Proceedings of the 6th International Conference on Computing and Informatics, ICOCI 2017
25-27April, 2017 Kuala Lumpur. Universiti Utara Malaysia (http://www.uum.edu.my)
Paper No. 217

How to cite this paper:

Noor Shahifah Muhamad & Aniza Mohamed Din. (2017). Exponential smoothing techniques on daily temperature level data in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference of Computing & Informatics (pp 62-68). Sintok: School of Computing.

EXPONENTIAL SMOOTHING TECHNIQUES ON DAILY TEMPERATURE LEVEL DATA

Noor Shahifah Muhamad[1] and Aniza Mohamed Din[2]

[1] *Universiti Utara Malaysia, Malaysia, merah_sha@yahoo.com*
[2]*Universiti Utara Malaysia, Malaysia,* anizamd@uum.edu.my

ABSTRACT. The changes of temperature level occur throughout the year. This event whether hot temperature or cold temperature can affect human life and nature. Such event is also known as extreme event due to the nature of the data produced. Usually the time series of extreme dataset is rarely linear. The existence of nonlinear pattern and high fluctuation in variation greatly affect the quality of forecasting performances. Three exponential smoothing techniques have been tested to study their ability in handling of temperature level data from three cities in Texas. Single Exponential Smoothing Technique (SEST), Double Exponential Smoothing Technique (DEST) and Holt's method were explored in preparing the temperature data. From the experiments, it was found that DEST is the most suitable technique to deal with the data compared to SEST and Holt's method.

Keywords: temperature level, extreme event, extreme data, exponential smoothing technique

INTRODUCTION

The changes of temperature level occur throughout the year. Both, very hot or very cold temperatures could be dangerous to our health. Hot temperatures can produce several health effects to human and the most common is dehydration. Extreme hot conditions can lead to heat exhaustion, heat stroke and heat cramps. The most serious concern is heat stroke which can lead to death. In cold temperatures, the risk of hypothermia or dangerous overcooling of the body is the most serious problem. Hypothermia can lead to death if there is no immediate medical attention. Another serious effect of cold exposure is frostbite or freezing of the exposed extremities such as fingers, toes, nose and ear lobes. This event can also affect the nature and agricultures. It can lead to damage to the plant and agriculture production. Extreme temperature event is one of dangerous extreme events which include volcano eruptions (Lara et al., 2015), earthquakes (Zambrano-Vizuete, Perez-Llopis, Palau, & Esteve-Domingo, 2015), hurricanes (Fthenakis, 2013), and typhoons (Pang & Li, 2013). Extreme events are volatile, rapidly changing and can have unpredictable outcomes.

Physically, extreme events can also be described through the effect of the event to community. Extreme weather events are expected to become more frequent due to climate changes. The occurrence of extreme events is known as *"disaster"*. This event can be represented by extreme dataset values in the verification measurement and usually the time series of extreme events dataset is rarely linear (Mishra et al., 2007).

Extreme data exist when there are data with nonlinear pattern, tremendous noise variations and complex dimensionality in signal. This produces high fluctuation in verification measurement between the dataset and large ambiguity with the quality of forecasting performances. However, it depends on whether they are extremely high or low. Therefore, smoothing techniques should be applied to overcome the extreme data handling in order to get an accurate result in making prediction.

Exponential Smoothing Technique (EST) is a forecasting technique which has been applied in many works. It has also been used to isolate 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 investigates three smoothing techniques applied on extreme data and identifies the technique that could best deal with this type of data.

EXPONENTIAL SMOOTHING TECHNIQUES

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

This technique is a popular statistical technique used for time series forecasting and in data pre-processing (Chan et al., 2011). It provides an advantage of simplicity, less costly to develop and easier to understand (Hussain & Jamel, 2013). In SEST, there is only one smoothing parameter to be determined. The general equation for SEST is as follows:

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

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 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.

Eq. (1) is the basis and simplest technique in the group of exponential smoothing. This technique is a univariate time series method and suitable for short range forecasting without time series trend and seasonality component. SEST is a technique that uses weighted moving average of past data as the basis of forecasting. The greater weights are given to the most recent data, while lesser weights are given to the past data. The accuracy of the SEST strongly depended on the optimal value of the smoothing parameter (Sahu & Kumar, 2013).

