Int. J. Advance Soft Compu. Appl, Vol. 9, No. 1, March 2017
ISSN 2074-8523

Tourism Forecasting using Hybrid Modified Empirical Mode Decomposition and Neural Network

Nurhaziyatul A. Yahya[1], Ruhaidah Samsudin[1], and Ani Shabri[2]

[1]Faculty of Computing
[2]Faculty of Science
Universiti Teknologi Malaysia,
81300, Johor Bahru, Johor, Malaysia
email: haziahadawiyah@gmail.com,
ruhaidah@utm.my, ani@utm.my

Abstract

Due to the dynamically increasing importance of the tourism industry worldwide, new approaches for tourism demand forecasting are constantly being explored especially in this Big Data era. Hence, the challenge lies in predicting accurate and timely forecast using tourism arrival data to assist governments and policy makers to cater for upcoming tourists. In this study, a modified Empirical Mode Decomposition (EMD) and Artificial Neural Network (ANN) model is proposed. This new approach utilized intrinsic mode functions (IMF) produced via EMD by reconstructing some IMFs through trial and error method, which is referred to in this research as decomposition. The decomposition and the remaining IMF components are then predicted respectively using ANN model. Lastly, the forecasted results of each component are aggregated to create an ensemble forecast for the tourism time series. The data applied in this experiment are monthly tourist arrivals from Singapore and Indonesia from the year 2000 to 2013 whereby the evaluations of the model's performance are done using two well-known measures; RMSE and MAPE. Based on the empirical results, the proposed model outperformed both the individual ANN and EMD-ANN models.

Keywords: *Big Data, Empirical Mode Decomposition, Forecasting, Neural Network, Time Series, Tourism Demand.*

1 Introduction

The tourism industry in Malaysia started to blossom around the year 1999 when Malaysia Tourism Promotion Board (MTPB) promoted Malaysia through a worldwide campaign ‘Malaysia, Truly Asia’. The campaign was a huge success, consequently attracting approximately 7.4 million tourists [1]. From then onwards, various marketing and campaigns were made by the agency to further promote Malaysia throughout the world. As a result, the number of tourists coming to Malaysia escalated in just a few years, gradually becoming one of the popular tourist destinations in ASEAN. Tourism’s impact on the economic and social development of a country can be massive. For years, the tourism industry has assisted in boosting commodities sales and other services such as transportations, accommodation, cultural, sports, recreational as well as retail trade. In other words, this industry has contributed immensely to gross domestic product (GDP) and employment. This is true not only in Malaysia, but also worldwide. Hence, systematic strategic is required so that government, entrepreneurs, organizations and businesses for local and international can be more prepared to cater for the upcoming tourists. The need to provide for the gradually growing market is vital to ensure the satisfaction of the tourists, especially in terms of transportations and accommodations. In order to do so, forecasting the arrival of tourists to Malaysia plays a crucial part in assisting with the decision making and critical planning.

Recently, Big Data has begun seeping into the tourism industry as many organizations have realized the potential of Big Data. Due to the advancement of science and technologies, enormous amount of real life data is being produced every day. The generation of large datasets which cannot be handled via traditional computing techniques is called Big Data. Nevertheless, big data does not only take account volume, but also the variety and velocity of data. These data are generated through various means, especially online sources such as TripAdvisor, Hotels.com, Flickr, Facebook, Instagram, and much more [2]. The accumulation of data such as tourism arrivals, airline reservations, hotel reservations, and car rental reservations over time also adds up an immense number of data over time in conventional databases [3]. By utilizing the data, it can aid in tactical decision making as well as assist in personalised marketing [4] to further increase the arrival of tourist by country. Hence this study aims to propose a model which can cater for the complexity of tourism big data, mainly the tourist arrival.

One of the notable AI techniques which have been applied in various research areas is the artificial neural network (ANN) model. This model is well known for its ability to perform non-linear modeling, which is a huge advantage over linear models because real world data often contains non-linear characteristics. Earlier empirical studies showed that when the performance of ANN in tourism demand forecasting are compared with naïve, moving average, decomposition, single exponential smoothing, ARIMA, multiple regression and genetic regression, ANN managed to outperform all the models [5-8]. Many of the empirical results also found that ANN is a suitable modeling tool for forecasting tourism demand [9-15]. Nevertheless, even though ANN has the advantages of producing an accurate forecast, it does have its limitations as it has the tendency to produce inconsistent performance in some situation. Additionally, ANN might suffer from some weakness such as overfitting and local minima.

Pre-processing of raw time series data is significant before conducting a forecast due to the existence of noise in the data. An empirical study by [16] found that raw data which has not been preprocessed caused failure for ANN in capturing seasonal and trend pattern in the data. It should be noted that most tourism demand data exhibits these patterns. The empirical mode decomposition (EMD), as proposed by [17] was designed to decompose any complex time series into a finite and often small number of components referred to as Intrinsic Mode Functions (IMFs). This technique is a self-adaptive decomposition technique, which among its advantages are extracting reduced number of components as well as its application to data which contains both linear and non-linear characteristics. Since the tourism time series have been mentioned to possess non-linear characteristic, it is believed that EMD technique is capable of analyzing tourism time series data. Empirical results from previous researchers have shown successful applications of EMD on various time series [9, 18-23]. Nonetheless, only a few types of researches have applied EMD on tourism demand [9, 21]. While [9] proposed hybridization of EMD with ANN, and [21] applied EMD with group method data handling (GMDH) model for tourism demand forecasting. Both empirical results proved that hybridizing EMD with other forecasting model significantly improves the accuracy of the forecasting model when compared to the single model and other notable forecasting models without applying EMD.

In this study, modified empirical mode decomposition (EMD) which is developed based on the conventional EMD technique is proposed. First, EMD is used to decompose the tourism time series into IMFs and a residue. Next, a new component is created by summing up some IMFs and the residue. The ANN model is then applied to forecast for each component. The final forecasted values are obtained by summing all the predicted values of the models. Additionally, the monthly tourist arrival data from Singapore and Indonesia are explored to ensure wider application of conclusions.

