RTCEBE

River Water Level Time-Series Forecasting by Using Smoothing Technique

Ahmad Mursyid Hamidon[1], Sabariah Musa[2*]

[1,2]Faculty of Civil Engineering and Built Environment,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, 86400, MALAYSIA

*Corresponding Author Designation

Abstract: The increasing of river water level usually happens during raining season and can lead to devastating flash floods. Therefore, forecasting river water level series using the exponential smoothing method was applied to predict accurate river water level series. 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 Pahang, Malaysia which is Sungai Jelai in Jeram Bungor which is case study 1, Sungai Tembeling which is case study 2 and Sungai Temerloh which is case study 3. Monthly data of Sungai Pahang water level was obtained from JPS Malaysia from January 2010 to February 2021. The IBM SPSS software was used to analyse this data. This method of forecasting is evaluated to determine the ability in the forecasting river water level for short-term forecast with seasonal and non-seasonal data. Based on the error generated from the analysis, Simple exponential smoothing technique from case study 1 was found to be the best model smoothing technique as it produced the lowest MAPE error which is 0.09 % as it suitable for short-term forecasting in 6 months ahead. The selection of seasonal data in cases studies 2 and 3 while non-seasonal data in case study 1 also showed different situations in the forecasting results. By finding the best smoothing technique for extreme data, more accurate predictions 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.

Keywords: Exponential Smoothing Technique, SPSS Software, Seasonal and Non-seasonal, MAPE Error.

1. Introduction

There is about 90 % of water resources in Malaysia are used for industrial, domestic use and daily life purposes [5]. The main source of water in Malaysia is mainly river water that comes from rainfall. Though the rainfall is considered abundant, flood problems often occurred in Malaysia due to the abundant rainfall in rivers in Malaysia. Natural disasters that often occur in Malaysia are floods and it

is considered a continued hazard for humanity. One of the major causes of floods in Malaysia is the high and continuous distribution of rainwater from heavy rainfall for days. If this condition continues it can cause the river water level to rise drastically, this incident can cause devastating flash floods.

Pahang river is one of the areas that received the highest total rainfall throughout the year. Therefore, the study of the prediction of river water level time series is very important in order to avoid flood events. Extreme event time series are difficult to study and even harder to be used for prediction because of their rare characteristics [4]. The exponential smoothing techniques applied on monthly river water level and identifies that can forecast and analyse Sungai Pahang monthly river water level time-series data for short-term forecast. Sungai Pahang was selected in this study because Sungai Pahang was the largest water source in peninsular Malaysia and flood was very often occurred in Sungai Pahang. For that reason, 3 stations in Sungai Pahang river were examined in this study. The basin of Sungai Pahang has an annual rainfall of about 2,170 mm, a large proportion of which occurs during the North-East Monsoon between mid-October and mid-January [5].

2. Overview on exponential smoothing techniques

Three smoothing techniques which are Single Exponential Smoothing Technique (SEST), Double Exponential Smoothing Technique (DEST) and Holt winter are discussed.

Single Exponential Smoothing Technique.

This smoothing scheme starts by converting S_1, to y_1 with:

S = smoothing observation data

y = actual observation data

$t = 1, 2, ..., n.$

In the third time period, $S_3 = \alpha y_{t-1} + (1- \alpha) S_2$ (1)

Double exponential smoothing

This scheme can be proved by introducing a second formula with a fixed value, y, which should be chosen to be combined with α. Here are two formulas presented with Double Exponential Smoothing:

$S_t = \alpha y_t + (1- \alpha) (S_{t-1} + b_{t-1})$ $\qquad$ $0 <= \alpha =< 1$ (2)

$b_t = y (S_t - S_{t-1}) + (1- y) b_{t-1}$ $\qquad$ $0 <= y =< 1$ (3)

Where is the latest from the series is used to calculate the smoothed value to be replaced into multiple exponential smoothing.

