

EXPONENTIAL SMOOTHING TECHNIQUES ON TIME SERIES RIVER WATER LEVEL DATA

Noor Shahifah Muhamad¹ and Aniza Mohamed Din²

¹Universiti Utara Malaysia (UUM), Malaysia, merah_sha@yahoo.com

²Universiti Utara Malaysia (UUM), Malaysia, anizamd@uum.edu.my

ABSTRACT. The increasing of river water level usually happens during raining season. This event can lead to devastating flash flood, which would eventually cause damage to properties and possibly, loss of human life. Such event is also known as extreme event due to the nature of the data produced, which mostly consist of nonlinear pattern of data. The existence of nonlinear pattern and noise data greatly affect the quality of prediction result. Three exponential smoothing techniques have been investigated to study their ability in handling extreme river water level time series data, which are Single Exponential Smoothing Technique, Double Exponential Smoothing Technique and Holt's Method. The techniques were performed on river water level data from three rivers in Perlis, Malaysia. From the experiments, it was found that all the three techniques have their own limitations in handling extreme data, with Double Exponential Smoothing Technique to perform better than its counterpart.

Keywords: extreme event, extreme data, exponential smoothing technique, holt's method

INTRODUCTION

Extreme event can be described as occurrences that occur either naturally or due to human factor, rarely happen, but significant in terms of its impacts, effects or outcomes (Singh, 2012; Sarewitz & Pielke, 2001; Easterling et al., 2000). The characteristics of extreme event are volatile, rapidly change and can have unpredictable outcomes (Turrof et al., 2011). An example of extreme events is heavy rainfall that occurs at the time that is not supposed to happen and not in raining season. Extreme rainfall is entirely different from the normal weather day (Nayak & Gosh, 2013). This event may lead to rivers and reservoir water level to increase drastically. This incident can lead to devastating flash flood, cause damage to properties and possibly, loss of human life.

Extreme event time series is difficult to study and even harder to be used for prediction because of its rare characteristics (Ghil et al., 2011). This is due to the data generated from this event is rarely linear (Mishra et al., 2007). The extreme data are usually in nonlinear pattern, tremendous noise and complex dimensionality in signal. The presence of extremely high or low values in data, while not all of them are outliers, can affect the accuracy of prediction (Koehler et al., 2012). Hence, the need to identify a good data processing technique that could handle this type of data is necessary.

Exponential Smoothing Technique (EST) is a forecasting technique which has been applied in many studies. It has also been used to isolates trends and seasonality from irregular

variation (Chan et al., 2011). The technique eliminates the high fluctuations in signal while maintaining the important patterns of data. This criteria and the fact that the technique is easy to understand and implement make it an attractive technique to be investigated in handling extreme data. This paper investigated three smoothing techniques applied on extreme data and identifies the technique that could best deal with this type of data.

The rest of the paper is organized as follows. Section 2 gives an overview on EST, followed by description of the case study and the type of data used in Section 3. The experimental result and findings are presented in Section 4 and with the conclusion in Section 5.

OVERVIEW ON EXPONENTIAL SMOOTHING TECHNIQUES

In this paper, three smoothing techniques which are Single Exponential Smoothing Technique (SEST), Double Exponential Smoothing Technique (DEST) and Holt's method are discussed.

Single Exponential Smoothing Technique

The SEST is also known as simple exponential smoothing (Chan et al., 2011). This technique is a popular statistical technique used for time series forecasting and also can be used in data preprocessing (Zheng & Xu, 2008; Chan et al., 2011). SEST provides an advantage of simplicity, less costly to develop and easier to understand (Hussain & Jamel, 2013). There is only one smoothing parameter to be determined in SEST. The general equation for SEST is as follows:

$$f_{t+m} = \alpha y_t + (1 - \alpha)f_t \quad (1)$$

where f_{t+m} is the single exponential smoothed value in period $t+m$, for $m = 1, 2, 3, 4, \dots$, α is the unknown smoothing parameter to be determined with underlying between 0 and 1, y_t is the actual value in time period t , and f_t is the smoothed value for period t .

