated, a modular fashion distributed computing systems, tools, or applications. Asim et al. [15] proposed a new fault management mechanism to deal with fault detection and recovery and proposed a hierarchical structure to properly distribute fault management tasks among sensor nodes by introducing more self-managing functions. Hood et al. [16] proposed an intelligent system using adaptive statistical approaches to detect unknown or unseen faults. The system learns the network's normal behavior, and the information on deviations gets combined in the probabilistic framework of a Bayesian network. Other works can be found in [18], [19], [20], [21] and [22].

The real-time data collected for this article presents challenges, and the approaches utilized (to tackle or solve them) differed from those of the researchers mentioned above. Faulty (telecommunication) cards or ports would raise the alarm (in a network management system), and when this happens, customers start to complain of not being able to access the Internet. The service provider takes immediate action, which includes replacing the faulty cards or ports. The speed of the mentioned actions (including the periods that the customers stay disconnected) depends on the fault's complexity (plus the availability of new cards or ports in the inventory). The main objective and the focus of this article are to predict fault in telecommunication networks. The benefits of predicting faults (with high accuracy) are reduced the number of customer complaints (due to quick actions taken by the service provider in accordance with Service Level Agreement, SLA) and protection of the service provider's image (as well as further service improvement), among others.

This article is divided into several sections: Introduction, Data, Methodology, Results, and Discussion (Lessons Learned).

II. Data & Methodology[1]

The Cochrane–Orcutt (or multiple regression with serial correlation) approach or procedure is represented by the following steps [3]:

Step One: The regression coefficients $y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_p x_{p,t} + \varepsilon_t$, are estimated using ordinary least squares;

Step Two: The serial correlation is estimated from the current residuals ($e_t = y_t - \hat{y}_t$) using the formula $\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=2}^{n} e_{t-1}^2}$;

Step Three: A new set of data is created using the formulas $y_t' = y_t - \hat{\rho} y_{t-1}$, $x_{h,t}' = x_{h,t} - \hat{\rho} x_{h,t-1}$, $h = 1,2,\dots,p$;

Step Four: Ordinary least squares is used to fit the following multiple regression to the transformed data $y_t' = \beta_0' + \beta_1' x_{1,t}' + \cdots + \beta_p' x_{p,t}'$;

Step Five: The regression equation of the untransformed data is created using the following equations $\beta_0 = \frac{\beta_0'}{(1-\hat{\rho})}, \beta_1 = \beta_1', \beta_2 = \beta_2', \dots, \beta_p = \beta_p'$;

Step Six: Step Two till *Four* are repeated until $\hat{\rho}$ stabilizes (usually, only four or five iterations are necessary);

Step Seven: The formula for forecast j periods into the future after the end of the series (period n is the final period on which we have data) is $F_{n+j} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,n+j} + \cdots + \hat{\beta}_p x_{p,n+j} + \hat{\rho}^j e_n$;

Note that in the above steps, $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$ AR(1) can be modified to tackle or handle $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + v_t$, AR(2), $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \rho_3 \varepsilon_{t-3} + v_t$ AR(3), and so forth.

We used the above steps on the following (simulated and seasonal behavior) data (representing real-time data, especially when the telecommunication network devices display faulty behavior periodically, i.e., when a certain condition is met). The symbol or coefficient ρ used in the above steps was replaced by θ in the following paragraphs.

Let *y* represent (simulated and seasonal behavior) observations collected from device *A* and *x* represent (simulated and seasonal behavior) observations collected from device *B* (refer to **Figure 1**). The total number of observations equals 277, and they (i.e., the observations) will reveal whether the device is faulty or not (by referring to the device's Standard Operating Procedure, SOP). Variables *y* and *x* observations are highly correlated at 0.55 [2], and devices *A* and *B* formed an *integrated* (telecommunication) system or network[2]. We aim to forecast *y* given *x*, using each of the models in **Table 1**.

Table 1. List of models used in this study.

MODEL (1)	MODEL (2)	MODEL (3)
$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$,[#] $t = 1,2,\dots,n$	$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, $\varepsilon_t = \theta_1 \varepsilon_{t-1} + v_t$,[##] $t = 1,2,\dots,n$	$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, $\varepsilon_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + v_t$,[###] $t = 1,2,\dots,n$

#AR (0), ##AR (1), ###AR (2), AR = Auto-Regressive

MODEL (1), MODEL (2), and MODEL (3) used Cochrane-Orcutt (or multiple regression with serial correlation) approaches. MODEL (1) skipped steps related to finding θ. MODEL (2) and MODEL (3) used all the above steps. *Step Two, Three,* and *Seven* formulae of the Cochrane-Orcutt procedure were replaced with the following to accommodate MODEL (3): $\hat{\theta} = (A^t A)^{-1} A^t b$,

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_q \end{bmatrix}, \quad A = \begin{bmatrix} e_q & \cdots & e_1 \\ \vdots & \ddots & \vdots \\ e_{n-1} & \cdots & e_{n-q} \end{bmatrix}, \quad b = \begin{bmatrix} e_{q+1} \\ \vdots \\ e_{n-q} \end{bmatrix},$$

$y_t' = y_t - \sum_{i=1}^{q} \hat{\theta}_i y_{t-i}$, $x_{h,t}' = x_{h,t} - \sum_{i=1}^{q} \hat{\theta}_i x_{h,t-i}$, $h = 1,2,\dots,p$, $F_{n+j} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,n+j} + \cdots + \hat{\beta}_p x_{p,n+j} + \sum_{h=1}^{q} \hat{\theta}_h \varepsilon_{n+j-h}$;

We reserved the last 15 observations of *y* for measuring the performance of the tabulated or mentioned models. A model's performance is measured via the following parameters (especially *SSE, MSE, MAPE,* and R^2, which measure the difference between actual and forecast values; and *n* in the parameters is replaced by 15). $SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, $MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, $MAPE = (100\%) \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$, $R^2 = 1 - \frac{SSE}{SST}$, $SST = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$ [2]. R^2 is considered good or better if closer to one. Other parameters are considered good or better if closer to zero.

(a)

(b)

---

[1] All the methods mentioned in this section were incorporated into "expertmlregx.jar." The results presented in this article are produced from the mentioned java program.

