Forecasting of Water Consumptions Expenditure Using Holt-Winter’s and ARIMA

S N A M Razali¹, M S Rusiman¹, N I Zawawi¹, N Arbin²

¹Department of Mathematics and Statistic, Faculty of Applied Science and Technology University Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia.
²Department of Mathematic, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Perak, Malaysia.

Abstract. This study is carried out to forecast water consumption expenditure of Malaysian university specifically at University Tun Hussein Onn Malaysia (UTHM). The proposed Holt-Winter’s and Auto-Regressive Integrated Moving Average (ARIMA) models were applied to forecast the water consumption expenditure in Ringgit Malaysia from year 2006 until year 2014. The two models were compared and performance measurement of the Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD) were used. It is found that ARIMA model showed better results regarding the accuracy of forecast with lower values of MAPE and MAD. Analysis showed that ARIMA (2,1,4) model provided a reasonable forecasting tool for university campus water usage.

1. Introduction

Water is crucial to survival of everything on the earth. Water demand doubles for about every two decades, but the amount of water on earth remains the same [2]. Hence, it is obvious that the amount of water available per capita will decrease in the future. In other words, more people would have to share the same amount of water. If current trends continue, by 2025 the demand for the fresh water is expected to increase by 56 percent more than the amount of water that is currently available. The global supply is not circulated evenly around the planet, and water is not equally accessible at all times throughout the year.

Although Malaysia is blessed with rich of water resources such as abundant of rainfall, however, inefficient management and ill-treatment of water usage have resulted in water crisis that caused hardship. At this moment, Malaysia uses an average of 226 liters of water per person daily, which is higher than other South-East Asia neighbours [3]. Past years, Malaysia had encountered some crucial problem. Drought had occurred in 1977 and 1978, divesting the paddy crop in Northwest Peninsular Malaysia. In 1982 and 1991, drought resulted in drop of the water level of Muda and Pedu dams dropping to critical levels, resulting in cancellation of the off-season crop. In 1998, an El Nino related drought also caused severe water mainly in Kedah and Penang, but caused severe water rationing in Kuala Lumpur and Petaling Jaya for many months. In 2002, drought destroyed thousands of hectares of paddy in Perlis and many areas also suffered water stress [4].

A precise forecasting of water consumption is very essential component in water resources planning and management. Water demand forecasting can guide the decision maker to form the basis for forecasting in the water sector. The prediction of future water demand can help decision-makers to take necessary actions according to the possible crises and restrictions [5]. The daily water demand
forecasting performance of double seasonal univariate time series models can be examined based on multi-step ahead forecast mean squared errors (MSE). It is shown that all the univariate time series models quite useful for short-term forecasting [6]. Holt-Winter method which in ad hoc procedures, have a clear basis in statistical theory and could be considered special education cases of the general class of structural time series models [7]. The optimal combined forecasts of Holt-Winter’s, ARIMA and GARCH models can be quite useful especially for short-term forecasting. However, the forecasting performance of this approach is not consistent over the seven days of week [6].

The univariate time series techniques based on ARIMA models are known to be optimal among linear models for short–term forecasting [8]. The combination of hybrid approach to time series forecasting using ARIMA and Artificial Neural Network models is the useful way to increase forecasting accuracy achieved by either of the two models used separately [9]. Nevertheless ARIMA forecast is suitable if the number of consumers and consumption per consumer contained valuable information [8]. In terms of univariate and multivariate forecast error variance, ARIMA modelling is the better model [10-11]. The accuracy of the water demand forecast model is important, to avoid suboptimal control of the system, and to prevent that operators overrule the control settings (necessary or unnecessary) in order to meet all operational conditions [12, 13]. Forecasting can answer any questions, but the answers are not always right. The method with the smallest percentage error should be selected [14, 15].

This study had compared two forecasting methods, Holt-Winter’s and ARIMA to identify the most suitable forecasting method for water consumption expenditure in UTHM. The consumption of water at UTHM, Batu Pahat campus has been increasing rapidly throughout the years as its population increase year to year. This situation cause several problems such as financial problem for the university management to acquire and optimise their annual budget. Factors that contributed to water shortage problems are the increasing number of students, staffs, new academic buildings and new office buildings [1]. In order to help the university to better plan future water consumption expenditure and fully utilize the budget, this study aims to predict the water consumption for the next 24 periods.

2. Methodology
The monthly data of water bill amount in Ringgit Malaysia for the period from 2006 to 2014 was used. All the reference data were obtained from the Property Administration Office, UTHM. This study applied Holt-Winters and ARIMA methods to forecast the water consumption’s expenditure at UTHM. Minitab statistical software was used to conduct the analyses.