In the case of Simple Exponential Smoothing, there are various schemes to provide the initial values of S_t and b_t for Multiple Smoothing. S_t is the basic preparation for y_t. Here are three types of suggestions for b_1:

$b_1 = y_2 - y_1$ (4)

$b = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]/3$ (5)

$b_1 = (y_n - y_1) / (n-I)$

Holt's winter model

This method is nearly the same as the simple exponential smoothing method, but has the advantage of reducing the update period of the trend component. The value of the data is a smoothed estimate of the

value of the data at the end of each period, and the growth in the data is a smoothed estimate of average growth in the data at the end of each period [9]. The model is based on three variables which consist of:

1) The level estimate

$$L_t = \alpha y1 + (1 - \alpha)(L_{t-1} + T_{t-1}) \tag{6}$$

2) The trend estimate

$$T_t = \beta(L_1 - L_{t-1}) + (1 - \beta)T_{t-1} \tag{7}$$

3) Forecast m period into the future

$$Y_{t+m} = pTt + Lt \tag{8}$$

Where Lt = New smoothed value, α = Smoothing constant for the level, Yt = Real value of the series in period t (actual value), β = Smoothing constant for trend estimation, Tt = Trend Estimate, p = Period to be predicted, Yt+p = Forecast for period p (estimated values).

3. Methods

The methods section, otherwise known as methodology, describes all the necessary information that is required to obtain the results of the study.

3.1 Case study

The data were collected and gathered from the Department of Irrigation and Drainage of Malaysia. The data used for experiment and testing were the historical data of river water level from January 2010 to February 2021. These data were collected in Pahang from three rivers of Sungai Pahang which were Sungai jelai in Jeram Bungor Kuala Lipis as case study 1, Sungai Tembeling in Kg Merting as case study 2 and Sungai Pahang in Temerloh as case study 3. The total number of time-series monthly river water level data will be obtained from JPS Malaysia selected for experiment and testing is 122 series data.

3.2 Result error

The results were simulated using the Exponential Smoothing Technique. First, the data is used to determine whether trends, seasonality, or both were included in the model. The performance evaluation of techniques was evaluated based on error measurement obtained by using performance metrics.

These measurements are based on the forecast errors, or how different the actual forecast is compared to the forecast. To test the model's ability to make accurate predictions, we utilise the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The formulas are as follows:

1) Percentage Error (PE)

$$PE = \frac{y1 - \hat{y}1}{y1} \times 100 \tag{9}$$

2) Mean Absolute Percentage Error identifies significant relationships between forecast data and actual monthly water level.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\{PE\} \tag{10}$$

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

$$RMSE = \sqrt{\frac{\sum_t^n e_t^2}{n}} \tag{11}$$

MAPE identifies significant relationships between forecast data and actual monthly water level. Models with a MAPE of around 30 % produce reasonable predictions while MAPEs between 5 % to 10 % produce very accurate predictions. Analyzes that produce MAPEs of around 5 % to 10 % can be considered as accurate predictions.

To determining the value of the smoothing parameter,α is based on trial and error. In order to find the best value, the value of ,α were tested which begin with 0.1, 0.2 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. The α that gives the smallest error will defined as the best value.

4. Results and Discussion

The results and discussion section presents data and analysis of the study.

4.1 Results

. The tentative models developed are used to estimate its parameters as scheduled in Table 1,2 and 3 are produced giving SSE (Sum of Squared Errors), MAPE and RMSE values respectively.

Table 1: Result Comparison using Sungai Jelai Dataset

Model	SSE	Parameter	MAPE (fit)	MAPE (forecast)	RMSE
SEST	57.00	$\alpha = 0.6$	0.441	0.280	0.490
DEST	76.90	$\alpha = 0.2$	0.517	0.227	0.522
Holt winter	41.22	$\alpha = 0.6$ $Y = 0.0$	0.557	0.323	0.441

Table 2: Result comparison using Sungai Tembeling dataset

Model	SSE	parameter	MAPE (fit)	MAPE (forecast)	RMSE
SEST	44.53	$\alpha = 0.93$	3.273	2.913	3.115
DEST	142.348	$\alpha = 0.8$	4.635	0.359	3.428
Holt winter	19.48	$\alpha = 0.5$ $Y = 0.0$	2.713	1.304	2.135

**Table 3: Result Comparison using Sungai Temerloh dataset**

Model	SSE	parameter	MAPE (fit)	MAPE (forecast)	RMSE
SEST	41.259	$\alpha = 0.48$	3.535	1.643	1.594
DEST	44.732	$\alpha = 0.1$	3.885	0.838	1.637
Holt winter	38.240	$\alpha = 0.5$ $Y = 0.0$	3.116	1.511	1.207

