International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 11 (2019), pp. 1864-1868

Forecasting Foreign Tourist Arrivals to Bali: Hybrid Double Exponential Smoothing Approach

Seng Hansun, Marcel Bonar Kristanda

Department of Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia.

Abstract

Bali is one of the most popular tourist destinations in the world. With its magnificent nature and enchanting culture, it has attracted many people to come and visit every year. The need to have a proper and accurate tourist flow prediction has become important since it could be a scientific reference for decision making process for the development of tourism sectors in Bali province. In this research, we try to forecast the foreign tourist arrivals to Bali by using its monthly distribution data recorded from January 2008 to December 2018. Two hybrid double exponential smoothing methods, i.e. B-WEMA and H-WEMA, were successfully implemented and could predict the future foreign tourist arrivals number. B-WEMA excels H-WEMA in terms of accuracy level which was calculated using MSE and MAPE error measurements. Furthermore, the prediction of foreign tourist arrivals to Bali province for 2019 was given which showed a slight increase compare to the number of foreign tourist arrivals in 2018.

Keywords: Bali, Forecasting, Foreign tourist arrivals, Hybrid double exponential smoothing, monthly distribution data.

I. INTRODUCTION

One of the key development factors and income sources for many developing countries is the tourism sector. It is the main factor for export income, job creation, and local development [1]. In Indonesia, it has become the third largest GDP contributor after oil and mining sectors [1].

Accurate prediction of tourist flow has become a key issue in tourism economic analysis and development planning [2]. In fact, it has attracted much attention from notable researchers and practitioners, as we can see from the increasing academic literature in this domain [3]. Some of them using statistical approach, as we can see in the works of Hopken et al. [4], Zhu et al. [5], and Tung [6]. Others used artificial intelligence and machine learning approach, such as integrated fuzzy time series model [7] and Long Short-Term Memory (LSTM) network [8]. Some others even proposed hybrid approaches as the works of Liu et al. [9], Sun et al. [10], and Binru et al. [11].

One of the most popular tourist destinations in the world is Bali province [12]. It is located in Indonesia, more exactly at 8°3'40" - 8°50'48" S dan 114°25'53" - 115°42'40" E [13]. Figure 1 shows the Bali island map.

Fig. 1. Bali island map [14]

With its magnificent nature and enchanting culture, Bali has attracted many people around the world to come [15]. Bali Government Tourism Office has reported an increasing trend of foreign tourist arrivals to Bali each year. In 2016, around 4.9 million foreign tourists had come to Bali and increased to 5.7 million people in 2017, and increased again in 2018 to 6.0 million people [16]. Therefore, there is a need for a proper and accurate foreign tourist flow prediction to Bali province, since it could become a scientific reference for decision making process of tourism related departments.

In this research, we try to forecast foreign tourist arrivals to Bali by its monthly distribution data. Two hybrid double exponential smoothing methods will be incorporated here, i.e. the Brown's Weighted Exponential Moving Average (B-WEMA) and Holt's Weighted Exponential Moving Average (H-WEMA) methods. Both of them are improved version of original WEMA method which was introduced in 2013 [17]. Moreover, to get the accuracy level of both methods applied in this study, we used Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE), two of the most common used forecast error measurement criteria in time series analysis. In the next section, we will give more detail of B-WEMA and H-WEMA followed by the explanation of MSE and MAPE in Section 3. The implementation results and analysis will be given in Section 4, and some conclusion remarks will end the organization of this paper.

II. B-WEMA and H-WEMA

Weighted Exponential Moving Average (WEMA) is a hybrid moving average (MA) methods that combines the weighting factor calculation as can be found in Weighted Moving Average (WMA) and the procedure of Exponential Moving Average (EMA). It was first published by Hansun in 2013 [17]. Since then, WEMA has been applied by many researchers in different real-cases, such as to generate differential physical layer secret key [18], to predict the function of transplanted kidney in long-term care process [19], and to forecast the big five ASEAN capital markets [20].

In this section, we discuss two improved versions of WEMA, i.e. B-WEMA and H-WEMA, which were applied in this study.