Many works showed that SEST has been used widely in forecasting model development such as wind energy predictions (Olaofe, 2015), landslide forecasting (Zhang et al., 2013), bus passenger traffic prediction (Ge et al., 2013), exchange rate forecasting model (Akincilar et al., 2011) and traffic flow forecasting model (Chan et al., 2011). The SEST was also used in data preprocessing to remove irregularity in raw data, for example, to process the traffic flow data before they could be trained using neural networks (Chan et al., 2011). To improve the accuracy in the prediction of river water level pollution, SEST was applied to deal with the original data (Zheng-wen & Kai-yu, 2010). Generally, SEST is a successful technique which can be used in many fields of study. The target of the study is to minimize the error measurement, reduce randomness, and to find the best method which is appropriate with the sample.

Even though SEST is the most commonly used in practice, this technique has its weaknesses when the time series show the nonlinear patterns especially with the presence of outliers (Croux et al., 2010; Gelper et al., 2007). According to Monfared et al. (2014), outliers cannot be detected when SEST is used. This is because SEST is a linear method which cannot handle nonlinearity and irregularity in raw data (Lai et al., 2006). Mostly, SEST is suitable for short term prediction (Ge et al., 2013) and extensively used for time series not exhibiting trend and seasonality (Akincilar et al., 2011).

Double Exponential Smoothing Technique

DEST, also known as Brown's method, 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:

The exponentially smoothed series value

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

The double exponentially smoothed series value

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

The difference exponentially smoothed series value trend estimate

$$\alpha_t = 2(S_t - S'_t) \tag{4}$$

$$b_t = \frac{\alpha}{1-\alpha} 2(S_t - S'_t) \tag{5}$$

Forecast m period into the future

$$f_{t+m} = \alpha_t + b_t * m \tag{6}$$

In previous works, it was found that the DEST is the appropriate technique to predict the number of male patients infected with asthma disease (Hussain & Jamel, 2013). This technique has also been selected to forecast the tuberculosis cases in Kelantan after the evaluation and comparison with other univariate forecasting models (Abdullah et al., 2012). The main difficulty when using this technique is to determine the value of smoothing parameter (Hussain & Jamel, 2013). The DEST can only handle time series data that exhibits trend, but sensitive with seasonality data.

Holt's Method

Holt's method, also known as 'Holt-Winters double exponential smoothing technique', was introduced 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 similar to SEST excluding two components which are level and trend estimates that must be updated on each period. Level estimate is a smoothed estimate of the value of the data at the end of each period while trend is a smoothed estimate of average growth at the end of each period (Kalekar, 2004). Holt's method requires three equations:

The level estimate

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

The trend estimate

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

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 are the values of the smoothing parameters, and T_{t-1} is the prediction value for period $t+1$while T_t is the forecast value for period t.

The ability of this technique had inspired Tikunov and Nishimura (2007) to use this method to evaluate future radio traffic of circuit switched services for near-term outlook and the result proved that this technique could be used for traffic prediction. Holt's method had also been tested in previous works such as tourists forecasting in Kenya (Akuno et al., 2015), forecasting of particleboard consumption (Tavakkoli et al., 2015) and univariate forecasting model of tuberculosis cases (Abdullah et al., 2012). These works share a common target which is to deal with the time series data that exhibit trend component and to find the best model fit with the data sample. Several works found that Holt's method gave a better performance since it produced the lowest error.

Though the Holt's method is good in handling time series trend forecasting data, it has difficulties in determining the best value for the smoothing parameter, since it requires two smoothing parameters (Hussain & Jamel, 2013) and only able to deal with trend variation.

METHODOLOGY

The generated result from SEST, DEST and Holt's method were compared based on forecast errors, or the differences between the actual and forecast errors, namely Sum of Squared Error (SSQ), Mean Squared Error (MSE) and Root Mean Squared error (RMSE).

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).

Benchmark data of temperature level for three cities in Texas, United States were taken from the University of Dayton academic site. The source of this data is from the National Climatic Data Center (NCDC). This daily data of temperature level is measured in Fahrenheit (F'). Table 1 shows a sample of data for the three cities' temperature level.

Table 1. Sample Data for Temperature Level (in Fahrenheit)

Time (day) \ City	1/1/1995	1/2/1995	1/3/1995	1/4/1995	1/5/1995	1/6/1995
Brownsville	58.90	49.50	46.60	45.50	42.60	52.10
Corpus Christi	53.10	44.60	44.30	44.60	42.00	49.80
San Antonio	48.50	39.50	41.90	43.20	36.80	48.60

Figure 1 presents the temperature graph of the actual data for the three cities in Texas. Based on the figure, the characteristics and behaviors of the data pattern for the three cities are almost the same. It also shows that there exists seasonal variation at every beginning and end of the year.

