Autocorrelated process control: Geometric Brownian Motion approach versus Box-Jenkins approach

R M Salleh, N I Zawawi, Z F Gan and M E Nor

Department of Mathematics and Statistics, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia

Abstract. Existing of autocorrelation will bring a significant effect on the performance and accuracy of process control if the problem does not handle carefully. When dealing with autocorrelated process, Box-Jenkins method will be preferred because of the popularity. However, the computation of Box-Jenkins method is too complicated and challenging which cause of time-consuming. Therefore, an alternative method which known as Geometric Brownian Motion (GBM) is introduced to monitor the autocorrelated process. One real case of furnace temperature data is conducted to compare the performance of Box-Jenkins and GBM methods in monitoring autocorrelation process. Both methods give the same results in terms of model accuracy and monitoring process control. Yet, GBM is superior compared to Box-Jenkins method due to its simplicity and practically with shorter computational time.

1.0 Introduction

There is a major change in modern manufacturing industries when ‘surpass customer expectation’ becomes a philosophy of quality since the late 1990s. Industries believe that the importance to stay competitive is by producing not only the high quality of process and products but also a creative, innovative and useful with pleasing unexpected features. However, those criteria are not likely to be static, and will certainly be changed based on time and demands. In practice, a fundamental idea to improve the quality of process and products is realized by reducing the process variability. If the process variability is small then the quality will be high.

The most popular control charts developed by Shewhart have wide application in industry for monitoring and improving the manufacturing process with the assumption that observations are uncorrelated. However, that assumption is not always valid and successive observations show autocorrelation. [1] mentioned that in manufacturing system, one of the most challenging issue is autocorrelation,[2] and [3] observed that the presence of autocorrelation among observations leads to false alarms and misleading conclusions about the control state of the process. So, it is necessary to eliminate the autocorrelation before use any statistical process control to detect the assignable causes [4]. If the existing of autocorrelation will not take care carefully, it will affect the quality of the products and cause huge loss for manufacturing industries. There are two major approaches in dealing with autocorrelated process data in process control; i) residual-based approaches and ii) methods that modify control limits to adjust for autocorrelation. That is the main focus for this paper.

In order to control autocorrelation in a process, usually, Box-Jenkins method will be considered first due to its popularity as can be seen in [5], [6], [7], [8], [9], [10] and [3]. Box-Jenkins method uses the concept of three components which are component of autoregressive (AR), the integration filter (I) and the moving average (MA). However, Box-Jenkins has its own disadvantages. First, this method need of time series data with a minimum of 50 observations which is sometimes difficult to obtain when
the value measured is the only year [11]. Besides that, Box-Jenkins method is more complicated to apply [12] and time-consuming [13] which is not suitable choice for a practitioners that need a fast result and prompt action. [6] mentioned that we just need suitable transformation instead of using sophisticated time series modelling. Hence, another alternative method is proposed to overcome this problem which is Geometric Brownian Motion (GBM). The reason why GBM is preferred due to it is relatively easy to implement and it does not use much data to predict the future [14]. The application of GBM can be seen in [15], [16], [14], and [17].

The objectives of this research are to develop GBM and Box-Jenkins model and the process quality will be monitored by using both methods. The accuracy of both models also will be compared and measured using mean absolute percentage error. To start with, in the next section we begin our discussion with the methodology of Box-Jenkins and GBM method. Section 3 presents the models obtained and monitoring process control. Conclusion in Section 4 will close this discussion.

2.0 Methodology

In this research, a set of univariate manufacturing data which is furnace temperature data was used. An oven that can heat to above 1000°C (1832°F) is in a term as a ceramic furnace. Hence, a total of eighty observations of furnace’s temperature was recorded. These observations record the consecutive hourly temperature reading of thermocouple that placed inside the large ceramics furnace. These eighty-hour time segment represented a stable period that the process engineers used as a baseline for a good process performance.

2.1 Box-Jenkins Method

Box-Jenkins analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models. This practice can be seen in the literature such as, for example, [18] and [19]. The general non-seasonal time series is denoted as ARIMA \((p,d,q)\):

\[
x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}
\]

where \(\phi\) and \(\theta\) are coefficients, \(d\) is the degree of the first differencing involved, \(p\) and \(q\) are the orders of autoregressive and moving average polynomials, respectively.