[2] We did not utilize boxplot because variable *x* satisfies normal distribution when tested using Kolmogorov-Smirnov One-Sample test [2].
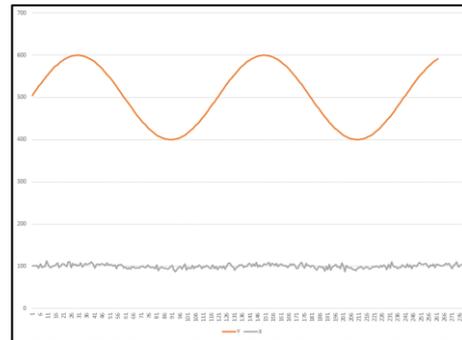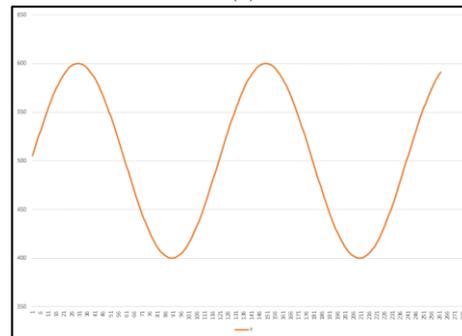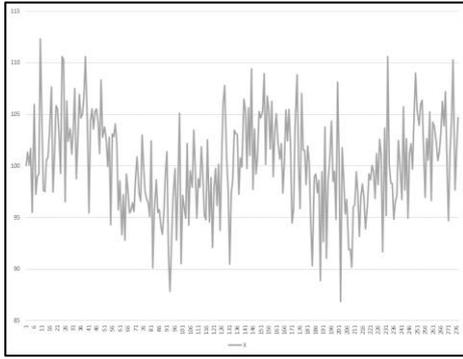
(c)

Figure 1. (a) *y* and *x* are plotted together using line plot (note that seasonal behavior of the latter is hidden and 15 observations omitted from the former); (b) *y* is plotted using line plot, and (c) *x* is plotted using line plot (note that seasonal behavior is displayed or observed).

III. Results

MODEL (1) or AR (0) was performed, and its residuals were plotted using the auto-correlation function (ACF) and partial autocorrelation function (PACF) plots to see if there's a valid reason to use the Cochrane–Orcutt approach. The results from the mentioned plots display AR behavior, and the results give an early indication of AR's *q* being greater than one (1); refer to **Figure 2** [1].
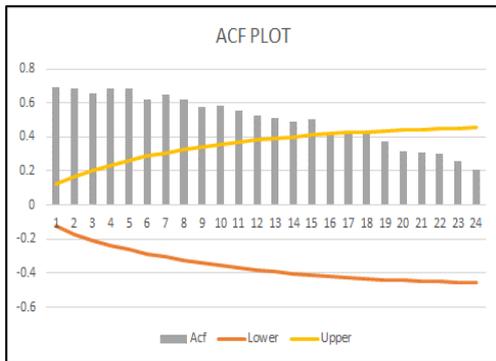
(a)
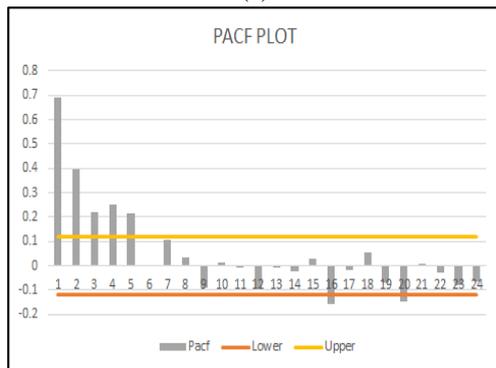
(b)

Figure 2. (a) ACF Plot for MODEL (1) residuals; (b) PACF Plot for MODEL (1) residuals.

MODEL (2) and MODEL (3) required more than (normal or typical) four or five iterations to achieve

convergence. MODEL (1), MODEL (2), and MODEL (3) results can be found in **Table 2**, **Table 3**, and **Figure 3**.

Table 2. Coefficients β_0, β_1, θ_1, and θ_2 are estimated for each model. Hypothesis testing and the Durbin-Watson test are not included in this article, and they are replaced by *SSE*, *MSE*, *MAPE*, and R^2.

MODEL	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
MODEL (1)	-327.54	8.32	.	.
MODEL (2)	3379.85	0.017	0.999885	.
MODEL (3)	499.9999984	7.77E-10	1.9972	-1.00

Table 3. *SSE*, *MSE*, *MAPE*, and R^2 were computed for each model.

MODEL	SSE	MSE	MAPE	R^2
MODEL (1)	89324.47	5954.97	177.71	-1176.52
MODEL (2)	295.67	19.71	10.12	-2.90
MODEL (3)	1.12E-10	7.46E-12	5.61E-06	1.00

(a)

(b)

(c)
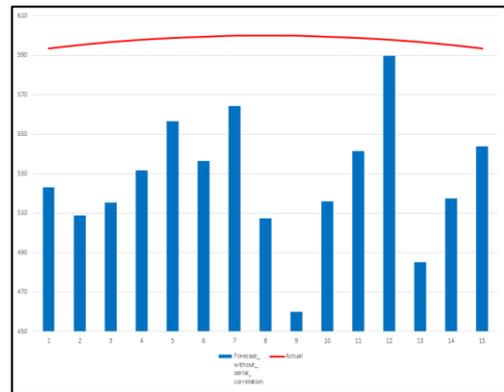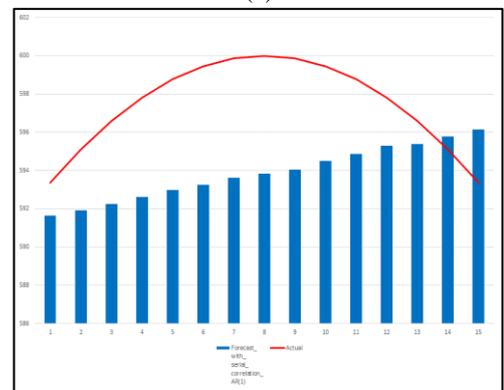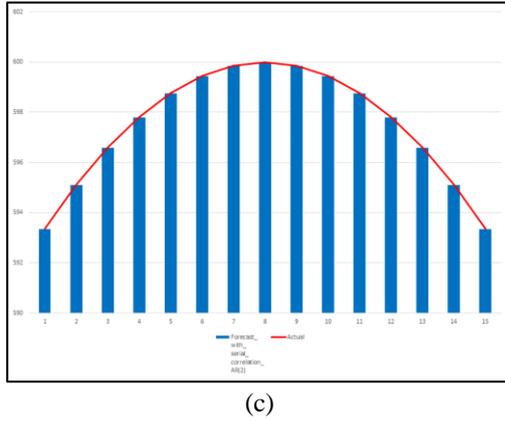
Figure 3. (a) Actual (line plot) versus MODEL (1): $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, t = 1,2,\ldots,n$, (clustered column plot); (b) Actual (line plot) versus MODEL (2): $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \varepsilon_t = \theta_1 \varepsilon_{t-1} + v_t, t = 1,2,\ldots,n$, (clustered column plot); (c) Actual (line plot) versus MODEL (3): $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \varepsilon_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + v_t, t = 1,2,\ldots,n$, (clustered column plot)