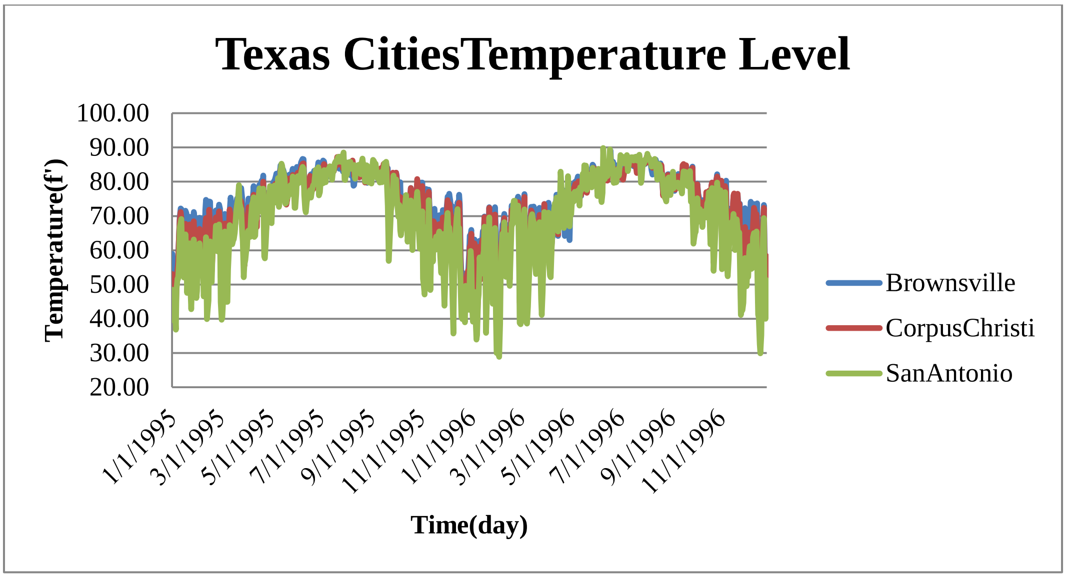

Figure 1. Temperature Level of Texas Cities

FINDINGS

Tables 2, 3, and 4 show the generated results obtained from the SEST, DEST and Holt's method applied on temperature level data from each city, respectively. It was found that the best technique to deal with Brownsville, Corpus Christi, and San Antonio temperature level data is DEST. The DEST produced the lowest error of MSE values which are 3.247008, 3.568950 and 4.338322 for the three temperatures level respectively. The DEST with α = 0.5 shows a better performance compared to SEST with α = 0.9 and Holt's method with α = 0.9 and β = 0.1 for the three temperatures level data. From the results, it was found that SEST and Holt's method had some limitations to deal with the pattern of temperature data from the three cities.

Table 2. Result Comparison using Brownsville Temperature Dataset

Techniques	SSQ	MSE	RMSE
SEST	18888.99	26.017894	5.100774
DEST	2357.33	**3.247008**	1.801946
Holt's Method	20708.18	28.523660	5.340755

Table 3. Result Comparison using Corpus Christi Temperature Dataset

Techniques	SSQ	MSE	RMSE
SEST	21269.92	29.297409	5.412708
DEST	2591.06	**3.568950**	1.889166
Holt's Method	23315.96	32.115643	5.667067

Table 4. Result Comparison using San Antonio Temperature Dataset

Techniques	SSQ	MSE	RMSE
SEST	25639.78	35.316507	5.942769

DEST	3149.62	**4.338322**	2.082864
Holt's Method	28079.92	38.677571	6.219129

These results have shown that DEST performed better compared to SEST and Holt's method. SEST and Holt's method did not perform well due to the limitation in handling seasonal data.

CONCLUSION

This study is conducted to find the most suitable smoothing technique which to be used in handling temperature level data that consist of seasonal variation. Three exponential smoothing techniques were tested to find the best technique which can handle the pattern of temperature level data from three cities located in Texas. Based on the result, it was found that DEST to be the best technique for handling the temperature level data where it produces the lowest value of MSE and RMSE compared to SEST and Holt's method. SEST was found to be suitable for the analysis of time series which is not exhibiting trend and seasonality while, Holt's method was able to deal with trend variations. These two techniques are sensitive with data that change rapidly and random variation.

ACKNOWLEDGMENTS

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 is 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)* (pp. 1–7). doi:10.1109/ICSSBE.2012.6396582

Akincilar, A., TEMİZ, İ., & Şahin, E. (2011). An application of exchange rate forecasting in Turkey. *Gazi University Journal of Science*, *24*(4), 817-828.