Box-Jenkins method refers to the application of three steps which are identification, estimation, and diagnostic checking. Model identification to select values of \(d\) and then \(p\) and \(q\) in the ARIMA \((p,d,q)\) model. Mann-Kendall test [20] and Dickey-Fuller test [21] will be used to test the data behaviors. Mann-Kendall test is used to identify trend behavior, meanwhile Dickey-Fuller test is used to verify the stationary characteristic.

After done the testing, autocorrelation plot (ACF) and partial autocorrelation plot (PACF) will be used to choose a suitable model for the data based on identifying the pattern of the plots. The following table shows the way to identify the model [22].

| Model               | Autocorrelation | Partial Correlation |
|---------------------|-----------------|---------------------|
| ARIMA \((p,d,0)\)   | Tails off.      | Cuts off after \(p\) lags. |
| ARIMA \((0,d,q)\)   | Cuts off after \(q\) lags. | Tails off. |
| ARIMA \((p,d,q)\)   | Tails off.      | Tails off.          |

Next step is an estimation process. In this step, the parameter for Box-Jenkins model will be estimated. The final step is diagnostic checking. The checking is based on studying the autocorrelation plots of the residuals. Another test will be conducted by using Minitab is Ljung-Box test. Ljung-Box test is used to determine the correlation of error in ARIMA model. Besides that, mean absolute
percentage error (MAPE) is used to check the percentage error or evaluate the accuracy of a model. Value of MAPE will be interpreted by referring to Table 2 [23].

**Table 2. Prediction accuracy**

| MAPE     | Meaning               |
|----------|-----------------------|
| Less than 10% | Highly accurate   |
| 11% to 20%    | Good forecast        |
| 21% to 50%    | Reasonable forecast  |
| More than 51% | Inaccurate forecast  |

Moreover, Akaike information criterion (AIC) will also be conducted by using Microsoft Excel add-in to find the best model. These steps are applied iteratively until step three does not produce any improvement in the model.

2.2 Geometric Brownian Motion (GBM)

GBM is an extension of Brownian Motion [14]. The details fundamental of GBM can be found in [23]. The variable $x_t$ follows GBM if it satisfies the Stochastic Differential Equation (SDE):

$$dx_t = \mu x_t dt + \sigma x_t dw_t$$

(2)

where

i. $dw_t = \epsilon_t \sqrt{dt}$ is the Wiener process,

ii. $\mu$ and $\sigma$ are the drift parameter and variance parameter respectively,

iii. $\epsilon_t$ has identical independent standard normal distribution.

Let $x_0$ be an initial value, the solution of equation (2) can be written as:

$$x_t = x_0 e^{\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma \epsilon_t}$$

(3)

Hence, for each $t$, $x_t$ has a lognormal distribution.

For dependent data, transformation of log-return is needed and shown as below,

$$R_t = \ln\left(\frac{x_t}{x_{t-1}}\right)$$

(4)

where $R_t$ is logarithmic return for any time $t$. However, in general $\{X_t\}$ is a GBM process if $\{R_t\}$ is an AR(1) process with constant $c$. Then GBM model implies that

$$R_{t+1} = c + \theta R_t + \epsilon_t$$

(5)

Eq (5) specifies the logarithmic return at time $n$ in terms of the logarithmic return one time period earlier, it is so called AR (1) which is autoregressive with the parameter, $p = 1$. Under this model, the predicted value of $X_t$ is

$$\hat{X}_t = \exp(\hat{c}).X_{t-1}\left(\frac{X_{t-1}}{X_{t-2}}\right)^{\hat{\theta}}$$

(6)

where $\hat{c}$ and $\hat{\theta}$ are the maximum likelihood estimates in the regression model in Eq (5).

The construction of GBM starts by checking whether the data set is under GBM properties which are i) test that the log-returns are independent and identically normally distributed, ii) test the normality of log
returns [23]. To confirm the absence of autocorrelation in (i) and (ii), Durbin Watson Test and Anderson Darling test are used, respectively. There is no need to monitor the data using complicated models if the two assumptions fulfilled by a time series data [24].