In **Figure 3**, a clustered column plot is used or chosen instead of a line plot for forecast values because of its ability to highlight or reveal forecast values *behavior*. Combine **Figure 3** with **Table 3**, and it is *clear or obvious*, MODEL (3) performed better as compared to MODEL (1) and MODEL (2).

IV. Discussion (Lessons Learned)

In the previous section, we highlighted several approaches taken by researchers to handle the fault in telecommunication, introduced the Cochrane–Orcutt approach or procedure (by listing all steps involved), and described (simulated and seasonal behavior) data used in the preparation of this article. The Cochrane–Orcutt approach should be *adjusted*, especially for the case of seasonal behavior data, to include AR (2), AR (3), and so forth (i.e., higher Auto-regressive, AR, parameters) in serial correlation. To measure the models' performance (as well as to introduce a *stopping* mechanism), it is highly recommended to *reserve* several actual observations for the dependent variable, *y*.

Referring to **Table 3,** where the whole process is terminated or stopped at AR (2) of MODEL (3), the models show significant improvement (in terms of *SSE, MSE, MAPE*, and *R^2*) when AR (#) of MODEL (#+1) is changed in the following manner or sequence (from left to right): AR (0), AR (1), and AR (2).

The whole process given in the previous section can be explained via the Cochrane-Orcutt *adjusted* procedure (refer to **Figure 4**), where the *Threshold* value is fixed at 1×10^{-3}. *RSqr*(q)* in **Figure 4** refers to the following formula *|R^2(q)-1|*. Other scenarios were considered, and their details can be found in **Appendix A**.
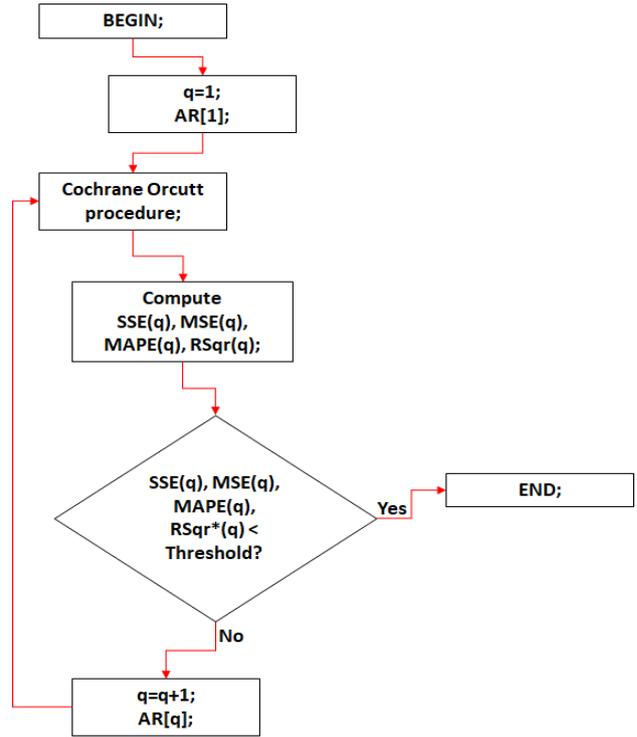
Figure 4. A flow diagram of the Cochrane-Orcutt *adjusted* procedure. *q* starts with one (1) because AR (0) represents multiple regression *without* serial correlation. SSE (0), MSE (0), MAPE (0), and RSqr*(0) derived from AR (0) are subjected to the same rule, i.e., they are checked whether they are less than a *threshold* value or not.

After performing the Cochrane-Orcutt *adjusted* procedure and getting the final AR (*q*), we can use the information (i.e., final AR (*q*)) plus all the observations (including the "reserved" ones) to forecast, say, *j*, periods into the future for variable *y*. We can use time series forecasting techniques such as ARIMA, Brown & Winters exponential smoothing, moving average. Simple regression to forecast the same number of periods into the future for variable *x*, if they (i.e., variable *x* observations) are not available, before proceeding to forecast into the future for variable *y*; further details are given in **Appendix B** [1,17].

We collected real-time data from multiple sources every hour from 1500 hours on June 2, 2016, till 0000 hours on June 19, 2016, and saved it in "csv" format. Several processes were performed on the real-time data (including dividing the real-time data according to log in ID, re-arranging the time when the observations were collected, and matching the alarm raised with the arranged observations) to get the desired format. The mentioned processes emphasized or focused on alarms closely related or linked to Customer Trouble Tickets (CTTs) and factors that caused problems to customers and led to the creation of CTTs. The behavior displayed by one of the variables under study is best described as follows: *all observations were fluctuating between 15 and 15.5 for several hours or days before dropping below 15, which trigged the alarm; the fluctuating continued between 13.5 and 14.5 for several hours or days before the alarm stopped; the*

observations were back or returned to normal (i.e., fluctuating between 15 and 15.5).

Unfortunately, the number of incidents (i.e., alarms raised) is not enough for us to forecast fault[3] (where they should follow either hotel's occupancy that shows seasonal behavior or simulated and seasonal behavior data presented in this article). Scatterplots derived from the real-time data showed consistent behaviors with the assumptions about the effects on the variables from the alarms raised. Boxplots and ACF plots (plus PACF plots) showed outliers and serial correlation, respectively. The correlation matrix helped determine which variables are useful in the next analysis (i.e., multiple regression with serial correlation). Using the selected (pair of) variables, multiple regression with serial correlation performed better than the conventional or ordinary method (i.e., multiple regression without serial correlation) in terms of *actual modeling* values (details not included in this article)[4].