Akuno, A. O., Otieno, M. O., Mwangi, C. W., & Bichanga, L. A. (2015). Statistical Models for Forecasting Tourists' Arrival in Kenya. *Open Journal of Statistics*, *5*(01), 60

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)* (pp. 376–381). Presented at the 2011 6th IEEE Conference on Industrial Electronics and Applications (ICIEA). doi:10.1109/ICIEA.2011.5975612

Croux, C., Gelper, S., & Mahieu, K. (2010). Robust exponential smoothing of multivariate time series. Comput. Stat. *Data Anal*, 54(12), 2999–3006. doi:10.1016/j.csda.2009.05.003

Fthenakis, V. (2013). The resilience of PV during natural disasters: The hurricane Sandy case. In *Photovoltaic Specialists Conference (PVSC), 2013 IEEE 39th* (pp. 2364–2367). http://doi.org/10.1109/PVSC.2013.6744949

Ge, S. Y., Zheng, C. J., & Hou, M. M. (2013). Forecast of Bus Passenger Traffic Based on Exponential Smoothing and Trend Moving Average Method. *Applied Mechanics and Materials*, *433-435*, 1374–1378. doi:10.4028/www.scientific.net/AMM.433-435.1374

Gelper, S., Fried, R., & Croux, C. (2007). Faculty of Economics and Applied Economics Robust forecasting with exponential and Holt-Winters smoothing Robust Forecasting with Exponential and Holt-Winters Smoothing.

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 Meadical and Health Sciences.* Vol. 1, No. 2.

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

Lai, K. K., Yu, L., Wang, S., & Huang, W. (2006). Hybridizing exponential smoothing and neural network for financial time series predication. In *Lecture Notes in Computer Science 3994* (pp. 493–500).

Lara, R., Benitez, D., Caamano, A., Zennaro, ., & Rojo-Alvarez, J. (2015). On Real-Time Performance Evaluation of Volcano Monitoring Systems with Wireless Sensor Networks. *IEEE Sensors Journal*, *PP*(99), 1–1. http://doi.org/10.1109/JSEN.2015.2393713

Lazim, M. A. (2012). *Introductory Business Forecasting a Practical Approach* (3rd ed.). Shah Alam, Selangor: UiTM Press.

Mishra, A. K., Desai, V. R., Singh, V. P., & Asce, F. (2007,December), Drought Forecasting Using a Hybrid Stochastic and Neural, 626–638.

Monfared, M., Rastegar, H., & Kojabadi, H. M. (2009). A new strategy for wind speed forecasting using artificial intelligent methods. *Renewable Energy*, *34*(3), 845–848. doi:10.1016/j.renene.2008.04.017

Olaofe, Z. O. (2015). Wind Energy Predictions of Small-Scale Turbine Output Using Exponential Smoothing and Feed-Forward Neural Network. *International Journal*, *5*(2), 28-42.

Pang, L. & Li, Y. L. (2013). The Application of Grey Model to the Prediction of Extreme Sea States Induced by Typhoon. In *International Conference on Remote Sensing, Environment and Transportation Engineering.* doi:10.2991/rsete.2013.136

Sahu, P., & Kumar, R. (2013). Survey of Demand fluctuation of Dairy products at raipur Dugdh Sangh, Chhattisgarh, India for selection of appropriate forecasting. *Uncertain Supply Chain Management*, *1*(3), 133–144.

Tavakkoli, A., Hemmasi, A. H., Talaeipour, M., Bazyar, B., & Tajdini, A. (2015). Forecasting of Particleboard Consumption in Iran Using Univariate Time Series Models. *BioResources*, *10*(2), 2032-2043.

Tikunov, D., & Nishimura, T. (2007). Traffic prediction for mobile network using Holt-Winter #x2019;s exponential smoothing. In *15th International Conference on Software, Telecommunications and Computer Networks, 2007. SoftCOM 2007* (pp. 1–5). doi:10.1109/SOFTCOM.2007.4446113

Zambrano Vizuete, A. M., Perez Llopis, I., Palau, C., & Esteve Domingo, M. (2015). Distributed Sensor System for Earthquake Early Warning Based on the Massive Use of Low Cost Accelerometers. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, *13*(1), 291–298. http://doi.org/10.1109/TLA.2015.7040661

Zhang, W. Z., Li, Y. J., Li, X. S., & Kang, Q. R. (2013). Forecasting of Landslide Displacement Based on Exponential Smoothing and Nonlinear Regression Analysis. *Applied Mechanics and Materials*, *353-356*, 1101–1105. http://doi.org/10.4028/www.scientific.net/AMM.353-356.1101

Zheng-wen, X., & Kai-yu, S. (2010). Improved Grey Model Base on Exponential Smoothing for River Water Pollution Prediction. In *2010 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)* (pp. 1–4). Presented at the 2010 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE). doi:10.1109/ICBBE.2010.5517830