2.3 Monitoring Process Control
A control chart is a graphical display of a quality characteristic that can let most people understand the meaning of the chart easily. The control chart for autocorrelated process can be monitored by using the residual $e_i$ obtained from

$$e_i = y_i - \hat{y}_i$$  \hspace{1cm} (7)

where $y_i$ and $\hat{y}_i$ represents the number of observation and predicted values of the response variable for each observation [24]. The two horizontal line in the chart which are called as upper control limit (UCL) and lower control limit (LCL). These control limits will ensure that the process is in control as long as the values of $e_i$ plotted within these two limits, and there is no further action taken. The equation of both control limits is shown below.

$$UCL = \bar{x} + 3\sigma$$  \hspace{1cm} (8)
$$LCL = \bar{x} - 3\sigma$$  \hspace{1cm} (9)

where $\bar{x}$ is average mean of residual and $\sigma$ is the standard deviation of residual.

3.0 Results and Analysis

3.1 Box-Jenkins Method
In order to find a suitable ARIMA model to control the error, it is necessary to identify the behavior of the data. Time series plot is used to evaluate the patterns of data and its behavior, as can be seen in Figure 1.

![Figure 1. Time series plot of furnace temperature.](image)

From Figure 1, a trend presence in this data and it is suspected has non-stationary characteristic based on visual inspection. The $p$-value obtained from Mann-Kendall test is 0.129 which greater than alpha value of 0.05 and it indicated that there is no trend in the series. Meanwhile, Dickey-Fuller test gives a $p$-value of 0.056 which is greater than the alpha value 0.05. Hence, it can be concluded that furnace temperature data is not stationary. Due to its unstable mean, differencing is needed in order to stabilize the data. The time series plot after first differencing is constructed in Figure 2 and visually revealed that the data distributes randomly around zero.
Next, Autocorrelated Function (ACF) plot and Partial Autocorrelated Function (PACF) plot are plotted to determine a suitable ARIMA model. There are a lot of possible models can be selected to monitor and control the process. These models can be referred to Table 3.

| Model       | AIC Value |
|-------------|-----------|
| ARIMA (0,1,1)| 91.1451   |
| ARIMA (0,1,2)| 90.6950   |
| ARIMA (1,1,1)| 92.7206   |
| ARIMA (1,1,2)| 81.6829   |
| ARIMA (2,1,1)| 80.1524   |
| ARIMA (2,1,0)| 89.9274   |
| ARIMA (2,1,2)| 82.1519   |
| ARIMA (3,1,0)| 89.4046   |
| ARIMA (3,1,1)| 82.1519   |
| ARIMA (3,1,2)| 84.1391   |

ARIMA (2,1,1) is the most suitable model to monitor the process since the AIC value is the smallest among all the models which is 80.1524. The fitted model in equation form for ARIMA (2,1,1) is as follows:

\[ \hat{y}_t = 0.0018 + 0.9741 x_{t-1} - 0.3830 x_{t-2} + 0.9663 e_{t-1} + e_t \]  

(10)

where all the values are coefficients, \( x_t \) is predicted value and \( e_t \) is residual for period \( t \). In order to determine whether there is any autocorrelation occurs in the residuals for ARIMA (2,1,1) model, ACF and PACF plots of residuals for furnace temperature are constructed in Figure 3 and Figure 4 respectively.
Both ACF and PACF plots show that the furnace temperature data by using this model is stationary because all the spikes are within the significance limits and these findings can be supported by Ljung-Box test. The result of Ljung-Box test shows that the residuals are independent to each other. It can be concluded that ARIMA (2,1,1) is the preferable model for furnace temperature data.

The accuracy of fitted ARIMA model is tested by using mean absolute percentage error (MAPE) which is 0.0189%. The value of MAPE is considered highly accurate based on Table 2. Hence, the ARIMA (2,1,1) model can be used to monitor and control the process.

3.2 Geometric Brownian Motion

In order to fulfill the two assumptions of GBM; normality test and independent test was conducted. Based on Figure 5, the value of Anderson-Darling Test, $A^2$ computed is 0.2737 which is less than the critical value 0.754 ($\alpha$ =0.05). Hence, random samples follow normal distribution.