Future research works will involve the analysis of the real-time data collected beyond the time period used in this article, the study of the weather data (which will be treated as external or predictor variable in multiple regression with serial correlation), the study of observations related to network performance collected from network cards that serve, say *n*, number of customers, and the use of Hidden Markov Models (a probabilistic model consisting of variables representing observations, variables that are hidden, the initial distribution, transition matrix, and parameters for all observation distributions) in predicting fault in telecommunication [4,5,6,7,8,13].

References
[1] GEP Box, G.M. Jenkins, Time series analysis, forecasting and control, *Golden-Day*, San Francisco, 1976.
[2] L. Sachs, Applied statistics: A handbook of techniques, Second edition, *Springer-Verlag*, New York, 1984.
[3] D. Cochrane, G.H. Orcutt, Application of least squares regression to relationships containing auto-correlated error terms, *Journal of the American Statistical Association*, 44(245): 32–61.doi:10.1080/01621459.1949.10483290, 1949
[4] M.R. Spiegel, Shaum's outline series: Theory and problems of advanced calculus: SI (Metric) edition, *McGraw-Hill*, Inc, 1974.
[5] K. V. Mardia, J. T. Kent, J. M. Bibby, Multivariate analysis, *Academic Press Inc. (London) Ltd*, 1979.
[6] J. A. Bilmes, A Gentle Tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models, Technical Report, *University of Berkeley*, ICSI-TR-97-021, 1998.
[7] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal Royal Statistics Society*, vol. 39, no. 1, pp. 1–38, 1977.

[8] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
[9] T. Wang, M. Srivatsa, D. Agrawal, L. Liu, Learning, indexing, and diagnosing network faults, *Proceedings of the 15th ACM SIGKDD International Conference on knowledge discovery and data mining*, pp. 857–866, 2009.
[10] E. Rozaki, Design and implementation for automated network troubleshooting using data mining, *International journal of data mining and knowledge management process*, vol. 5., no. 3, 2015.
[11] K. S. Umadevi, R. Jagadeesh Kannan, F-MAC: An optimal clustering-based fault-tolerant system for underwater acoustic sensor networks, *International Journal of Pure and Applied Mathematics*, vol. 109, no. 8, pp. 177–184, 2016.
[12] M. I. M. Yusoff, I. Mohamed, M. R. Abu Bakar, Hidden Markov models: An insight, *Proceedings of the 6th International Conference on Information Technology and Multimedia*, doi: 10.1109/ICIMU.2014.7066641, 2014.
[13] P. Stelling, I. Foster, C. Kesselman, C. Lee, G.V. Laszewski, A fault detection service for wide-area distributed computations, *Cluster Computing*, vol. 2, no. 2, pp. 117–128, doi: 10.1023/A:101907040, 1999.
[14] M. Asim, H. Mokhtar, M. Merabti, A self-managing fault management mechanism for wireless sensor networks, *International Journal of Wireless & Mobile Networks*, vol. 2, no. 4, pp. 184–197, 2010.
[15] C.S. Hood, C. Ji, Proactive network fault detection, *Proceedings of INFOCOM '97*, doi: 10.1109/INFCOM.1997.631137, 1997.
[16] D.C. Montgomery, Forecasting and time series analysis, *McGraw-Hill*, New York, 1976.
[17] Dr. S Gajanana, Yayavaram Revanth Sai, K Rahul, S Rohith Yadav, Inflation Based replacement model for cutting tools using Markov Stochastic process, *International Journal of Engineering Trends and Technology,* vol. 68, no. 2, pp. 29-35, 2020.
[18] Mbamaluikem Peter O, Bitrus, Irmiya, Okeke, Henry S., Asymmetrical Fault Recognition System on Electric Power Lines Using Artificial Neural Network, *International Journal of Engineering Trends and Technology,* vol. 67, no. 11, pp. 61-66, 2019.
[19] Ankesh Dabhade, Rahul Kale, Saurabh Fulzele, Prof. K.S.Kalkonde, Railway Track Crack Fault Detection using Method of Histograms in Image Processing of MATLAB, *International Journal of Engineering Trends and Technology (IJETT)*, vol. 58, no. 3, pp. 130-136, April 2018.
[20] Ayokunle A. Awelewa, Peter O. Mbamaluikem, Isaac A. Samuel, Artificial Neural Networks for Intelligent Fault Location on the 33-Kv Nigeria Transmission Line, *International Journal of Engineering Trends and Technology (IJETT)*, vol. 54, no. 3, pp. 147-155, December 2017.
[21] Fauzy Che Yayah, Khairil Imran Ghauth, Choo-Yee Ting, Adopting Big Data Analytics Strategy in Telecommunication Industry, *Journal of Computer Science & Computational Mathematics*, vol. 7, no. 3, pp. 57-67, 2017.

Appendix A

Four scenarios were considered where for each scenario, 1000 samples (each sample consists of 277 observations) were generated using a random number generator.

Scenario 1: The correlation between variables *x* and *y* is "weak" (at 0.4)[5]. **Figure A.1** shows the behavior of variables *x* and *y,* especially the "weak" correlation (or linkage) between them (the observations are clustered together, which resembles an ellipse shape). *SSE, MSE, MAPE,* and R^2, r_{xy}, μ_x, σ_x, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\theta}_1$, and $\hat{\theta}_2$ are calculated for each of the 1000 samples. The *1000* calculations are

---

[3] At least two incidents are required to forecast fault.
[4] Flow diagram in **Figure 4** was utilized where AR's were limited to AR(0) and AR(1) only. Real-time data is strictly confidential.

[5] Based on rule of thumb, correlation above 0.5 is considered "strong" and "weak" otherwise. Hypothesis testing $\rho = 0$ is not included in this article. Here ρ refers to correlation and not to be confused with the one given in **Data & Methodology** section.

summarized in the form of minimum, average, standard deviation, and maximum; refer to tables **A.1**, **A.2**, and **A.3**. Note that *SSE, MSE*, and *MAPE*, minimum, average, and maximum are close to zero, whereas R^2 close to one. The said behaviors, which represent the flow diagram's performance in **Figure 4**, are further verified (or visualized) in part by **Figure A.2**.