2.1 Holt-Winter’s
Holt-Winter’s model can continue to provide forecast with the same accuracy over time. The Holt-Winter’s has an additive and a multiplicative form. The difference between those forms is in the nature of the seasonal component. The additive model is preferred when the variation of seasonal component is almost stable through the series while the multiplicative is used when seasonal variations changes proportionally to the level of the series. Holt winter’s have three parts which is \( L_t \) is the level at time \( t \), \( T_t \) is the trend at time \( t \), \( S_t \) is the seasonal component at time \( t \). The additive model is represented by the model as in (1) [10];

\[
\hat{Y}_t = L_{t-1} + T_{t-1} + S_{t-p}
\] (1)

while, the multiplicative model is represented as in (2);

\[
\hat{Y}_t = (L_{t-1} + T_{t-1})S_{t-p}
\] (2)

where, the smoothing constants are represented by \( \alpha, \gamma \) and \( \delta \). \( Y_t \) is the value at time \( t \) and \( \hat{Y}_t \) is the fitted value or one-period ahead forecast, at time \( t \).

2.2. ARIMA
ARIMA consist of three parts: (i) AR, autoregressive part; (ii) I, differencing part, and (iii) MA, moving average part. The ARIMA \( (p,d,q) \) model is written in intercept form as in (3) [10]:

\[
\Delta^d y_t = \varphi_0 + \varphi_1 \Delta^d y_{t-1} + \ldots + \varphi_p \Delta^d y_{t-p} + \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}
\]

In the first step, examine the time series plot to see whether the data need to transform, to stabilize variance and use the differencing technique in order to obtain stationary series. When the stationary has been achieved, autocorrelation function (ACF) and partial autocorrelation function (PACF) are being used. The ACF and PACF plots are useful aids in identifying the appropriate ARIMA model series. The second step, after the model has been identified, tests the chosen model in checking ACF and PACF of residuals. If the residuals have no large spikes, model is adequate and ready for the third step. If large spikes remain, consider changing the model. The final step is application of the chosen model in forecasting [10].

3. Result
This section shows the result obtained using Holt-Winter’s and ARIMA techniques.

3.1 Holt-Winter’s

The Holt-Winter’s calibration assigned smoothing parameters of 0.15, 0.20 and 0.25. Different combination of these values had been experimented. The pattern in the data is not obvious since it contains both, constant seasonality pattern and some seasonal fluctuations as the time increase with the series level. Therefore, both methods, multiplicative and additive were applied and the one with smaller accuracy measures were selected. The lowest MAPE and MAD values are multiplicative model with smoothing parameters; \( \alpha = 0.25, \gamma = 0.15, \delta = 0.15 \). The MAPE value is 22 while MAD value is 23486. Figure 1 below shows the Minitab output when \( \alpha = 0.25, \gamma = 0.15, \delta = 0.15 \). The actual values are seemed quite different from fitted. However, it falls within the interval.

![Winters' Method Plot for Actual Multiplicative Method](image)

**Figure 1.** MINITAB Graphical Forecast when \( \alpha = 0.25, \gamma = 0.15, \delta = 0.15 \)
3.2 ARIMA
Firstly, the data characteristics have been emphasized to see whether any transformation is needed. The data trend is clearly shown in Figure 2 below.

![Figure 2. Time Series of Water Consumption’s Expenditure at UTHM.](image)

It is noticed that there is general increasing trend with no clear seasonal pattern. The random fluctuations in the data are roughly increased in size over time. This referred to non-stationary in the variance of the data and must be corrected. For that purpose, natural log transformation has been used; a Box-Cox transformation with zero value of \( \lambda \). Figure 3 below shows the time series transformation of data.

![Figure 3. Time Series of Natural Log Water Consumption’s Expenditure.](image)
It is seen that the value of the fluctuations does not vary with time. Therefore, the transformation has achieved a time series that are stationary in its variances. The next thing to do is to make the mean series stationary. Based on the Figure 3, the series seems non stationary as it increased over time. It requires differencing. The logged and differenced series are shown in Figure 4 below.

**Figure 4.** Time Series Plot of Differenced Natural Log Water Consumption’s Expenditure

As the graph appears to be stationary so it do not require more differencing. Next step is identifying the model of ARIMA by observing ACF and PACF plot. Figure 5 below shows ACF and PACF plot of logged and differenced series.

**Figure 5.** Autocorrelation Function and Partial Autocorrelation Function of Differenced Natural Log Water Consumption’s Expenditure

ACF and PACF shows exhibiting large spikes that gradually die out. Based on Figure 5, the model was probably (2,1,2) or (2,1,4). However, after examine the significant of parameters and the Box
Pierce (Ljung-Box) Chi Square statistics, the best model is ARIMA (2,1,4). After the models are estimated, their adequacy is needed to be checked. One way to check their adequacy is by verifying that there is no large spikes in the ACF of residuals.