In Figure 5, the pattern of the QQ plot indicates that it is normal which satisfy the first assumption of GBM.
The scatter plot in Figure 6 strongly indicates that it has a trend and the coefficient correlation $r$ is 0.7103. This result can be supported by using Durbin-Watson test to test the independent which the value is 1.5015, less than $D_u = 1.662$ and $D_l = 1.611$ at 5% significant level. The value of $D_u$ and $D_l$ can be obtained from [https://www3.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf](https://www3.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf). Hence, it is correlated to each other and dependent. Therefore, log-returns need to be computed which is used to transform the furnace temperature data into independent form.

After computed the log-return, the run chart was constructed as in Figure 7 and shows that all the points approach to zero and there is a stationary pattern. The normality and independency of log-return will be tested by using Q-Q plot and scatter plot respectively.

**Figure 6**: Scatter plot for furnace temperature.

**Figure 7**: Run chart of log-returns for furnace temperature.
Based on Figure 8, Q-Q plot shows that log-return is normal since p-value greater than the critical value 0.05. The presence of autocorrelation in the data is visualized again by using lag-1 scatter plot as presented in Figure 9. This figure clearly indicates that it is independent due to the random pattern of the scatter plot. Meanwhile, Durbin-Watson test is conducted too in order to ensure log-return is independent. Durbin-Watson test is 1.8952 which is greater than $D_L = 1.662$ and $D_U = 1.611$ at 5% significant level. This meant no correlation exists. To find the fitted model, we calculate the estimates parameter of $\hat{c}$ and $\hat{\theta}$ as in Eq (5). Accordingly, then the fitted model is

$$
\hat{\mu}_t = \exp(8.49E - 06)x_{t-1}^{1.056} \left( \frac{x_{t-1}}{x_{t-2}} \right)^{0.156}
$$

(11)

After the fitted model had been formed, the accuracy of the model needs to be measured. The accuracy of fitted GBM model is tested by using MAPE which is 0.0197%. The value of MAPE is considered highly accurate based on Table 2. Hence, the GBM model is suitable to monitor the process.

3.3 Monitoring process of two methods
Fitted model for Box-Jenkins method and GBM method are used to compute residuals as in Eq (7) in order to plot control chart. Table 4 consists of the information to construct the residual chart for ARIMA(2,1,1) and GBM.
Table 4: Control limits of ARIMA and GBM residual chart.

| Control limits     | ARIMA residual chart | GBM residual chart |
|--------------------|----------------------|--------------------|
| Upper Control Limit| $UCL = \bar{x} + 3\sigma$ |
| Where $\bar{x}$ is 0.0353 and $\sigma$ is 0.3731 |
| UCL is 1.1546 | $UCL = \bar{x} + 3\sigma$ |
| Where $\bar{x}$ is -1.4973E-05 and $\sigma$ is 0.4250 |
| LCL is -1.0841 | UCL is 1.2750 |
| Lower Control Limit| $LCL = \bar{x} - 3\sigma$ |
| Where $\bar{x}$ is 0.0353 and $\sigma$ is 0.3731 |
| LCL is -1.0841 | $LCL = \bar{x} - 3\sigma$ |
| Where $\bar{x}$ is -1.4973E-05 and $\sigma$ is 0.4250 |
| LCL is -1.2750 |

When the control limits and center line are computed, the residual chart for ARIMA model will be plotted in Figure 10.

![Figure 10. Control chart of residual for ARIMA(2,1,1) model](image)

Based on Figure 10, there is no signal of out of control because all the point is between the upper and lower limit. The process is in control. Meanwhile for GBM since all the points also within the control limits, then the process is in control as can be seen in Figure 11.

![Figure 11. Control chart of residual for GBM model](image)

The residual control chart for GBM method shows that it is in control and the pattern is same as Box-Jenkins method.
4.0 Conclusion

Based on the results, furnace temperature data is fit into ARIMA (2,1,1) model with MAPE of 0.0189% and the fitted model is shown below.

\[ \hat{x}_t = 0.0018 + 0.9741x_{t-1} - 0.3830x_{t-2} + 0.9663\epsilon_{t-1} + \epsilon_t \]

The furnace temperature data is also fitted into GBM model with MAPE of 0.0197% and the fitted model is also shown below.

\[ \hat{x}_t = \exp(8.49E-06) \cdot x_{t-1} \left( \frac{x_{t-1}}{x_{t-2}} \right)^{0.1596} \]