The organization of the paper is as follow: Section 2 described the proposed hybrid ensemble learning paradigm in detail. The predicted results, as well as the efficiency of the proposed method are discussed in Section 3. Finally, Section 4 presented the conclusion of the paper.

2 Methodology Formulation

2.1 Artificial neural network

Artificial neural networks (ANN) has consistently been a notable model which was thoroughly studied and applied in various areas such as science and engineering due to its ability to learn the non-linear patterns from past data as well as its flexibility. The most popular ANN structure is the single hidden layer feedforward ANN because it is known to be suitable for modeling time series data. The general structure of a feedforward ANN can be classified into; (a) the input layer (b) the hidden layer and (c) the output layer as shown in Fig. 1.

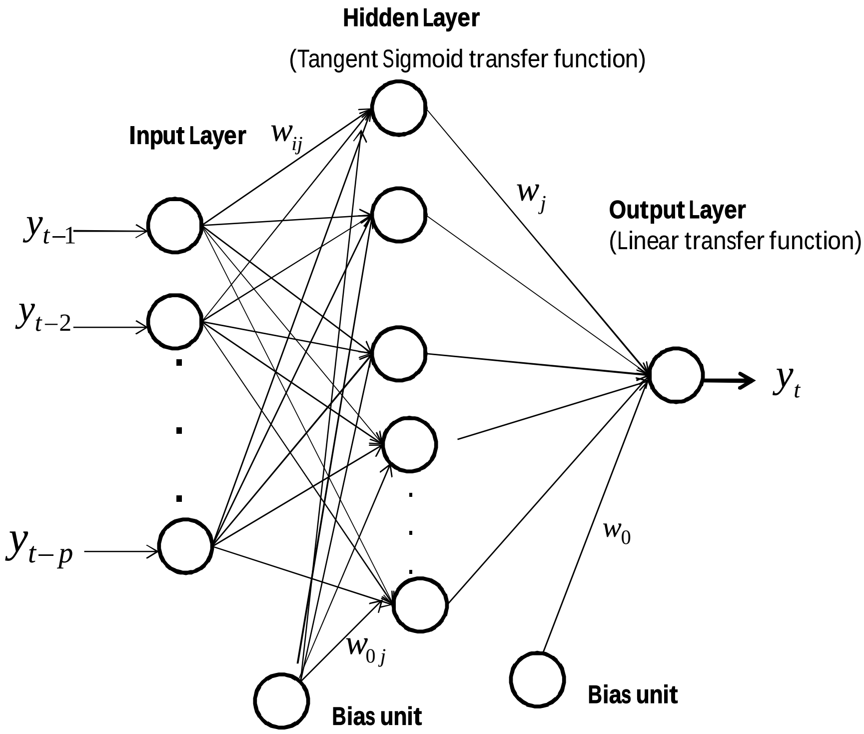

Fig. 1: Architecture of three layer feed-forward ANN

Given an output y_t and a set of input $(y_{t-1}, y_{t-2}, \cdots, y_{t-p})$ where p is the number of observations. The relationship between the input and the output can be deduced based on the following Equation (1):

$$y_t = g((b_h + \sum_{i=1}^{p} w_i f(x_i)) \quad (1)$$

Where the bias of the neuron h is represented by b_h, w_i is the connection weight, while g and f are the layer's activation functions.

2.2 Empirical mode decomposition

EMD is a novel disintegration of data method applied in Hilbert-Huang Transformation (HHT) to process data which are not linear as well as data that are not stationary. Generally, EMD works by decomposing an original time series data into smaller components called intrinsic mode function (IMF) and also a residual. The residue depicts the trend of the time series data, hence, in some empirical work, they tend to ignore this component due to its small and insignificant value [18]. The EMD algorithm is further explained below:

1. Classify every local extrema, including local maxima and minima for data series $y(t) \quad (t = 1,2,...,n)$
2. Connect every local extrema with a spline interpolation in order to obtain its upper and lower envelopes, $y_U(t)$ and $y_L(t)$ respectively.
3. Calculate the average of upper and lower envelopes whereby $m(t) = [y_L(t) + y_U(t)]/2$.
4. Extract the details by evaluating the dissimilarity between $y(t)$ and $m(t)$ as $z(t)$ whereby $z(t) = y(t) - m(t)$.
5. To ensure that $z(t)$ satisfies the requirements of IMF;
6. When $z(t)$ accomplished the required conditions, an IMF is produced. Hence, $y(t)$ is substituted with the residual $r(t) = y(t) - z(t)$
7. However, if $z(t)$ does not satisfy the requirements, replace $y(t)$ with $z(t)$ whereby $y(t) = z(t)$.
8. The steps mentioned above are repeated until the stopping criterion is reached. The process from steps (1)-(8) can be repeated until no further IMFs can be extracted from the data.

By applying the procedure above, the data series $y(t)$ can be expressed as shown in Equation (2):

$$y(t) = \sum_{i=1}^{n} z_i(t) + r_n(t) \tag{2}$$

Where n is the quantity of IMFs, $z_i(t)$ is the IMFs and $r_n(t)$ symbolized residue.

2.3 Hybrid EMD-ANN model

For the process of hybrid EMD-ANN forecasting procedure, the technique of EMD are applied at the early stage by dealing with time series data, followed by ANN which will model the time series as shown in Fig. 2.