Referring to Table 1, the MAPE on the forecast data and fit values data generated by the SEST model is lower than the other. If seen in Table 1, also the RMSE for adaptation data shows the second lowest is SEST model. SEST model can be selected as the best model, since the SEST model has the second lowest SSE and RMSE values. Based on Table 2, the MAPE values on the forecast data and fit values data generated by the Holt winter model is lower than the other. However, all MAPE values of those models are less than 10 %. But between this three model Holt winter has the lowest value which is 0.3 %. Even so, with the RMSE also in Table 2 shows Holt winter model gives the lowest. As such, the Holt winter model was chosen as the best model in this case. Based on Table 3, the MAPE values on the forecast data and fit values data generated by the Holt winter model is lower than the other. However, all MAPE values of those models are less than 10 %. But between this three model Holt winter has the lowest value which is 0.42 %. Even so, with the RMSE also in Table 3 shows Holt winter model gives the lowest which is 1.207. As such, the Holt winter model was chosen as the best model in this case.
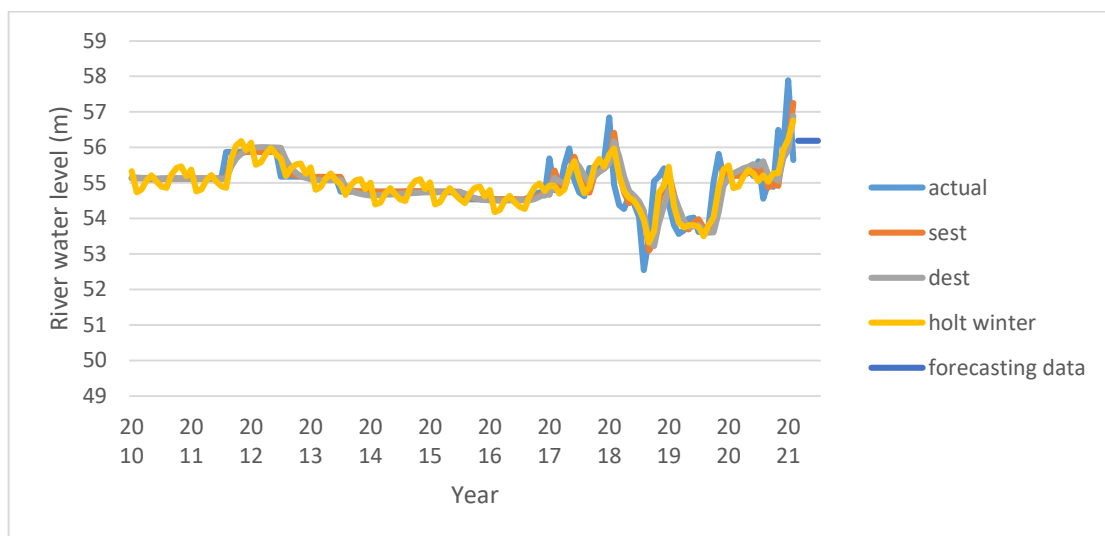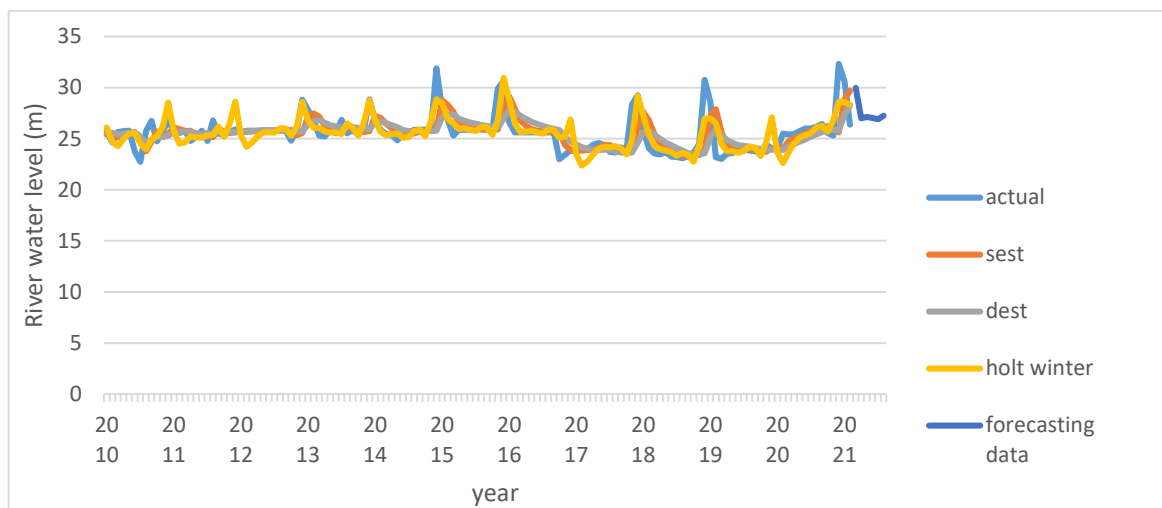