II.I Brown's Weighted Exponential Moving Average

B-WEMA is a hybrid MA method that combines the weighting factors calculation in WMA with the procedures of Brown's Double Exponential Smoothing (B-DES). Since it originally comes from B-DES procedures, B-WEMA procedures look likely with B-DES except for the initial step of finding the base value, B_t [21].

The first step of B-WEMA is finding its base value using

$$B_t = \frac{\sum_{t=k-n+1}^{k} w_t A_t}{\sum_{t=k-n+1}^{k} w_t} \quad (1)$$

where n is the span data, k is the relative position of the period currently being considered within the total number of periods, A_t is the actual value at time t, and w_t is the weight of A_t.

Next, we can find the prediction value by calculating the single-smoothed and double-smoothed series using

$$S_t' = \alpha Y_t + (1-\alpha) S_{t-1}' \quad (2)$$

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

The prediction value is formulated as

$$F_{t+k} = L_t + kT_t, \quad (4)$$

where

$$L_t = 2S_t' - S_{t-1}'' \quad (5)$$

$$T_t = \frac{\alpha}{1-\alpha}(S_t' - S_{t-1}'') \quad (6)$$

We start the model by defining

$$S_{t-1}' = S_{t-1}'' = B_t \quad (7)$$

and all of those processes will be repeated for all data point in the time series data.

II.II Holt's Weighted Exponential Moving Average

H-WEMA method is another version of hybrid double exponential smoothing method. It has the same underlying background and building methods as B-WEMA, i.e. WMA, but instead of combining it with B-DES, we combine it with Holt's Double Exponential Smoothing (H-DES) method. Here is the description of H-WEMA procedures [22].

Find the base value of H-WEMA using

$$H_t = \frac{\sum_{t=k-n+1}^{k} w_t A_t}{\sum_{t=k-n+1}^{k} w_t} \quad (8)$$

Using H_t value obtained, next we calculate the prediction value using

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (9)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta) T_{t-1} \quad (10)$$

$$F_{t+k} = L_t + kT_t \quad (11)$$

where

$$L_{t-1} = H_{t-1} \quad (12)$$

$$T_{t-1} = H_t - H_{t-1} \quad (13)$$

All those processes will be repeated for all remaining data point in the time series data.

III. MSE AND MAPE

To get the accuracy results of B-WEMA and H-WEMA methods implemented in this study, two error measurements were used, i.e. Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE), in order to get the forecast error rate. Those two measurements are widely known and used in time series analysis domain. MSE can be found using

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \quad (14)$$

where n refers to the total number of data, A_t is the actual value of data, and F_t is the forecasted value of data.

Moreover, MAPE can be calculated using

$$MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|\right) \cdot 100\% \quad (15)$$

As can be clearly seen in Eq.(15), the accuracy in MAPE is expressed as a percentage value.

IV. RESULTS AND ANALYSIS

As has been stated before, we try to forecast the foreign tourist arrivals to Bali by using its monthly distribution data. The distribution data was then collected from Bali Government Tourism Office which recorded the data from January 2008 to February 2019. Table 1 shows the monthly distribution data of foreign tourist arrivals to Bali province [16].