(a)

(b)
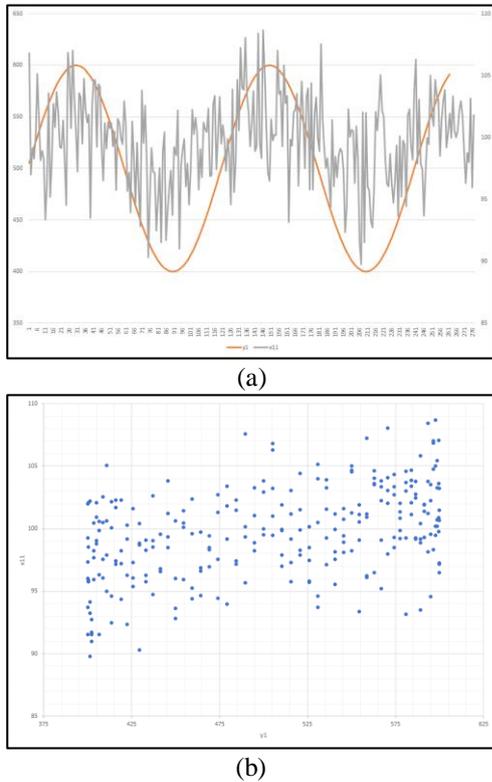
Figure A.1. Variables *y* and *x* of the selected sample are plotted together using (a) line plot where the left y-axis is reserved for variable *y* and right y-axis for variable *x* (b) scatter plot where y-axis and x-axis are reserved for variables *x* and *y*, respectively.

Table A.1. Minimum, average, standard deviation and maximum are calculated for each *SSE, MSE, MAPE*, and R^2.

	SSE	*MSE*	*MAPE*	R^2
Minimum	*1.72E-13*	*1.14E-14*	*2.31E-07*	*1.00E+00*
Average	*9.73E-12*	*6.49E-13*	*1.42E-06*	*1.00E+00*
Standard deviation	*1.2297E-11*	*8.2E-13*	*7.6E-07*	*1.6E-13*
Maximum	*7.78E-11*	*5.18E-12*	*4.74E-06*	*1.00E+00*

Table A.2. Minimum, average, standard deviation and maximum are calculated for each $r_{y,x}$, μ_x, and σ_x. Note that $r_{y,x}$, represents correlation coefficient (based on average, 0.39); μ_x, and σ_x represent the average and standard deviation for variable *x*.

	$r_{y,x}$	μ_x	σ_x
Minimum	*0.25653*	*99.019202*	*3.0408782*
Average	*0.39364*	*99.619892*	*3.5511861*
Standard deviation	*0.04965*	*0.2004639*	*0.1482002*
Maximum	*0.53929*	*100.22643*	*3.9370558*

Table A.3. Minimum, average, standard deviation and maximum are calculated for each $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\theta}_1$, and $\hat{\theta}_2$.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$
Minimum	*499.9999987*	*-3.3E-09*	*1.997259*	*-1*
Average	*500.0000005*	*1.9E-12*	*1.997259*	*-1*
Standard deviation	*5.84931E-07*	*7.83E-10*	*3.52E-09*	*3.4E-09*
Maximum	*500.0000019*	*2.28E-09*	*1.997259*	*-1*

(a)

(b)
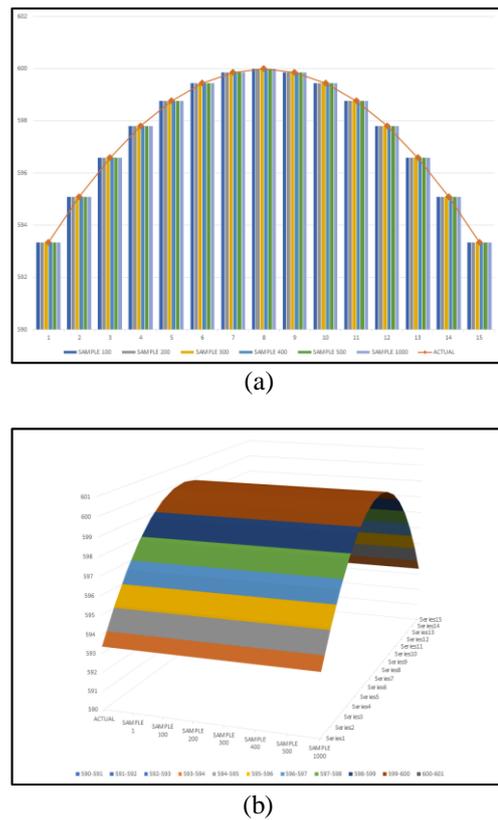
Figure A.2. Actual values are plotted against forecast values, using (a) line combined with clustered column plot and (b) surface plot, where 6 samples were selected for the latter, namely Sample 100, 200, 300, 400, 500, and 1000.

Scenario 2: The correlation between variables *x* and *y* is "strong" (at 0.8). **Figure A.3** shows the behavior of variables *x* and *y*, especially the "strong" correlation (or linkage) between them (the observations are clustered together, which resembles a linear trend shape). *SSE, MSE, MAPE,* and R^2, r_{xy}, μ_x, σ_x, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\theta}_1$, and $\hat{\theta}_2$ are calculated for each of the 1000 samples. The *1000* calculations are *summarized* in the form of minimum, average, standard deviation, and maximum; refer to tables **A.4**, **A.5**, and **A.6**. Note that *SSE, MSE*, and *MAPE*, minimum, average, and maximum are close to zero, whereas R^2 close to one. The said behaviors, which represent the flow diagram's performance in **Figure 4**, are further verified (or visualized) in part by **Figure A.4**.
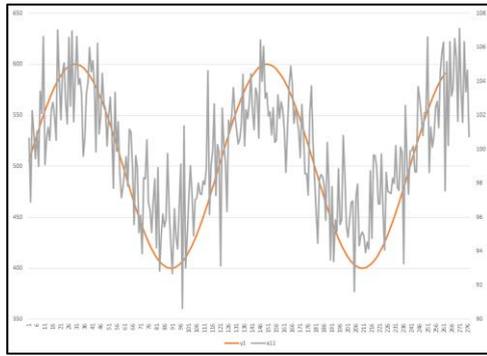
(a)
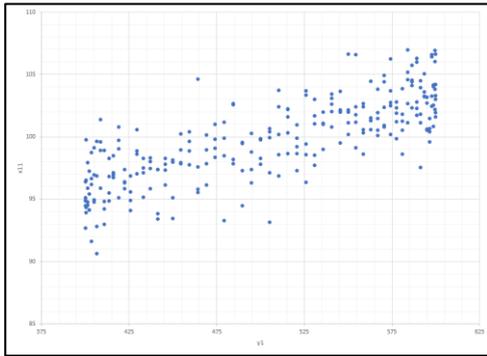
(b)

Figure A.3. Variables *y* and *x* of the selected sample are plotted together using (a) line plot where the left y-axis is reserved for variable *y* and right y-axis for variable *x* (b) scatter plot where y-axis and x-axis are reserved for variables *x* and *y*, respectively.