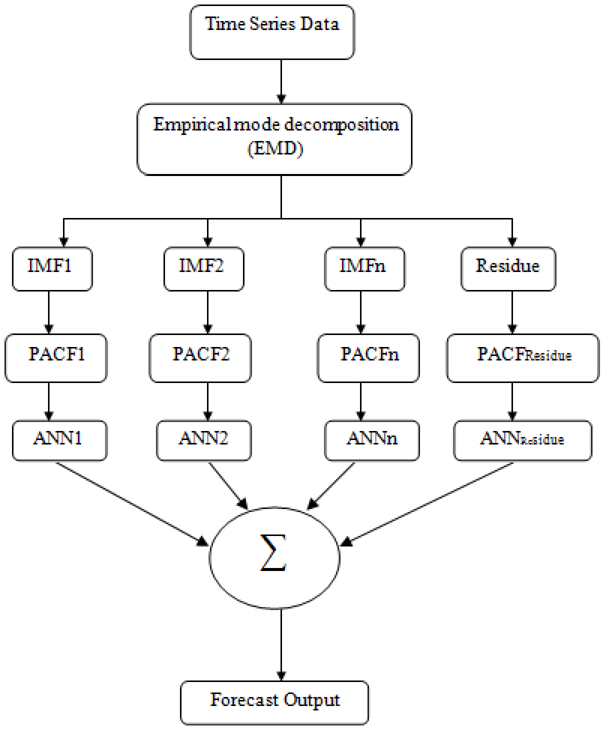

Fig. 2: Hybrid EMD-ANN architecture

The technique can be further expressed as follows:

1. Given a time series data, the EMD technique is applied on the data to disintegrate the complex original time series into simpler *k*-IMFs components and a residue $r_k(t)$.
2. The ANN model is then used as a forecasting model for each component which was previously extracted from the original time series.
3. The forecasting values of all IMFs and the residual components obtained via ANN model are summed to generate the final predicting result for the original time series.

2.4 Modified hybrid EMD-ANN model

The procedures of the proposed technique are almost similar to hybrid EMD-ANN. The only difference is, the EMD components with lower frequency (starting with residue), is summed with a higher frequency component sequentially, consequently reducing the number of components. It should be noted that this technique is done through trial and error method.

1. The first process is done by summing up the residue with the last IMF, creating the first decomposition.
2. Then, ANN model is used to forecast the remaining IMFs and the first decomposition.

3. The forecasting values are then summed up to produce the final forecasting result. This is the first modified hybrid EMD ANN model.

The processes are repeated in the same manner, cumulating the components each time. The number of decompositions models created in this research depends on the number of IMFs produced by EMD as shown in Equation (3):

$$\textit{Number of decompositions model} = \textit{number of IMFs} - 1 \qquad (3)$$

3 Experimental Results

3.1 Data sets

The real data used are the monthly data of tourism arrivals from two neighbouring countries of Malaysia; Singapore and Indonesia which are obtained through Malaysian's tourism authorities. The arrival of tourists from Singapore and Indonesia are used in this study because these countries generated the highest number of tourists that came to Malaysia in the year 2015 [24]. The data are taken from January 2000 until December 2013 (168 observations).

A training and test set are essential in building an ANN forecasting model because it could affect the performance of the model [25]. While training set is used for developing the network, the testing set is used validate the network. Usually, the division of data into training and testing set are based on the rule of 90% training vs. 10% testing, 80% training vs. 20% testing or 70% training vs. 30% testing [25]. In this study, the monthly tourism arrival for both Singapore and Indonesia data are divided into 144 for training and 24 for testing. Fig. 3 illustrated the graphs of both Singapore and Indonesia arrival series in Malaysia.

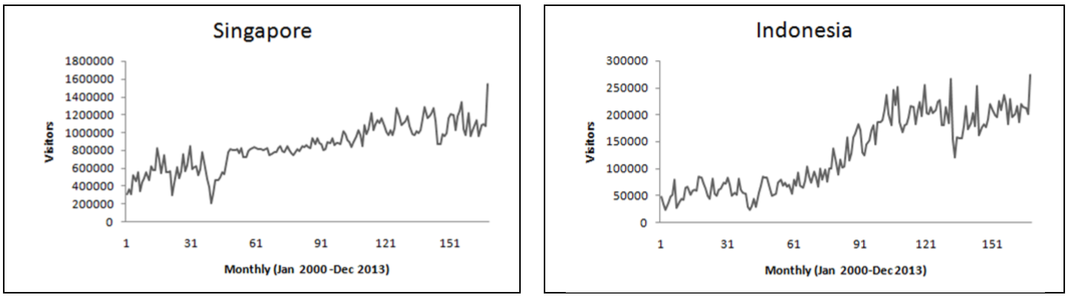

Fig. 3: Monthly Singaporean and Indonesian tourists' arrival in Malaysia

3.2 Performance criteria

Performance criteria are essential in measuring the accuracy of prediction as well as to evaluate the models. The training and forecasted data are evaluated using root mean squared error (RMSE) and mean absolute percentage error (MAPE). These are the frequently used performance measure [25]. The mathematical equations for RMSE and MAPE are defined as shown in Equation (4) and (5):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{5}$$

where y_t is the actual value, $\hat{y}_t$ is the forecasted value and n is the sample size. It should be noted the best model is chosen based on the relatively small values of RMSE and MAPE.

3.3 Forecasting results

In this research, the implementation of EMD algorithm is done using EMD library in R software package. According to the previous step in section 2.2, the monthly tourism arrivals data are first inserted into the EMD in R. The data series are then decomposed into several independent IMFs and a residue as shown in Fig. 4 and 5. Nevertheless, the number of IMFs produced by each data differs according to the stopping criterion. As shown clearly in Fig. 4, the Singapore Visitors data are decomposed into seven IMFs and one residue. Meanwhile, Fig. 5 illustrates the decomposition of Indonesia Visitors into six IMFs and one residue. Based on the figures, all of the IMFs extracted are aligned from components of high frequency to components with low frequency. Additionally, the last IMF component (residue) represents the trend of the time series. These components are important in this research as it helps to improve the forecasting performance through the “divide and conquer” approach.