Double Exponential Smoothing Technique

The DEST is also known as Brown's method (Lazim, 2012; Abdullah et al., 2012). It is useful for time series that exhibits linear trend characteristic. The main advantage of DEST is its ability to generate multiple-ahead-forecast but difficult to determine the value of α (Hussain & Jamel, 2013). The formulation can be shown as follows:

Let, S_t be the exponentially smoothed value of y_t at time t

S'_t be the double exponentially smoothed value of y_t at time t

There are four main equations involved:

- 1) The exponentially smoothed series value

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1}) \quad (2)$$

- 2) The double exponentially smoothed series value

$$S'_t = \alpha S_t + (1 - \alpha)(S'_{t-1}) \quad (3)$$

- 3) The difference exponentially smoothed series value trend estimate

$$\alpha_t = 2(S_t - S'_t) \quad (4)$$

$$\alpha_t = 2(S_t - S'_t) \quad (5)$$

- 4) Forecast m period into the future

$$f_{t+m} = \alpha_t + b_t * m \quad (6)$$

Holt's Method

Holt's method is also known as 'Holt-Winters double exponential smoothing technique', originally presented by Charles C. Holt and Peter Winters (Gelper et al., 2007). It uses two smoothing parameter and is good in handling trends. This method is much like SEST excludes the two components which are level and trend estimate that must be updated on each period. The level is a smoothed estimate of the value of the data at the end of each period while trend is smoothed estimate of average growth at the end of each period (Kalekar, 2004). Holt's method requires three equations:

- 1) The level estimate

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (7)$$

- 2) The trend estimate

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (8)$$

- 3) Forecast m period into the future

$$f_{t+m} = \alpha_t + b_t * m \quad (9)$$

where y_t is the actual value for period t , α and β is the value of the smoothing parameter, and T_{t-1} is the prediction value for period $t+1$ while T_t is the forecast value for period t .

CASE STUDY

The data used for experiment and testing were the historical data of river water level from year 2001 to 2013 which were acquired from the Department of Irrigation and Drainage of Perlis. These secondary data were collected from three rivers which were Sungai Arau, Sungai Korok and Sungai Jarum. The total number of observation selected for experiment and testing is 4386 series for Sungai Arau, 906 series for Sungai Korok and 912 series for Sungai Jarum.

EXPERIMENT AND RESULT

The evaluation on techniques performance has been done based on error measurement obtained by using performance metrics. The performance metrics is a normal criterion which has been used to evaluate the technique performance based on error measurement (Lazim, 2012). These measurements are based on forecast errors, or the differences between the actual and forecast errors. The measure is called Mean Squared Error (MSE) and Root Mean Squared error (RMSE).

- 1) Mean Squared Error is an estimator to measure the average of the squares of the error.

$$MSE = \frac{\sum_t^n e_t^2}{n}$$

- 2) Root Mean Squared Error is used to measures the differences between fitted value and actual value.

$$RMSE = \sqrt{\frac{\sum_t^n e_t^2}{n}}$$

In the experiment, determining the value for the smoothing parameter (denoted by α) is based on trial and error. The value of α , which begins with 0.1, 0.2 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 have been tested in order to find the best value. The best value of α is defined as the

one that gives the smallest error. A small α is suitable to be used for stable time series data while a large α is to deal with a series that change rapidly (Lazim, 2012).

Table 1. Result Comparison using Sungai Arau Dataset

Techniques	SSQ	MSE	RMSE
DEST	41.93	0.009559	0.097771
SEST	265.13	0.060450	0.245866
Holt's	285.91	0.065188	0.255319

Table 2. Result Comparison using Sungai Korok Dataset

Techniques	SSQ	MSE	RMSE
DEST	1.22	0.001343	0.036650
SEST	3.12	0.003448	0.058723
Holt's	3.15	0.003476	0.058958

Table 3. Result Comparison using Sungai Jarum Dataset

Techniques	SSQ	MSE	RMSE
DEST	22.27	0.024415	0.156253
SEST	107.46	0.117826	0.343258
Holt's	113.71	0.124683	0.353104

Tables 1, 2, and 3 shows the generated result obtained from the DEST, SEST and Holt's method applied on the data from each river, respectively. As shown, the best technique to deal with Sungai Arau, Sungai Korok, and Sungai Jarum water level data is DEST. The DEST produced the lowest MSE values which are 0.009559, 0.001343 and 0.024415 for the three rivers respectively. For the three river water level data, the DEST with $\alpha = 0.5$ shows a better performance compared to SEST with $\alpha = 0.9$ and Holt's method with $\alpha = 0.9$ and $\beta = 0.1$. From the result, it was found that SEST and Holt's method had some limitations to deal with the pattern of data from the three rivers.