Table 1. 2008-2018 monthly distribution data of foreign tourist arrivals to Bali

Month	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Jan	139,872	164,643	168,923	202,660	248,289	232,935	279,257	301,748	350,592	460,824	358,065
Feb	155,153	139,370	187,781	201,320	219,475	241,868	275,795	338,991	375,744	453,985	452,423
Mar	153,929	161,169	194,482	201,833	227,846	252,210	276,573	305,272	364,113	425,499	492,678
Apr	147,515	179,879	178,549	221,014	219,984	242,369	280,096	313,763	380,767	477,464	516,777
May	159,877	181,983	196,719	204,489	215,868	247,972	286,033	295,973	394,557	489,376	528,512
Jun	170,994	190,617	219,574	240,154	238,296	275,667	330,396	359,702	405,835	504,141	544,550
Jul	183,122	224,636	247,778	278,041	258,781	297,878	361,066	382,683	484,231	592,046	624,366
Aug	187,584	222,441	236,080	250,835	254,020	309,219	336,763	303,621	438,135	601,884	573,766
Sep	181,033	208,185	229,573	251,737	243,722	305,629	354,762	389,060	445,716	550,520	555,903
Oct	180,944	210,935	223,643	241,232	255,709	266,562	341,651	369,447	432,215	465,085	517,889
Nov	142,014	163,531	194,152	216,384	241,985	307,276	296,876	270,935	413,232	361,006	406,725
Dec	166,855	182,556	215,804	246,880	268,044	299,013	347,370	370,640	442,800	315,909	498,819
Total	1,968,892	2,229,945	2,493,058	2,756,579	2,892,019	3,278,598	3,766,638	4,001,835	4,927,937	5,697,739	6,070,473

To predict the foreign tourist arrivals to Bali, we used two hybrid double exponential smoothing methods, i.e. B-WEMA and H-WEMA. Next, we divided the dataset we have into two parts, one for the training data and one for the test data. The data from January 2008 to December 2015 will be our training set, while the data from January 2016 to December 2018 will be our test set. Figure 1 and 2 show the forecasting graphs of foreign tourist arrivals to Bali on the training data using five span period and 12 initial data for B-WEMA and H-WEMA consecutively.

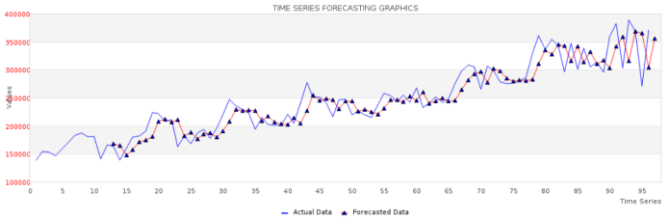

Fig. 2. B-WEMA forecasting results

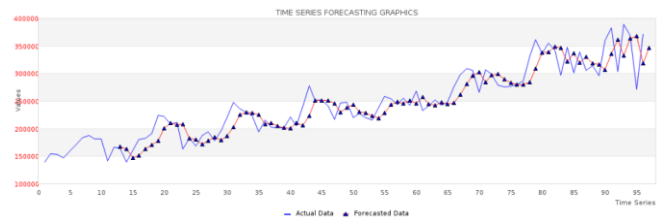

Fig. 3. H-WEMA forecasting results

Next, we try different parameter values for the span period data to get the best alpha in B-WEMA and the best alpha and gamma in H-WEMA for this dataset. Table 2 shows the parameters being used and its corresponding alpha and error rate values for B-WEMA, while Table 3 shows the parameters being used and its corresponding alpha and gamma values for H-WEMA.

Table 2. B-WEMA parameters finding

Span no.	Best alpha	MSE	MAPE
2	0.09	756710761.32602	7.8635118956942
3	0.16	695825577.78671	7.6461205560419
4	0.21	754310388.38797	7.8723146334738
5	0.23	755013753.11102	7.8336363117554
6	0.23	744492143.54229	7.7876944518195
7	0.25	748029242.79958	7.7887610891421
8	0.26	752253771.72044	7.8005203026556
9	0.27	746236395.36853	7.7465857170653
10	0.27	738003951.40048	7.6958255670862
11	0.26	724869614.31292	7.6399546522178
12	**0.26**	**713157232.28013**	**7.5881903277946**

Table 3. H-WEMA Parameters Finding

Span no.	Best alpha	Best gamma	MSE	MAPE
2	0.54	0.03	761518458.50749	8.0572803262077
3	0.53	0.02	749215154.05092	7.9088081827676
4	0.55	0.02	757077138.19602	7.9270770628331
5	0.52	0.02	739009352.81943	7.8419910135061
6	0.56	0	739281071.70919	7.7520371463652
7	0.53	0	716821769.05352	7.7376918211077
8	0.48	0	700773912.64579	7.7380368914531
9	0.51	0	710571451.38354	7.7206640854425
10	0.53	0	715564385.98797	7.71451962034
11	**0.53**	**0**	**714273902.63146**	**7.712006760961**

From the parameters finding phase, we choose to use span period of 12 for B-WEMA and span period of 11 for H-WEMA which have the smallest MAPE values. After we found the parameters for best scenario of each method, we used them on the test set as can be seen on Figure 3 and 4.