Table A.4. Minimum, average, standard deviation and maximum are calculated for each *SSE, MSE, MAPE,* and R^2.

	SSE	*MSE*	*MAPE*	R^2
Minimum	*1.26E-13*	*8.38E-15*	*2.07E-07*	*1.00E+00*
Average	*6.03E-12*	*4.02E-13*	*1.25E-06*	*1.00E+00*
Standard deviation	*1.136E-11*	*7.6E-13*	*5.7E-07*	*1.5E-13*
Maximum	*1.15E-10*	*7.64E-12*	*5.68E-06*	*1.00E+00*

Table A.5. Minimum, average, standard deviation and maximum are calculated for each $r_{y,x}$, μ_x, and σ_x. Note that $r_{y,x}$, represents correlation coefficient (based on average, 0.8); μ_x, and σ_x represent the average and standard deviation for variable *x*.

	$r_{y,x}$	μ_x	σ_x
Minimum	*0.73468*	*99.270568*	*3.2056572*
Average	*0.79673*	*99.811673*	*3.5687636*
Standard deviation	*0.01875*	*0.1317101*	*0.1178905*
Maximum	*0.85432*	*100.23361*	*4.0237565*

Table A.6. Minimum, average, standard deviation and maximum are calculated for each $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\theta}_1$, and $\hat{\theta}_2$.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$
Minimum	*499.9999985*	*-4.4E-09*	*1.997259*	*-1*
Average	*500.0000001*	*-2.9E-11*	*1.997259*	*-1*
Standard deviation	*3.12866E-07*	*1.22E-09*	*1.84E-09*	*1.78E-09*
Maximum	*500.0000017*	*3.8E-09*	*1.997259*	*-1*
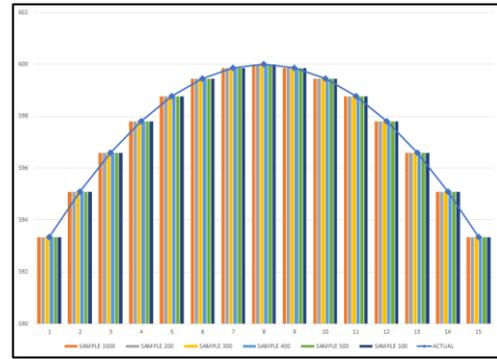
(a)
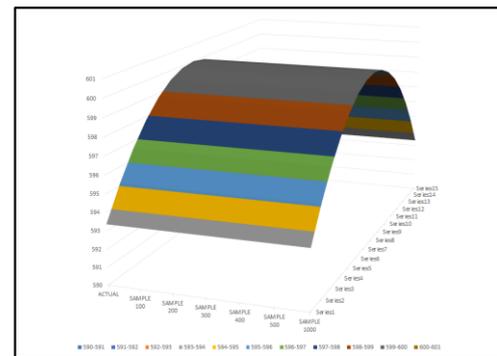
(b)

Figure A.4. Actual values are plotted against forecast values, using (a) line plot combined with clustered column plot and (b) surface plot, where 6 samples were selected for the latter, namely Sample 100, 200, 300, 400, 500, and 1000.

Scenario3: The correlation between variables *x* and *y* are "weak" (at 0.4). Here the device is considered faulty when its measurements dipped below the threshold value. **Figure A.5** shows the behavior of variables *x* and *y,* especially the "weak" correlation (or linkage) between them (the observations clustered together, which resembles an ellipse shape). *SSE, MSE, MAPE,* and R^2, r_{xy}, μ_x, σ_x, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\theta}_1$, and $\hat{\theta}_2$ are calculated for each of the 1000 samples. The *1000* calculations are *summarized* in the form of minimum, average, standard deviation, and maximum; refer to tables **A.7**, **A.8**, and **A.9**. Note that *SSE, MSE*, and *MAPE*, minimum, average, and maximum are close to zero, whereas R^2 close to one. The said behaviors, which represent the flow diagram's performance in **Figure 4**, are further verified (or visualized) in part by **Figure A.6**. Notice that the forecast values managed to preserve or capture the *convex* shape of actual values.
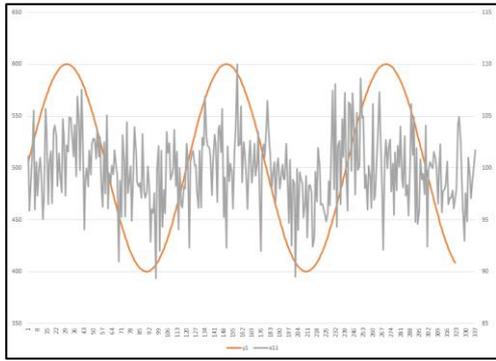
(a)
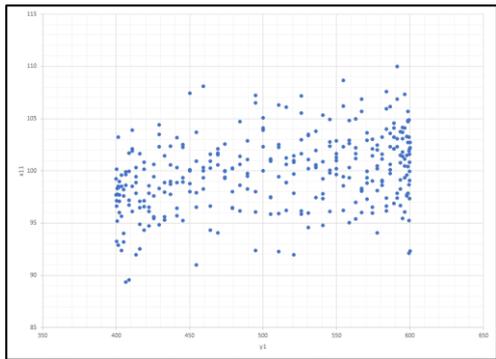
(b)

Figure A.5. Variables *y* and *x* of the selected sample are plotted together using (a) line plot where the left y-axis is reserved for variable *y* and right y-axis for variable *x* (b) scatter plot where y-axis and x-axis are reserved for variables *x* and *y*, respectively.