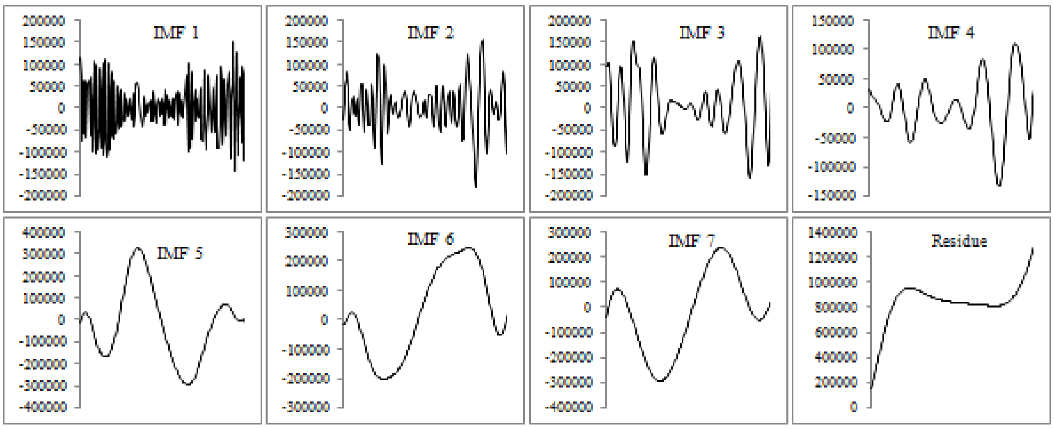

Fig. 4: The IMFs and residual produced via EMD of Singapore data

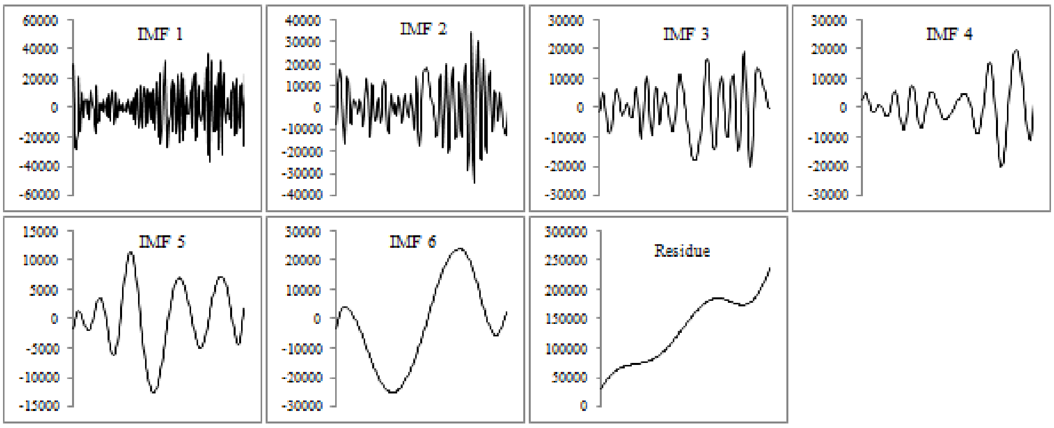

Fig. 5: The IMFs and residual produced via EMD of Indonesia data

Despite ANN's reliability, one of the prominent issues regarding this model is that there are no general rules for determining the optimal number of input and hidden nodes. According to [26], PACF plays a vital role in determining the number of inputs for artificial intelligent models. The implementation of PACF in this research is done using R software package. In the execution of PACF on the input sets, only the original training sets are applied. As a result, graphs with data lags will be displayed in the R platform. The selection of inputs is done by analysing the graphs. A blue dotted line will act as a threshold, and only lags which exceed the threshold will be kept. An example of this process is shown in Fig. 6. These steps are implemented on the original data, IMFs components and residual components for Singapore and Indonesia visitors respectively. By observing the figures mentioned, the input variables for ANN modelling are selected as shown in Table 1.

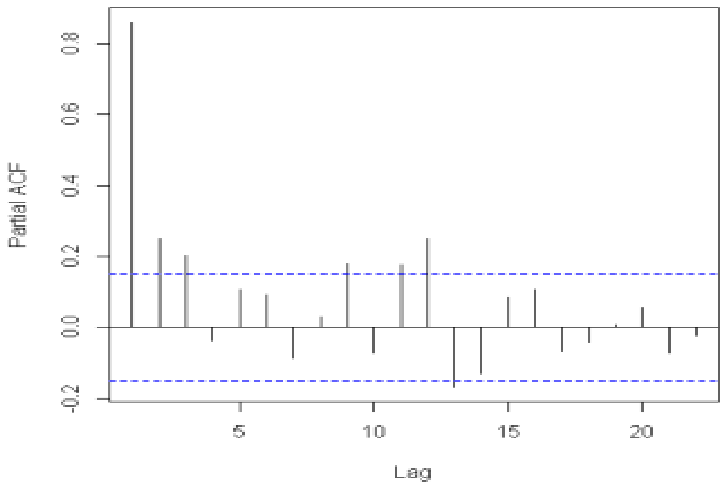

Fig. 6: An example of PACF plot for original Singapore data

Table 1: Input variables for ANN modelling based on PACF

	Singapore	Indonesia
Original data	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-9}, y_{t-11}, y_{t-12}, y_{t-13}	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-5}, y_{t-12}, y_{t-13}
IMF 1	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-9}, y_{t-12}	y_{t-1}, y_{t-2}, y_{t-4}, y_{t-5}, y_{t-8}, y_{t-11}, y_{t-12}, y_{t-13}
IMF 2	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}, y_{t-6}	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}
IMF 3	y_{t-1}, y_{t-2}, y_{t-3}	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}
IMF 4	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}	y_{t-1}, y_{t-2}
IMF 5	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}	y_{t-1}, y_{t-2}, y_{t-3}
IMF 6	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}
IMF 7	y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}	
Residual	y_{t-1}	y_{t-1}

For issues concerning the determination of hidden layer's number of nodes, rather than performing tedious trials and errors method, this research applied the guidelines prepared by [25]. In the case where *n* is the number of input nodes, the number of hidden layers could be:

1. $2n+1$
2. $2n$
3. n
4. $n/2$

In building an ANN model, from the input layer to the hidden layer, the transfer function applied in this study is hyperbolic tangent sigmoid. While from the hidden layer to the output layer, a linear function is applied. Additionally, the ANN network is trained using Levenberg-Marquardt which is a supervised learning algorithm

For this research, a modified EMD ANN is proposed. The development of modified hybrid EMD ANN model in this study begins by summing up the residue with the component exhibiting the lowest frequency (the last IMF). This new component is labeled as 'Decomposition 1'. In the next step, all the IMFs in addition to Decomposition 1 are preprocessed individually by arranging it into lags. Then input selections are done on each IMFs and Decomposition 1 to extract significant inputs. Each component is forecasted using ANN model, and the final results of the components are aggregated. To create the next modified EMD ANN model, the second lowest frequency (second last IMF) is summed with Decomposition 1 (aggregation of the residue with the last IMF). This component is then labeled as 'Decomposition 2'. The whole processes are repeated in the same manner, and the best model is chosen. Based on equation (3), the number of modified EMD ANN models created for Singapore tourist arrivals series is six as shown in Fig. 7. Meanwhile, the number of modified EMD ANN model created for Indonesia arrival series is five as shown in Fig. 8. Because this experiment is done through trial and error method, the IMFs are summed up sequentially. Nevertheless, it is believed that any other combinations of IMFs are possible.