Figure 1: Time-series plot for Sungai Jelai

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 Sungai Jelai in Jeram Bungor, Sungai Tembeling in Kg Merting and Sungai Pahang in Temerloh.. The x-axis is for date and the y-axis is the monthly water level data which is measured in meters. Based on the figure 1, the values data using SEST was found to mostly lay down to the actual data. This result has shown that SEST performed better compared to DEST and Holt's method. Based on the figure 2, the data using Holt winter was found to mostly lay down to the actual data. This result has shown that the Holt winter model performed better compared to SEST and DEST. This proves that the Holt winter models are suitable for forecasting. Based on the figure 3, the fit values data using Holt winter was found to mostly lay down to the actual data. This result has shown that the Holt winter model performed better compared to SEST and DEST. This proves that the Holt winter models are suitable for forecasting.
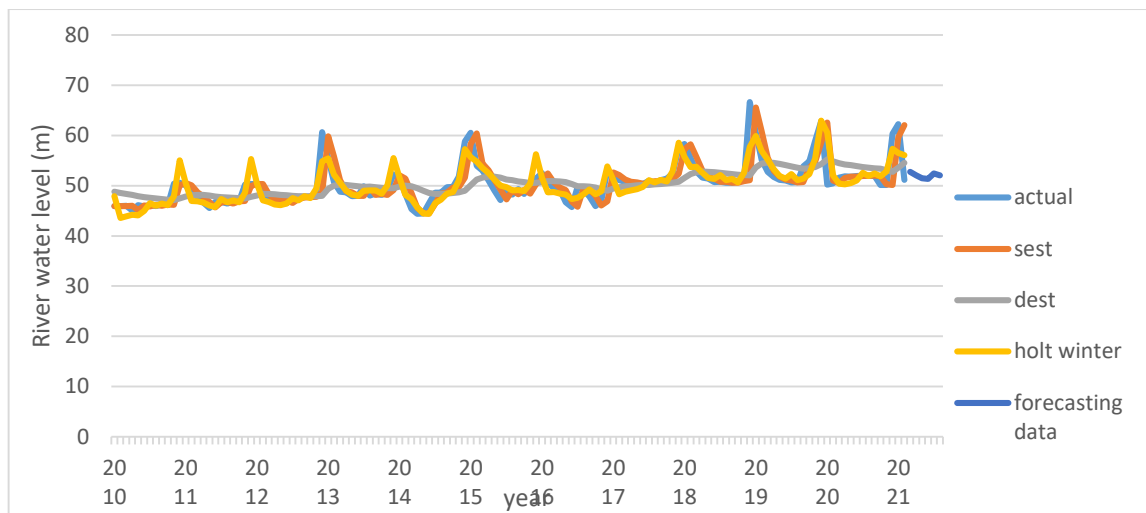
Figure 2: Time-series plot for Sungai Tembeling

Figure 3: Time-series plot for Sungai Temerloh

3.2 Discussions

Based on table 4, if we look at study case 1, the three models find that the MAPE produced is lower than studies case 2 and 3. The exponential Smoothing model that has a low MAPE and value of below 10%. This category will produce accurate predictions. It is likely that the monthly river water level time series data of this study case is more uniform and does not contain seasonal variation components in the time-series. The case model of study 2 and 3 is also a simplified model with the lowest number of parameters and MAPE values.

The models that have been analysed need to be updated after having the latest data. To produce a good forecast, the model was reconstructed using all-time series data including new data every six months. This is because the Exponential Smoothing model forecasting methods are suitable for short-term forecasting with 6-month forecasting ahead.