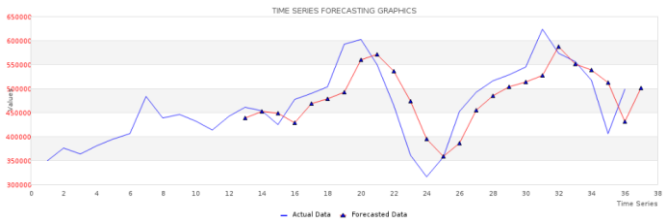

Fig. 4. B-WEMA forecasting results on test set using span=12

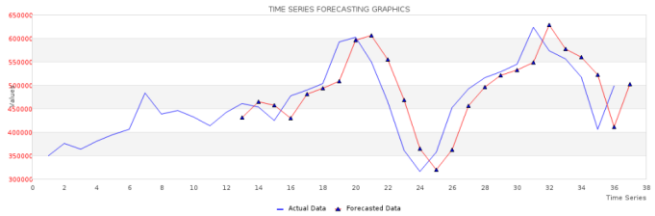

Fig. 5. H-WEMA forecasting results on test set using span=11

The MSE and MAPE values for B-WEMA on the test set are 3135624841.3253 and 9.6290702364251, while the values for H-WEMA on the test set are 3346943278.1265 and 10.226705561589. Clearly, B-WEMA outperforms H-WEMA in this case and therefore, it is advisable to use B-WEMA in predicting foreign tourist arrivals to Bali. Furthermore, we try to predict the foreign tourist arrivals to Bali in 2019 using B-WEMA as can be seen in Table 4.

Table 4. 2019 B-WEMA prediction results

Month	Forecasted Data
Jan'19	501,373
Feb'19	505,747
Mar'19	509,801
Apr'19	512,979
May'19	515,157
Jun'19	516,453
Jul'19	516,849
Aug'19	515,119
Sep'19	512,789
Oct'19	510,320
Nov'19	508,421
Dec'19	509,011
Total	**6,134,019**

IV. CONCLUSION

We have successfully forecast foreign tourist arrivals to Bali province using its monthly distribution data. Two hybrid double exponential smoothing methods, i.e. B-WEMA and H-WEMA, have been applied and using different parameter settings, we found that B-WEMA outperformed H-WEMA for the given dataset.

Furthermore, we also have tried to forecast the foreign tourist flows in 2019. It seems that there is a slight increase in the number of foreign tourist arrivals compare to the total number of tourist arrivals in 2018, i.e. around 1.1%.

We can explore a deeper analysis and prediction on foreign tourist arrivals to Bali based on their nationality in the near future. Comparison of the prediction results with other variants of hybrid moving average methods can also be conducted in the next research.

ACKNOWLEDGMENT

This research is supported and funded by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia (Kemenristekdikti) through the Research Grant funding under 'Penelitian Dasar Unggulan Perguruan Tinggi' scheme with Contract No.1581/HD/LPPM-UMN/IV/2019.

REFERENCES

[1] P. Sugiartawan, S. Hartati, and A. Musdholifah, "Tourist Visits Prediction with Fully Recurrent Neural Network," in *International Conference on Information Technology, Engineering, Science & its Applications*, 2018, pp. 1–10.

[2] C. Kang and J. Gu, "Analysis of Tourist Flow Forecasting Model Based on Multiple Additive Regression Tree," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 490, pp. 1–5, Apr. 2019.

[3] Y. Yao *et al.*, "A paired neural network model for tourist arrival forecasting," *Expert Syst. Appl.*, vol. 114, pp. 588–614, 2018.

[4] W. Höpken, T. Eberle, M. Fuchs, and M. Lexhagen, "Search Engine Traffic as Input for Predicting Tourist Arrivals," in *Information and Communication Technologies in Tourism 2018*, Cham: Springer International Publishing, 2018, pp. 381–393.