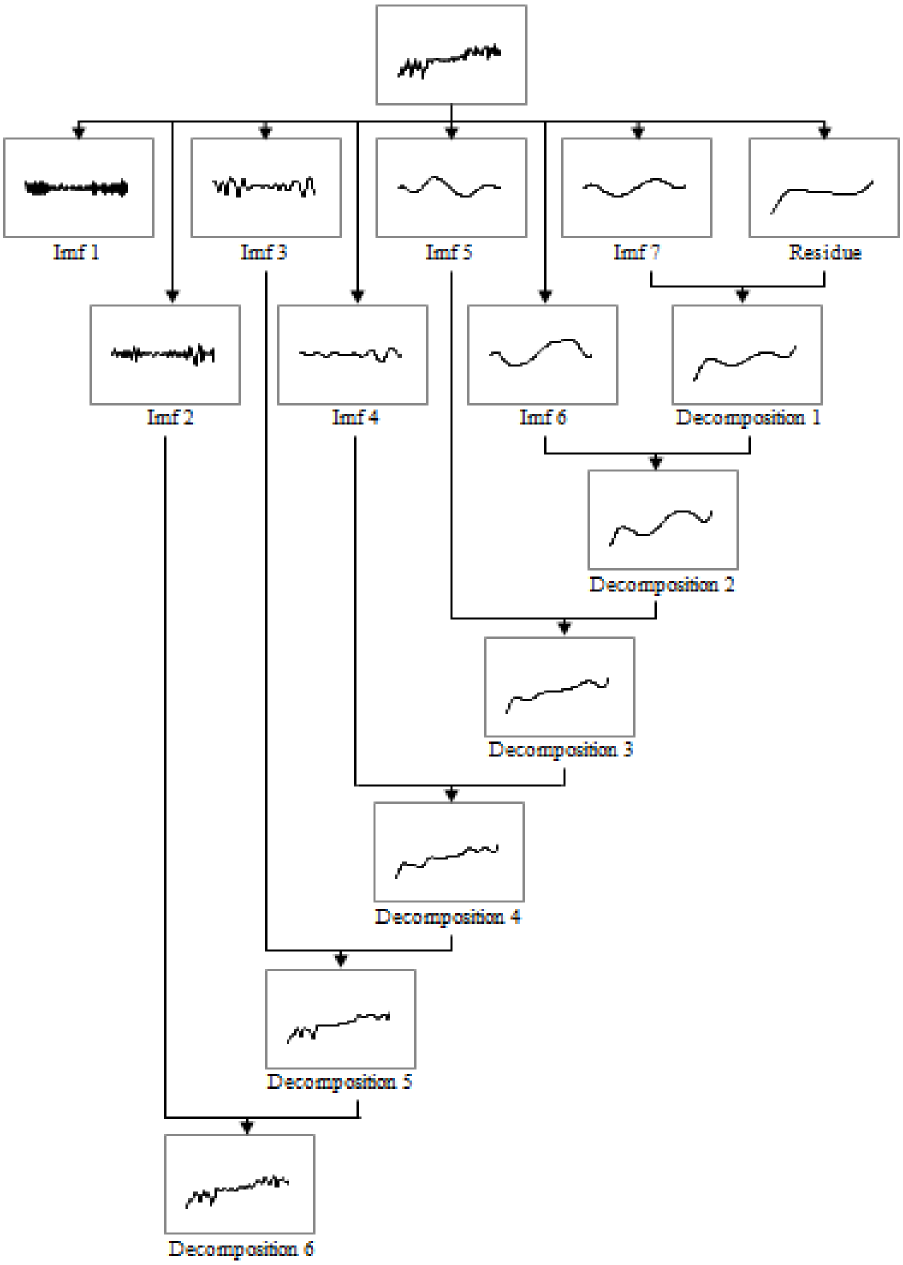

Fig. 7: New decompositions based on IMF and residual of Singapore data

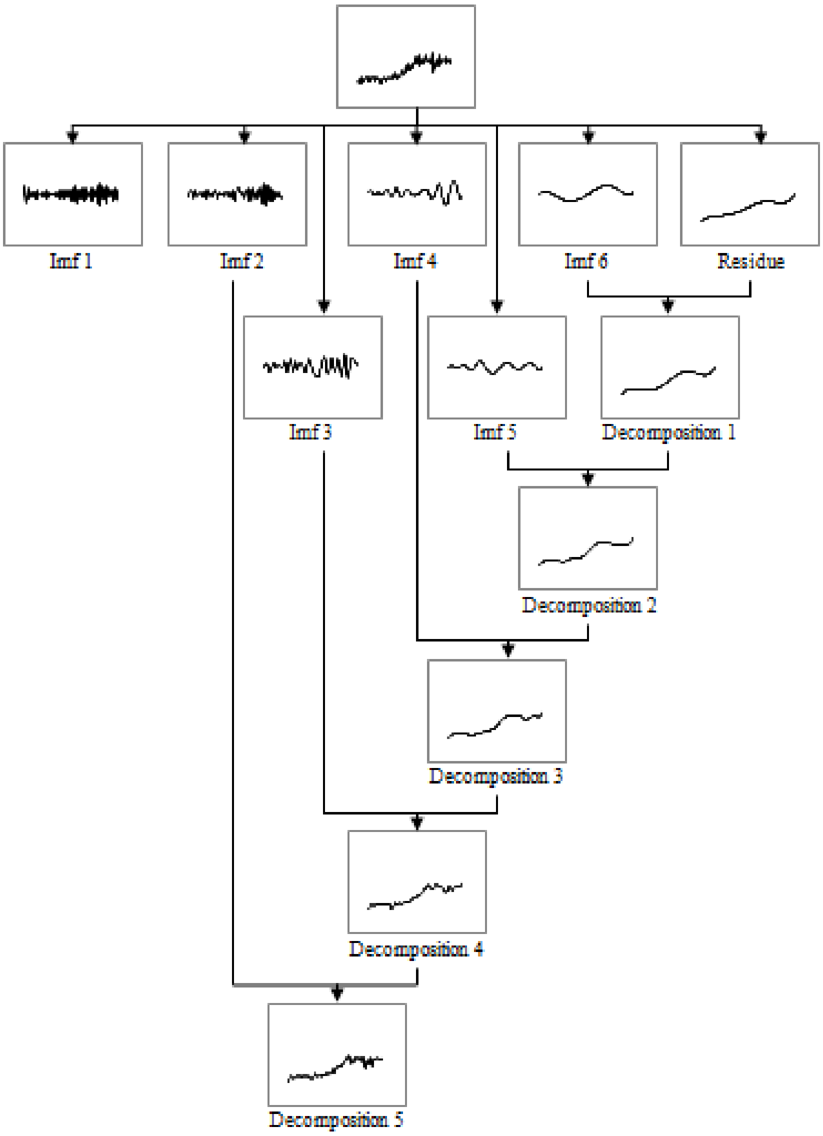

Fig. 8: New decompositions based on IMF and residual of Indonesia data

To evaluate the precision of the proposed model, individual ANN and hybrid EMD-ANN models are built to serve as benchmark models for comparison purposes. The forecasting accuracy of the model is calculated by comparing the predicted values with the actual values using the standard performance measurement, RMSE and MAPE. It should be noted that the best model is chosen based on the smaller values of RMSE and MAPE. The performance results for the tourist arrivals from Singapore and Indonesia based on the selected forecasting models are as shown in Table 2. It can be seen that EMD significantly improves the RMSE and MAPE results of an individual ANN model by 36.5% and 37.23% for Singapore, 43.9% and 34.3% for Indonesia.

As shown in Table 2, for the monthly tourist arrivals from Singapore, the modified EMD ANN model (Decomposition 2) achieved the best RMSE and MAPE of 75491.53 and 5.87% respectively. This indicated that the modified EMD-ANN for Decomposition 2 managed to slightly improve the results of EMD-ANN by 4.5% and 4.9% respectively. As seen in Fig. 9, the modified model using Decomposition 1 and Decomposition 5 also improved the accuracy of hybrid EMD ANN even though the improvements are quite subtle, that is by 2.0% and 1.5% (Decomposition 1), 2.5% and 1.3% (Decomposition 5). Regarding accuracy, Modified EMD ANN using Decomposition 2 performed best, followed by Decomposition 1 and 5, and hybrid EMD ANN. Based on Table 2, it is noticeable that the individual ANN model ranked last compared to the other models.

As for the monthly tourist arrivals from Indonesia, the modified EMD ANN model (Decomposition 4) produced the best forecast, with 11562.16 for RMSE, and 5.05% for MAPE. In detail, the modified model using Decomposition 4 boosted the accuracy of EMD-ANN by 6.3% and 8% for RMSE and MAPE respectively. According to Fig. 10, it is also evident that modified EMD ANN model using Decomposition 3 slightly improved the results of EMD-ANN by 2.2% and 2.6%. Overall, it can be seen that modified EMD ANN (Decomposition 4) performed the best, followed by Decomposition 3 and hybrid EMD ANN in terms of precision. Similar to Singapore arrival series data, the single ANN model performed the worst in this study.

Table 2: Performance of the forecasting methods

	Singapore		Indonesia	
	RMSE	MAPE	RMSE	MAPE
ANN	124515.00	9.83	21975.90	8.36
EMD-ANN	79090.02	6.17	12333.80	5.49
Modified EMD-ANN	75491.53	5.87	11562.16	5.05