Table A.7. Minimum, average, standard deviation and maximum are calculated for each $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\theta}_1$, and $\hat{\theta}_2$.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$
Minimum	499.999992	-2.5E-09	1.997259	-1
Average	499.9999983	-2.5E-11	1.997259	-1
Standard deviation	1.95118E-06	7.11E-10	5.17E-09	5.6E-09
Maximum	499.9999999	2.27E-09	1.997259	-1

Table A.8. Minimum, average, standard deviation and maximum are calculated for each $r_{y,x}$, μ_x, and σ_x. Note that $r_{y,x}$, represents correlation coefficient (based on average, 0.39); μ_x, and σ_x represent the average and standard deviation for variable *x*.

	$r_{y,x}$	μ_x	σ_x
Minimum	0.231888	98.9162	3.169821
Average	0.393431	99.49665	3.555339
Standard deviation	0.046069	0.17712	0.133561
Maximum	0.520473	100.0899	3.98267

Table A.9. Minimum, average, standard deviation and maximum are calculated for each *SSE, MSE, MAPE,* and R^2;

	SSE	*MSE*	*MAPE*	R^2
Minimum	6.78E-12	4.52E-13	2.16E-06	1.00E+00
Average	1.67E-10	1.12E-11	7.19E-06	1.00E+00
Standard deviation	3.13885E-10	2.09E-11	6.46E-06	4.14E-12
Maximum	1.46E-09	9.74E-11	2.82E-05	1.00E+00
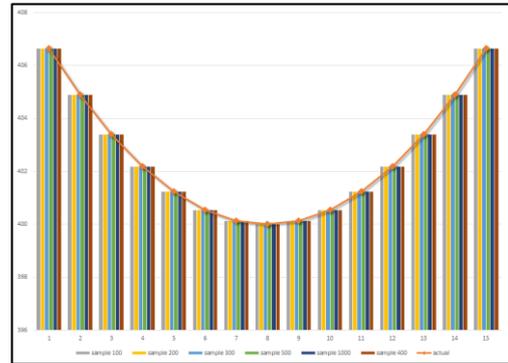
(a)
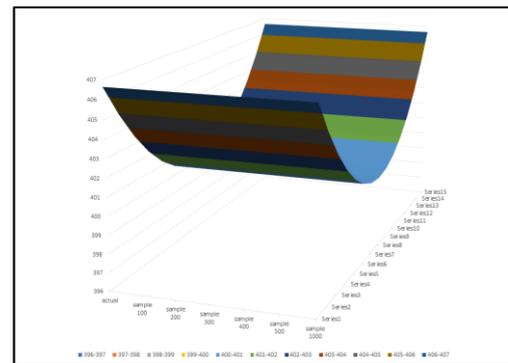
(b)

Figure A.6. Actual values are plotted against forecast values, using (a) line plot combined with clustered column plot and (b) surface plot, where 6 samples were selected for the latter, namely Sample 100, 200, 300, 400, 500, and 1000.

Scenario 4: Variables *x1, x2, x3, x4,* and *x5*, are collected from *B1, B2, B3, B4,* and *B5*. All of the said devices are linked to device *A*. Variable *y* is collected from device *A*. Correlation between variables *x1* and *y*; *x2* and *y*; *x3* and *y*; *x4* and *y*; varies from 0.2 to 0.6 (i.e., from "weak" to "strong"). Correlation between *x1* and *x2*; *x1* and *x3*; *x1* and *x4*; *x1* and *x5*; *x2* and *x3*; *x2* and *x4*; *x2;* and *x5*; *x3* and *x4*; *x3* and *x5*; *x4;* and *x5* are fixed at zero to comply with multiple regression rules, thus ensuring hypothesis testing could be performed on the estimated coefficients; refer to **Figure A.7**. Out of 1000 samples, 996 samples succeeded in following the diagram in **Figure 4** using all the said variables at once. They are *summarized* in the form of minimum, average, standard deviation, and maximum refer to tables **A.10**, **A.11**, and **A.12**. Out of 996 successful

samples, 4 samples require more than the normal AR(2) to achieve the desired results; refer to table **A.12 (b)**. Note that *SSE, MSE,* and *MAPE*, minimum, average, and maximum are close to zero, whereas R^2 close to one. The said behaviors, which represent the flow diagram's performance in **Figure 4**, are further verified (or visualized) in part by **Figure A.8**. Out of 1000 samples, 4 samples failed due to the denominator of $\beta_0 = \frac{\beta'_0}{(1-\sum\hat{\theta})}$ i.e. $(1-\sum\hat{\theta})$ (refer to **Data & Methodology** section) equals to zero. The variables of 4 failed samples are split or sliced pairwise, i.e., *x1* and *y; x2* and *y; x3* and *y; x4* and *y; x5* and *y* where they (i.e., pairwise variables) managed to follow the diagram in **Figure 4** successfully. Their results are similar to the previous scenarios; therefore, they are not included in this article.
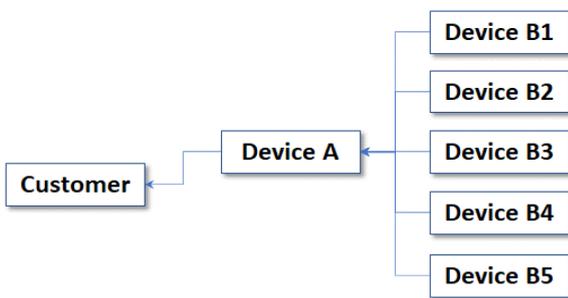
Figure A.7. Telecommunication devices *B1, B2, B3, B4, B5,* are linked to devise *A*, and device *A* is linked to the customer.

Table A.10. Minimum, average, standard deviation and maximum are calculated for each *SSE, MSE, MAPE,* and R^2.