[5] L. Zhu, C. Lim, W. Xie, and Y. Wu, "Modelling tourist flow association for tourism demand forecasting," *Curr. Issues Tour.*, vol. 21, no. 8, pp. 902–916, 2018.

[6] L. T. Tung, "Forecasting the Foreign Tourist Arrivals to Vietnam using the Autoregressive Integrated Moving Average (ARIMA) Method," *J. Environ. Manag. Tour.*, vol. 9, no. 30, pp. 1135–1144, 2018.

[7] N. Ramli, S. N. A. Nor Azemi, N. S. Omar, and S. M. Ab Mutalib, "Integrated Fuzzy Time Series Model for Forecasting Tourist Arrivals," in *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, Singapore: Springer Singapore, 2018, pp. 851–858.

[8] Y. Du, J. Liu, and S. Wang, "Consulting and Forecasting Model of Tourist Dispute Based on LSTM Neural Network," in *Green, Pervasive, and Cloud Computing*, 2019, pp. 495–500.

[9] H.-H. Liu, L.-C. Chang, C.-W. Li, and C.-H. Yang, "Particle Swarm Optimization-Based Support Vector Regression for Tourist Arrivals Forecasting," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Sep. 2018.

[10] S. Sun, Y. Wei, K.-L. Tsui, and S. Wang, "Forecasting tourist arrivals with machine learning and internet search index," *Tour. Manag.*, vol. 70, pp. 1–10, Feb. 2019.

[11] Z. Binru, P. Yulian, H. Rong, and T. Runzhi, "Analysis on the Role of Daily Consumer Search Data in Forecasting Monthly Tourist Flow A Mixed Data Sampling Approach," in *Proceedings of the 4th International Conference on Economics, Management, Law and Education (EMLE 2018)*, 2018.

[12] M. Khamdevi and H. Bott, "Rethinking tourism: Bali's failure," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 126, p. 012171, Mar. 2018.

[13] P. P. Bali, "Letak Geografis, Batas Administrasi, dan Luas Wilayah," 2010. [Online]. Available: http://www.baliprov.go.id/id/Geographi. [Accessed: 15-May-2019].

[14] BaliTrekkingTour, "Geography of Bali." [Online]. Available: https://www.balitrekkingtour.com/geography-of-bali/. [Accessed: 15-May-2019].

[15] D. P. E. Pratiwi, I. K. Sulatra, and K. D. P. Candra, "Bali tourism advertisements: a linguistic analysis," *Int. J. Linguist. Lit. Cult.*, vol. 5, no. 1, pp. 43–53, 2019.

[16] D. P. P. P. Bali, "Statistik," 2019. [Online]. Available: http://www.disparda.baliprov.go.id/id/Statistik4. [Accessed: 15-May-2019].

[17] S. Hansun, "A new approach of moving average method in time series analysis," in *2013 Conference on New Media Studies (CoNMedia)*, 2013, pp. 1–4.

[18] M. Chen, T. Jiang, and W. Zou, "Differential physical layer secret key generation based on weighted exponential moving average," in *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2015, pp. 1–4.

[19] P. Rashidi Khazaee, J. Bagherzadeh M., Z. Niazkhani, and H. Pirnejad, "Predicting the function of transplanted kidney in long-term care processes: Application of a hybrid model," *J. Biomed. Inform.*, vol. 91, p. 103116, Mar. 2019.

[20] S. Hansun, M. B. Kristanda, and P. M. Winarno, "Big 5 ASEAN capital markets forecasting using WEMA method," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 1, pp. 314–319, Feb. 2019.

[21] S. Hansun, "A New Approach of Brown's Double Exponential Smoothing Method in Time Series Analysis," *Balk. J. Electr. Comput. Eng.*, vol. 4, no. 2, pp. 75–78, 2016.

[22] S. Hansun and S. Subanar, "H-WEMA: A New Approach of Double Exponential Smoothing Method," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 2, pp. 772–777, Jun. 2016.