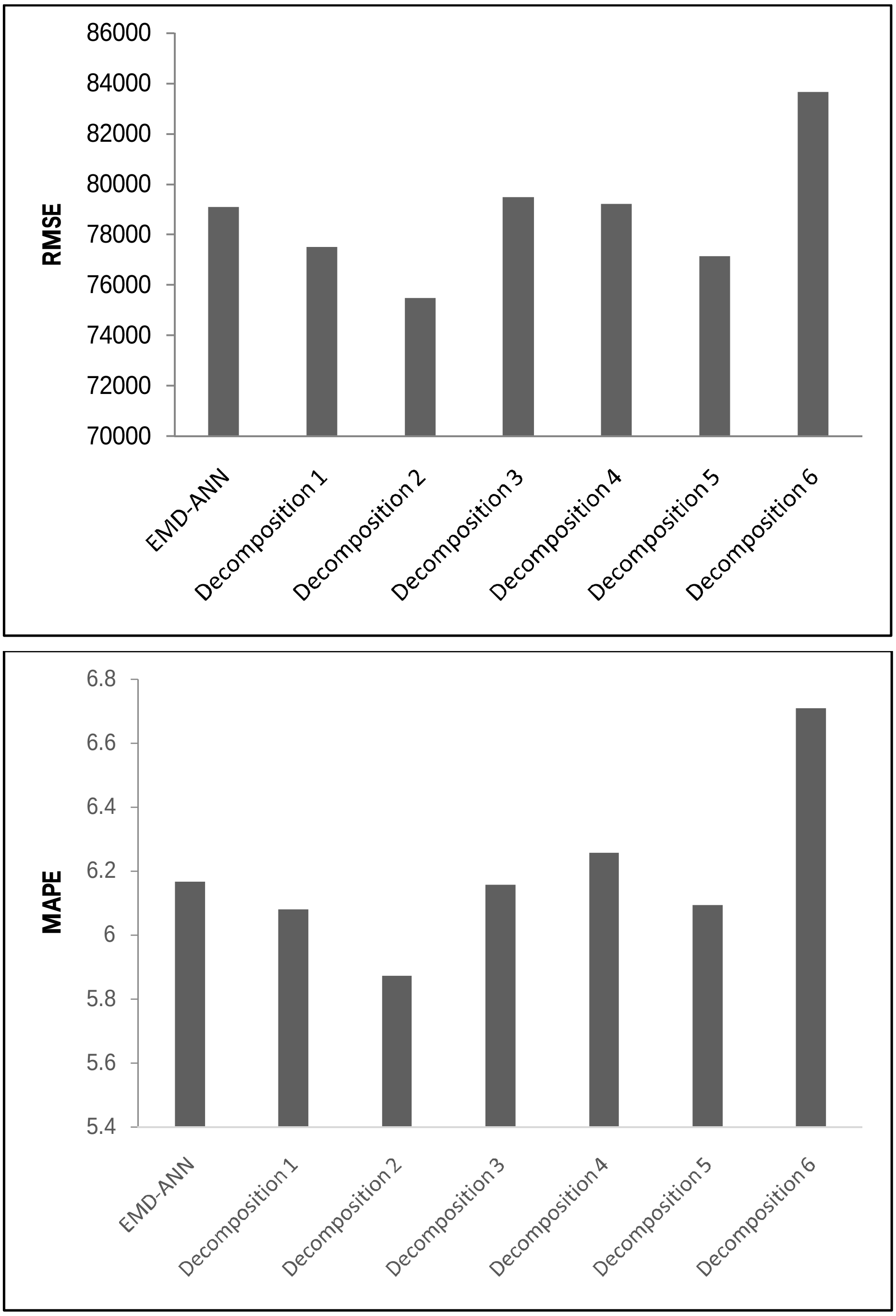

Fig. 9: Performance comparison between EMD-ANN and the new decompositions for Singapore data

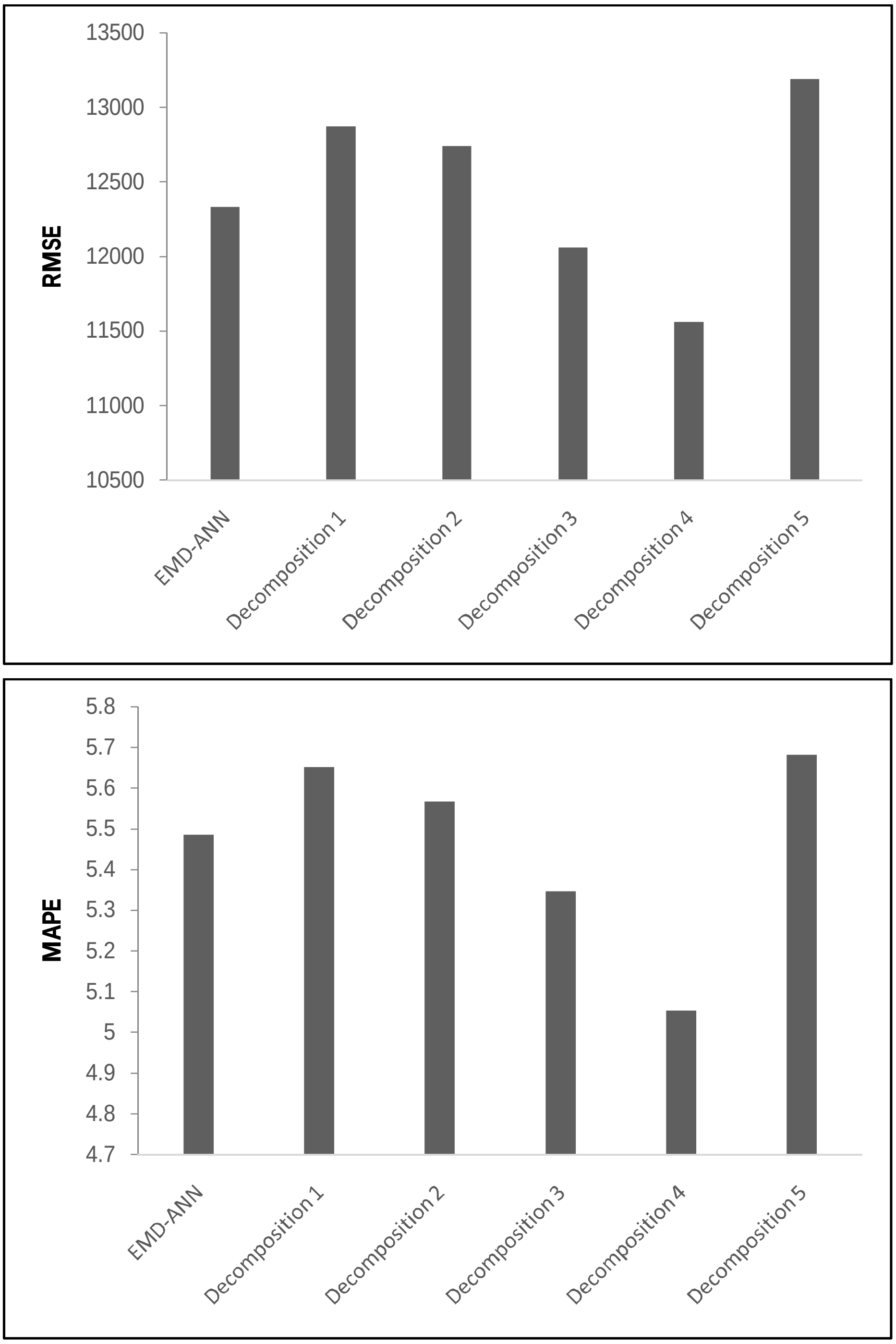

Fig. 10: Performance comparison between EMD-ANN and the new decompositions for Indonesia data