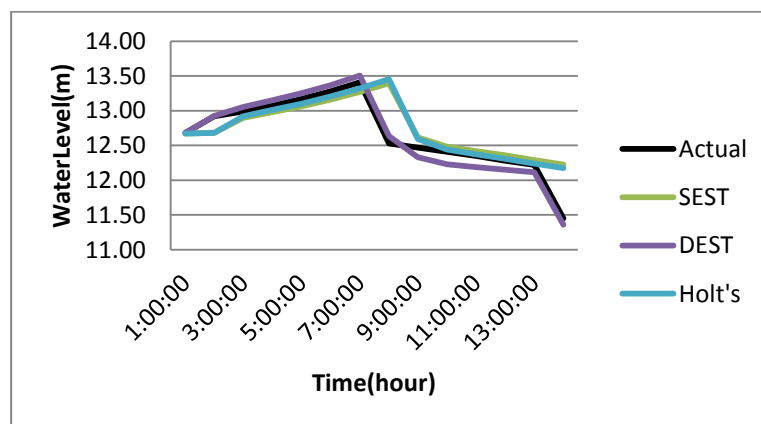


Figure 1. Time Series Plot for Sungai Arau

Figures 1, 2, and 3 illustrated the time series plot between the actual data and smoothed data using DEST, SEST and Holt's method for three rivers. The y-axis is for time and the x-axis is the water level data which is measured in meters. Based on the three figures, the forecasted data using DEST was found to be mostly lay down to the actual data. This result has shown that DEST performed better compared to SEST and Holt's method. However, Figure 2 shows the performances of the SEST and Holt's method are equal with DEST, which means they are also closer to the actual data. This is because the SEST and Holt's method can still capture the pattern of Sungai Korok water level data, where the data is normal without high fluctuations.