Since this study analyses seasonal and non-seasonal series data, it gives the impression that the accuracy of the forecast results depends heavily on the appropriateness and number of time series data available. Similarly, the time data component greatly influences the shape of the river water level forecast direction.

Table 4: The value comparison result of exponential smoothing technique

	Model	SSE	MAPE (%)	RMSE	Parameter
Case study 1: Sungai Jelai in Jeram Bungor	SEST	57.00	0.09	0.490	$\alpha = 0.6$
	DEST	76.90	0.46	0.522	$\alpha = 0.2$
	Holt winter	41.22	0.34	0.441	$\alpha = 0.6$
Case study 2: Sungai Tembeling in Kg	SEST	44.53	0.43	3.115	$\alpha = 0.93$
	DEST	142.35	0.76	3.428	$\alpha = 0.8$
	Holt winter	19.48	0.3	2.135	$\alpha = 0.5$
Case study 3: Sungai Pahang in Temerloh	SEST	41.30	0.44	1.594	$\alpha = 0.48$
	DEST	44.73	0.64	1.637	$\alpha = 0.1$
	Holt winter	38.24	0.42	3.116	$\alpha = 0.5$

5. Conclusion

This study has fulfilled the objectives of the study by developing and evaluating the model of Exponential Smoothing to predict the monthly water level of the river. All models developed are able to provide good accurate predictions. The simple exponential smoothing technique is the best model with the lowest MAPE value error of 0.09 %. However, all models formed gave MAPE the lowest values because it has error below 5 %. This study shows that the ability of the Exponential Smoothing methods to predict accurately is reasonable. The Exponential Smoothing model shows its ability to be more accurate for 6 months forecasting. The selection of seasonal data in cases study 2 and 3 while non-seasonal data in case study 1 also showed different situations in the forecasting results. The accuracy of the data is an important role in determining the accuracy of the forecast results. The amount of data must be sufficient and complete to form a good model. By finding the best smoothing technique for big data, more accurate prediction can be produced.

Acknowledgement

References

[1]   Chan, K. Y., Dillon, T. S., Singh, J., & Chang, E. (2011). Traffic flow forecasting neural

networks based on exponential smoothing method. (2011) 6th IEEE Conference on Industrial Electronics and Applications (ICIEA), 376–381. doi:10.1109/ICIEA.2011.5975612

[2]   Gelper, S., Fried, R., & Croux, C. (2010). Robust forecasting with exponential and Holt– Winters smoothing. Journal of Forecasting, 29(3), 285–300.

[3]   Georgia Papacharalampousa, & Hristos Tyralis. (2020), Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. Department of Water Resources and Environmental Engineering, School of Civil

Engineering, National Technical University of Athens, Heroon Polytechneiou 5, 157 80 Zographou, Greece

[4]   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.

[5]   Faisal,A. Azmin,N.S, Heshmatpoor A., and Hafiz,R.(2019), Identification of Flood Source Areas in Pahang River Basin, Peninsular Malaysia. Department of Environmental Science, Faculty of Environmental studies, University Putra Malaysia (UPM), 43400, UPM Serdang, Selangor, Malaysia. *The international journal published by the Thai Society of Higher Education Institutes on Environment*. Retrieved from https://www.researchgate.net/.

[6]   Hyndman, R.J., Athanasopoulos, G. (2018), Forecasting: Principles and Practice. Melbourne, Australia: Monash University

[7]   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. 1(2).

[8]   Jheison Contreras Salinas & Fernando López. (2020), Analysis of Energy Consumption in Colombia Using the Holt Method. International Journal of Energy Economics and Policy ISSN: 2146-4553

[9]   Kalekar, P. S. (2004). Time series forecasting using Holt-Winters Exponential Smoothing Under the guidance of Bernard, P., (04329008).

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

[11]   Nayak, M. A., & Ghosh, S. (2013). Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. Theoretical and Applied Climatology.

[12]   Noor Shahifah Muhamad & Aniza Mohamed Din. (2015) exponential smoothing techniques on time series river water level data. Universiti Utara Malaysia (UUM), Malaysia

[13]   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.

[14]   Yuk Feng Huang, Shin Ying Ang, Khia Min Lee and Teang Shui Lee. (2015). Quality of Water Resources in Malaysia.IntechOpoen DOI: 10.5772/58969