Based on the empirical results, it is clear that using EMD to pre-process the time series data will enhance the accuracy of the forecast generated by single ANN model because it reduces the complexities of the data, making it simpler for the modelling of a forecast. As shown in the results, implementations of modified

EMD technique have the potential to further improve the accuracy of the predictions, depending on the combinations of IMFs. In other words, if correct combinations are used, the model will produce results better than the conventional EMD technique. In contrast, if unsuitable combinations are used, the results generated will deteriorate. Hence, the future challenge lies in selecting the best combinations for the modified EMD technique.

4 Conclusion

The rapid growth of the tourism industry worldwide has caused new approaches for big data tourism demand forecasting to be explored extensively. The accuracy of the models in anticipating tourist arrivals is essential for effective policy planning. To improve the precision of the forecast, this study proposes a modified hybrid forecasting model based on empirical mode decomposition (EMD) and ANN using monthly tourist arrivals from Singapore and Indonesia to Malaysia. Individual ANN and hybrid EMD ANN model are used in this research as benchmark models for comparison purposes. The empirical results showed that the selected modified EMD ANN model slightly improved the result generated by the conventional hybrid EMD ANN model for both Singapore and Indonesia tourism data series. Therefore, it can be said that the proposed model is a promising tool for forecasting big data series provided that the best set of combinations are used.

ACKNOWLEDGEMENTS
This research is supported by Universiti Teknologi Malaysia under grant number R.J130000.7826.4F681.

References

[1] Munan and Heidi. 2002. Malaysia, New York: Benchmark Books.

[2] Chareyron, G., J. Da Rugna, and T. Raimbault. 2014. Big data: A new challenge for tourism. In *International Conference on Big Data*, pp. 5-7.

[3] Davenport, T.H. 2013. At the Big Data Crossroads: turning towards a smarter travel experience. Amadeus IT Group.

[4] Sagiroglu, S. and D. Sinanc. 2013. Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference*, pp. 42-47.

[5] Burger, C.J.S.C, Dohnal, M., Kathrada, M. and Law, R. 2001. A practitioners guide to time-series methods for tourism demand forecasting—a case study of Durban, South Africa. *Tourism management*, 22(4), pp. 403-409.

[6] Cang, S. 2014. A Comparative Analysis of Three Types of Tourism Demand Forecasting Models: Individual, Linear Combination and Non-linear Combination. *International Journal of Tourism Research*, 16(6), pp. 596-607.

[7] Cho, V. 2001. Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality & Tourism Research*, 25(4), pp. 399-420.

[8] Cho, V. 2003. A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, 24(3), pp. 323-330.

[9] Chen, C.F., Lai, M.C. and Yeh, C.C. 2012. Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, 26, pp. 281-287.

[10]Chen, K.Y. 2011. Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38(8), pp. 10368-10376.

[11]Claveria, O. and Torra, S. 2014. Forecasting tourism demand to Catalonia: neural networks vs. time series models. *Economic Modelling*, 36, pp. 220-228.

[12]Claveria, O., Monte, E. and Torra, S. 2015. Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*, pp. 1-4.

[13]Kon, S.C. and Turner, L.W. 2005. Neural network forecasting of tourism demand. *Tourism Economics*, 11(3), pp. 301-328.

[14]Pai, P.F., Hung, K.C. and Lin, K.P. 2014. Tourism demand forecasting using novel hybrid system. *Expert Systems with Applications*, 41(8), pp. 3691-3702.

[15]Palmer, A., Montano, J.J. and Sesé, A. 2006. Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27(5), pp. 781-790.

[16]Zhang, G.P. and Qi, M. 2005. Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 160(2), pp. 501-514.

[17]Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 454, No. 1971, pp. 903-995. The Royal Society.

[18]Guo, Z., Zhao, W., Lu, H. and Wang, J. 2012. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renewable Energy*, 37(1), pp. 241-249.

[19]Lin, C.S., Chiu, S.H. and Lin, T.Y. 2012. Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling*, 29(6), pp. 2583-2590.

[20]Sang, Y.F., Wang, Z. and Liu, C. 2014. Comparison of the MK test and EMD method for trend identification in hydrological time series. *Journal of Hydrology*, 510, pp. 293-298.

[21]Shabri, A. 2015. A novel hybrid ensemble learning paradigm for tourism forecasting. In *THE 2ND ISM INTERNATIONAL STATISTICAL CONFERENCE 2014 (ISM-II): Empowering the Applications of Statistical and Mathematical Sciences*. AIP Publishing.

[22]Wang, X.L. and Li, H. 2010. Short-term wind speed forecasting of wind farm based on EMD and LS-SVM. *Computer Engineering and Design*, 31(10), pp. 2303-2307.

[23]Yu, L., Wang, S. and Lai, K.K. 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30(5), pp. 2623-2635.

[24]MALAYSIA TOURISM PROMOTION BOARD,. *MALAYSIA REGISTERS 25.7 MILLION TOURISTS IN 2015*, 2016.Web. 10 July 2016.

[25]Zhang, G., Patuwo, B.E. and Hu, M.Y. 1998. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1), pp. 35-62.

[26]Hu, J., Wang, J. and Zeng, G. 2013. A hybrid forecasting approach applied to wind speed time series. *Renewable Energy*, 60, pp. 185-194.