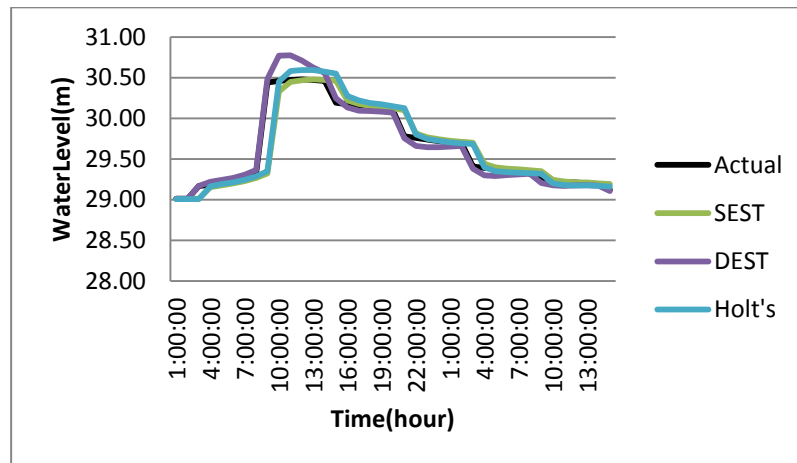


Figure 2. Time Series Plot for Sungai Korok

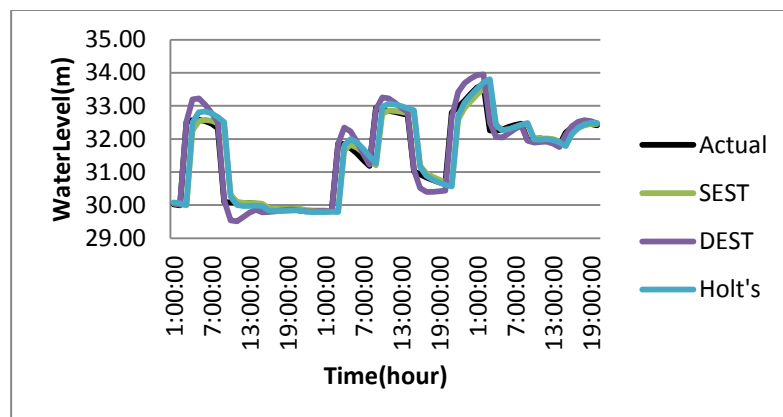


Figure 3. Time Series Plot for Sungai Jarum

CONCLUSION

Three exponential smoothing techniques were investigated to find the best technique to deal with the pattern of river water level data from three rivers located in Perlis. Based on the error generated from the experiment, DEST was found to be the most suitable smoothing technique as it produced the lowest error compared to SEST and Holt's method. This technique could handle the data from all the three rivers. SEST was found to be suitable for the analysis of time series with smooth fluctuations and produced a higher error if dealt with big-

ger fluctuations. From the experiment, it was found that Holt's method was able to deal with trend variations, but this technique is sensitive with data that change rapidly. By finding the best smoothing technique for extreme data, more accurate prediction can be produced. An accurate prediction is likely to be able to help the authority and the public in reducing the impact of flood disasters, and to act as an early warning system to inform the public about upcoming events.

ACKNOWLEDGEMENT

This project is under Long Term Grant Scheme (LRGS/b-u/2012/UUM/Teknologi Komunikasi dan Informasi). The authors would like to thank the Ministry of Higher Education Malaysia for supporting this research. The author would also like to thank the Department of Irrigation and Drainage Malaysia for supplying hydrology and river water level data for this study.

REFERENCES

- Abdullah, S., Sapii, N., Dir, S., & Jalal, T. M. T. (2012). Application of univariate forecasting models of tuberculosis cases in Kelantan. In *2012 International Conference on Statistics in Science, Business, and Engineering (ICSSBE)*, 1–7. doi:10.1109/ICSSBE.2012.6396582
- Chan, K. Y., Dillon, T. S., Singh, J., & Chang, E. (2011). Traffic flow forecasting neural networks based on exponential smoothing method. In *2011 6th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 376–381. doi:10.1109/ICIEA.2011.5975612
- Gelper, S., Fried, R., & Croux, C. (2010). Robust forecasting with exponential and Holt–Winters smoothing. *Journal of Forecasting*, 29(3), 285–300. doi:10.1002/for.1125
- Ghil, M., Yiou, P., Hallegatte, S., Malamud, B. D., Naveau, P., Soloviev, A., ... Zaliapin, I. (2011). Extreme events: dynamics, statistics and prediction. *Nonlin. Processes Geophys.*, 18(3), 295–350. doi:10.5194/npg-18-295-2011
- Hussain, M. M. F. & Jamel, R. A. (2013). Statistical Analysis & Asthmatic Patients in Sulaimaniyah Governorate in the Tuber-Closes Center. In *International Journal of Research in Medical and Health Sciences*. 1(2). Retrieved from http://www.ijsk.org/uploads/3/1/1/7/3117743/ashtehmatic_patients_4.pdf
- Kalekar, P. S. (2004). Time series forecasting using Holt-Winters Exponential Smoothing Under the guidance of Bernard, P., (04329008), 1–13. Retrieved from http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf
- Lazim, M. A. (2012). *Introductory Business Forecasting a Practical Approach (3rd ed.)*. Shah Alam, Selangor: UiTM Press.
- Nayak, M. A., & Ghosh, S. (2013). Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theoretical and Applied Climatology*. doi:10.1007/s00704-013-0867-3
- Wong, W. K., & Guo, Z. X. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, 128(2), 614–624. doi:10.1016/j.ijpe.2010.07.008
- Zheng, Y., & Xu, R. (2008). An Adaptive Exponential Smoothing Approach for Software Reliability Prediction. In *4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08*, 1–4. doi:10.1109/WiCom.2008.2946