	SSE	*MSE*	*MAPE*	R^2
Minimum	*1.94E-13*	*1.29E-14*	*2.67E-07*	*1.00E+00*
Average	*3.06E-09*	*2.04E-10*	*3.63E-06*	*1.00E+00*
Standard deviation	*4.94848E-08*	*3.3E-09*	*2.47E-05*	*6.52E-10*
Maximum	*1.01E-06*	*6.74E-08*	*4.54E-04*	*1.00E+00*
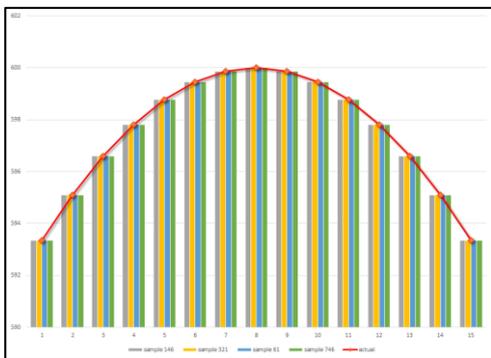
Figure A.8. Actual values are plotted against forecast values, using (a) line plot combined with clustered column plot and (b) surface plot, where 4 samples were selected for the latter, namely Sample 146, 321, 61, and 746. The said samples require more than the normal AR(2) to achieve the desired results.

Table A.11. Minimum, average, standard deviation, and maximum are calculated for each *(a)* μ_z, and σ_z (b) $r_{a,b}$. Note that $r_{a,b}$, represents correlation coefficient; μ_z, and σ_z represent the average and standard deviation for variable *z*.

(a)

	μ_{x1}	σ_{x1}	μ_{x2}	σ_{x2}	μ_{x3}	σ_{x3}
Minimum	995.0291	52.41646	898.2732	42.23347	798.0564	33.91225
Average	1004.863	60.13526	907.0653	50.18718	805.4563	40.18768
Standard deviation	2.952437	2.425973	2.600199	2.088673	2.259255	1.67034
Maximum	1013.683	67.37091	914.8974	58.1891	812.9204	45.16435

	μ_{x4}	σ_{x4}	μ_{x5}	σ_{x5}
Minimum	698.7547	26.5503	598.8464	17.14297
Average	704.0971	30.20018	603.0362	20.14551
Standard deviation	1.717054	1.279315	1.202868	0.846544
Maximum	710.4054	34.77309	606.7339	23.31273

(b)

	$r_{y,x1}$	$r_{y,x2}$	$r_{y,x3}$	$r_{y,x4}$	$r_{y,x5}$
Minimum	0.439166	0.334311	0.226186	0.097315	0.03823
Average	0.595367	0.496509	0.392397	0.29473	0.194312
Standard deviation	0.035726	0.042794	0.0502	0.057786	0.057073
Maximum	0.712198	0.615725	0.554684	0.473513	0.373082
		$r_{x1,x2}$	$r_{x1,x3}$	$r_{x1,x4}$	$r_{x1,x5}$
Minimum		-0.16549	-0.16212	-0.20519	-0.17987
Average		0.005126	0.000255	0.00138	-0.0019
Standard deviation		0.053711	0.057545	0.060186	0.058719
Maximum		0.163311	0.185158	0.180422	0.162974
			$r_{x2,x3}$	$r_{x2,x4}$	$r_{x2,x5}$
Minimum			-0.1755	-0.16357	-0.20547
Average			0.000624	0.002144	0.00244
Standard deviation			0.058468	0.058793	0.057729

Maximum			0.20218	0.177522	0.266674
				$r_{x3,x4}$	$r_{x3,x5}$
Minimum				-0.16662	-0.1786
Average				-0.00055	-0.00079
Standard deviation				0.061262	0.060788
Maximum				0.186432	0.212014
					$r_{x4,x5}$
Minimum					-0.18529
Average					-0.00176
Standard deviation					0.059689
Maximum					0.194791

Table A.12. Minimum, average, standard deviation, and maximum are calculated for each $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$, $\hat{\beta}_5$, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$ derived from (a) 992 samples, and (b) 4 samples that require more than the normal AR(2) to achieve the desired results.

(a)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Minimum	499.9999977	-3E-10	-3.6E-10	-3.9E-10	-4E-10	-6.2E-10
Average	499.9999998	2.99E-12	1.37E-12	2.18E-12	3E-12	1.77E-12
Standard deviation	5.51278E-07	9.18E-11	9.74E-11	1.05E-10	1.14E-10	1.45E-10
Maximum	500.0000023	2.94E-10	3.46E-10	3.9E-10	3.42E-10	4.96E-10

	$\hat{\theta}_1$	$\hat{\theta}_2$				
Minimum	1.997259	-1				
Average	1.997259	-1				
Standard deviation	2.67E-09	2.58E-09				
Maximum	1.997259	-1				

(b)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Minimum	503.9659	-8.2E-11	-2.2E-10	-2.3E-10	-2E-10	-1.6E-10
Average	504.4082	-2.1E-11	-1.4E-10	-8.8E-11	-1.4E-10	1.42E-11
Standard deviation	0.509503	6.55E-11	9.52E-11	1.06E-10	6.47E-11	1.39E-10
Maximum	505.1422	6.3E-11	-4.5E-12	9.97E-12	-7.7E-11	1.59E-10

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$			
Minimum	2.997258	-2.99726	0.999999			
Average	2.997259	-2.99726	1			
Standard deviation	4.31E-07	8.66E-07	4.32E-07			
Maximum	2.997259	-2.99726	1			

Appendix B
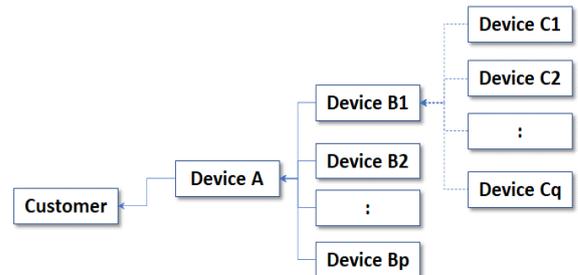
Figure B.1. An example of a telecommunication network (or system) where the customer is being "served" by device *A* where it (i.e. device *A*) depends on (or it is supported by) devices *B1*, *B2*,…,*Bp*.

Time series forecasting techniques, such as ARIMA, Brown & Winters exponential smoothing, moving average, and simple regression, are utilized to forecast into the future for device *B1* if the dotted lines (or links) in **Figure B.1** are not available [1,17]. If they (i.e. the dotted lines or links) are available, before proceeding to forecast into the future for device *A*, the flow diagram in **Figure 4** is utilized to forecast into the future for device *B1* using measurements (or data) collected from devices *C1*, *C2*,…, and *Cq*.