Both methods are suitable to monitor the autocorrelated process and give a highly accurate model based on MAPE values are below 10%. Hence, both methods can give almost similar better results. This shows that there is no difference in terms of accuracy and monitoring residuals in the process between these two methods. However, in manufacturing industry when people always working with a tremendous task and data that need to be monitored, dealing with a complicated method is unpractical. Although some people prefer to use time series model method, it is still difficult for the person who is lack of forecasting knowledge to apply. During the computational process, it shows that Box-Jenkins method is difficult to apply since there is a lot of possible models can be selected. Besides, Box-Jenkins method is a time-consuming due too many steps are required to be conducted. When it is compared to GBM method, GBM is easier to implement and time saver. This is because it does not need special statistical skills except logarithmic transformation and parameter estimation for regression. Moreover, it only requires short computation time. In conclusion, GBM method is easier to implement, time-saving and maintain high accuracy when compared with the Box-Jenkins method. Thus, a simple mathematical model as GBM is required to assist the practitioners to make a prompt action.

Acknowledgement

This research is supported by Short Term Grant Vot. U369 from Universiti Tun Hussein Onn Malaysia.

References

[1] Pereira D C, Vilas R and Garcia A 2012 Proc. of the Winter Simulation Conf., p 123
[2] Montgomery, D.C. and Mastrangelo, C.M. 1991 Journal of Quality Technology, 23, p179-193.
[3] Russo, S. L., Camargo, M. E., and Fabris, J. P. 2012. Ed. INTECH Open Access Publisher.
[4] Tong, L., Yang, C., Huang, C., and Shou, C. 2005 Proceedings of the Fifth International Conference on Electronic Business, Hongkong, p 692–696.
[5] Psarakis, S., and Papaleonida, G. E. A. 2007. Quality Technology & Quantitative Management, 4(4), p 501–540.
[6] Li, L. S., and Djauhari, M. A. 2013 AIP Conference Proceedings, 1557(1), p 556–560.
[7] Cheng, S. W., and Thaga, K. 2005 Statistica Sinica, 15(2), p 527–546.
[8] Sheu, S., Tai, S. Hsieh, Y., and Lin, T. 2009 Computers & Industrial Engineering, 57(1), p 401–407.
[9] Kandananond, K. 2014 Procedia Engineering, 69, p 1449–145.
[10] Vanhatalo, E., & Kulahci, M. 2015 Quality and Reliability Engineering International, 32(4), p1483-1500.
[11] Mečiarová Z 2007 J. of Info. Control and Management Systems 5, p 73-80.
[12] Armstrong. J.S. 1985 From Crystal Ball to computer 2nd Ed, Wiley-Interscience, New York.
[13] Makridakis, S. A., Andersen, R., Carbone, R.F and Hibon, M. 1982 J. Forecast, 1, p 111-153.
[14] Abidin S N Z and Jaffar M M 2012 World Appl. Sci. J. 17(1), p 87-93.
[15] Lefebvre, M. (2002). Hydrological Process, 16(7), p 1373–1381.
[16] Brewer, K. D., Feng, Y., and Kwan, C. C. Y. 2012 Spreadsheets in Education, 5(3).
[17] Venugopal, S. 2015 International Journal of Research in Engineering and Technology, 4(8), p 2319–2322.
[18] Box, G.E.P and Jenkins, G.M 1976 *Time Series Analysis: Forecasting and Control 1st Ed*, Holden Day, San Francisco.

[19] Box, G.E.P, Jenkins, G.M., and Reinsel, G.C. 2008 *Time Series Analysis: Forecasting and Control 4th Ed*, John Wiley and Sons, Hoboken.

[20] Chen Y, Guan Y, Shao G and Zhang D 2016 *Water* 8, p 77

[21] Stadnytska T 2010 *Europ. J. of Research Methods for the Behavioral and Social Sci.* 6, p 83-92

[22] Hyndman R J and Athanansopoulos G 2014 *OTexts*

[23] Djauhari M A, Li S L and Salleh R M 2016 *American J. of Appl. Sci.* 13, p 860-869

[24] Sagadavan R and Djauhari M.A 2013 *AIP Conf. Proc.* 1557 p 571-575

[25] Mason, R L and Young, J C 2002 *Multivariate Statistical Process Control With Industrial Application*, ASA-SIAM, Philadelphia.