Since there is no spike outside the insignificant zone, the fit are satisfied. Hence, the ARIMA (2,1,4) model is working fine. Forecasting with ARIMA(2,1,4) model has been used. The notation for ARIMA(2,1,4) model is:

\[ \Delta^1 y_t = \phi_0 + \phi_1 \Delta^1 y_{t-1} + \phi_2 \Delta^1 y_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} \]

| MODEL          | MAPE | MAD   |
|----------------|------|-------|
| Holt-Winter’s  | 21.7830 | 23485.65 |
| ARIMA          | 19.7085 | 22694.42 |

Based on Table 1, the ARIMA model obtained better results and got a better performance comparing Holt-Winter’s model, according to MAPE and MAD. It has been chosen to forecast water expenditure at UTHM.

By using ARIMA(2,1,4) model, Table 2 illustrates monthly forecasting value with 95% confidence limits for water consumption’s expenditure in UTHM for 24 future values ahead.

| PERIOD       | FORECAST | LOWER | UPPER  |
|--------------|----------|-------|--------|
| January 2015 | 181652   | 118181| 245123 |
| February     | 215370   | 141637| 289103 |
| March        | 190198   | 112727| 267668 |
| April        | 190245   | 109348| 271143 |
4. Conclusion
Holt-Winter’s and ARIMA models were applied to demand forecasting of the water consumption’s expenditure. Then the models were compared using accuracy measurements of MAPE and MAD. The Holt-Winter’s was used when the data show both seasonality and trend. The two main Holt-Winter’s models are additive model for time series exhibiting additive seasonality and multiplicative model for time series exhibiting multiplicative seasonality. The Holt-Winter’s model is simple and can gives accurate forecasting results as those obtained with more complex and try and error techniques. This method is popular and easy to use in practical application.

The ARIMA forecasting model has many assumptions that usually cannot be fulfilled. However, if the model assumptions are true, an excellent forecast is obtained. Despite the constraint, many evidences show that ARIMA methodology is better in terms of accurateness. Therefore, this method is studied to give support for every assumptions. Overall, this paper concludes that ARIMA model is better than Holt-Winter’s model as the MAPE and MAD values is lower. ARIMA(2,1,4) model is the most suitable forecasting model for water consumption’s expenditure at UTHM.

Acknowledgments
Authors would like to thank Universiti Tun Hussein Onn Malaysia (UTHM) and Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), UTHM for kindly providing us with the internal funding.
References

[1] Katimon A and Demun A S 2004 Water use at Universiti Teknologi Malaysia: Application of ARIMA model Journal Teknologi 41 47-56.

[2] Chan N W 2004 A Critical Review of Malaysia’s Accomplishment on water resources management under AGENDA 21 Malaysia Journal of Environment Management 5 55-78.

[3] Choong M Y 2011 Malaysia faces looming water crisis. The Star Online http://www.thestar.com.my/story/

[4] Chan N W 2009 Issues and challenges in water governance in Malaysia Iran Journal of Environmental Health Science and Engineer UCI 9 143-152.

[5] Behboudian S, Tabesh M, Falahnezhad M and Ghavanini F A 2014 A long-term prediction of domestic water demand using preprocessing in artificial neural network Journal of Water Supply: Research and Technology-AQUA 63 31-42.

[6] Caiado J 2007 Forecasting water consumption in Spain using univariate time series models MPR A 6610 1-9.

[7] Lepojevic V and Andelkovic-Pesic M 2011 Forecasting electricity consumption by using Holt-Winters and seasonal regression models Series: Economics and Organization 8 421 – 431.

[8] Guerrero V M and Berumen E 1998 Forecasting electricity consumption with extra model information provided by consumers Journal of Applied Statistics 25 283 – 299.

[9] Zhang G P 2003 Time series forecasting using a hybrid ARIMA and neural network model Neurocomputing 50 159-175.

[10] Preez J D and Witt S F 2003 Univariate versus multivariate time series forecasting: An application to international tourism demand International Journal of Forecasting 19 435-451.

[11] Rusiman M S, Hau O C, Abdullah A W, Sufahani S F, Azmi N A 2017 An Analysis of Time Series for the Prediction of Barramundi (Ikan Siakap) Price in Malaysia Far East Journal of Mathematical Sciences 102(9) 2081-2093.

[12] Bakker M, Duist H V, Schagen K V, Vreeburg J, Rietveld L, 2014 Improving the performance of water demand forecasting models by using weather input Procedia Engineering 70 93-102.

[13] Arbin N, Sabri S R M, Idris M S and Razali S N A M 2016 Development of Algorithms and Serration Model to add details to the Leaf Margin of the Leaf Shape Modeling by using Offset Approach Far East Journal of Mathematical Sciences 100(12) 2081-2088.

[14] Rusiman M S, Nasibov E and Adnan R 2011 The Optimal Fuzzy C-regression Models (OFCRM) in Miles per Gallon of Cars Prediction, Proceedings – 2011 IEEE Student Conference on Research and Development, SCOR eD 2011, 6148760 333-338.

[15] Shafi M A and Rusiman M S 2015 The Use of Fuzzy Linear Regression Models for Tumor Size in Colorectal Cancer in Hospital of Malaysia Applied Mathematical Sciences 9 (56) 2749-2759.