International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-8 Issue-6S3, September 2019

Forecasting ASEAN Tourist Arrivals in Malaysia using Different Time Series Models

A. Rafidah, Ernie Mazuin, Ani Shabri

Abstract: In this study three time series models are used for forecasting monthly ASEAN tourist arrivals in Malaysia from January 1999 to December 2015. Brunei, Thailand and Vietnam of ASEAN country selected as case study. This paper compares the forecasting accuracy of seasonal autoregressive integrated moving average (SARIMA), Support Vector Machine (SVM) and Wavelet Support Vector Machine (WSVM) and Empirical Mode Decomposition with Wavelet Support Vector Machine (EMD_WSVM) using root mean square error (RMSE) and mean absolute percentage error (MAPE) criterion. Moreover, correlation test has also been carried out to strengthen decisions, and to check accuracy of various forecasting models. Based on the forecasting performance of all four models, hybrid model SARIMA and EMD_WSVM are found to be best models as compare to single model SVM and hybrid model WSVM.

Index Terms: Forecasting; tourist arrivals; SARIMA model; SVM model; WSVM model and EMD_WSVM model.

I. INTRODUCTION

Malaysia is a significant ASEAN tourist destination, and tourism is a significant source of profit for a wide range of activities. According to the Malaysia Ministry of Tourism in 2012 there were 1 billion of tourist's nights registered and in January 2013, international tourist arrivals exceeded 1 billion for the first time ever in 2012, reaching a total of 1.035 billion tourists, 39 million more than 2011. Since 2007 the number of foreign tourist nights reveals as lightly upward trend with a growth rate of 4.5%. Foreign tourists account for a total of 85% of the total Malaysia tourist activity. The World Tourism Organization (WTO) predicted that there will be 1.6 billion international tourist arrivals worldwide by 2020 and that these tourists are also expected to spend over two trillion US dollars (WTO Report, 2012). Therefore, more accurate and reliable prediction of tourism demand is required for efficiently and accurately estimation of ASEAN tourist arrival which in turn assures the future planning of infrastructure in tourism sector.

From the past several decades, study on tourism forecasting has attracted much of attention of the researchers across the world, for instance [1] studied the forecasting tourist demand in Croatia using various forecasting approaches.

According to [2] studied the forecasting tourist arrivals on economic impact. Recently, [3][4][5] compared the hybrid forecasting models in predicting the tourist arrivals in Singapore.

Since component of tourism demand is seasonality, researchers found that among the competing time series models, SARIMA found to be more appropriate model for non-seasonal and seasonal time series forecasting [6] and [7][8]. Additionally, SARIMA model is capability to deal with both stationary and non-stationary series. However, demonstrates that SARIMA model is the best one for forecasting foreign tourist arrivals to India. Among different time series forecasting models, SVM has also gained more popularity for forecasting seasonal time series [9]. SVM is frequently adopted in empirical studies, because it often outperforms many time series models [10]. SARIMA and SVM are seems to be more accurate models for seasonal tourist arrivals prediction.

The standard SVM is solved using quadratic programming methods. However, this method is often time consuming and has higher computational burden because of the required constrained optimization programming and the application of SVMs for tourist forecasting has not been widely explored and SVM method also challenging to estimate qualitatively how many time steps into the past would allow the greatest efficiency, the values of time for tourist arrival are not known in advance[11]. To the best of our knowledge there is not much of literature available on the study of tourism demand in Malaysia. Though there are few, like [11] showed that SARIMA is the best model to forecast international tourists in Malaysia under RMSE and MAPE criterion. In [12][13][14] the author developed a hybrid model wavelet and empirical mode decomposition (EMD) of in different area using SVM model analysis.

Therefore, there is a lot of scope to study the forecasting of tourism demand in Malaysia using different types of forecasting models. On this note, we suggest studying the performance of SARIMA, SVM, WSVM and EMD_WSVM model under different comparison criteria to forecast ASEAN tourist arrivals in Malaysia in the coming years. Fig. 1-3 gives an overview of the number of three ASEAN tourist arrivals to Malaysia.

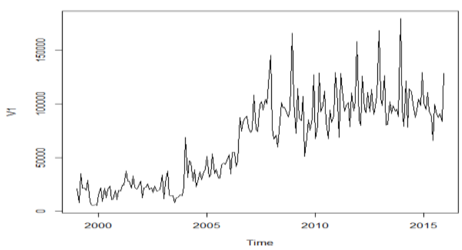

Fig. 1. . Monthly Brunei tourist arrivals (Jan 1999 to Dec 2015)

Revised Manuscript Received on September 22, 2019.

A. Rafidah, Lecturer: Technical Foundation Department, Universiti Kuala Lumpur, Bandar Seri Alam, johor Bahru Malaysia. rafidahali@unikl.edu.my.

Ernie Mazuin, Lecturer: Instrumentation and Control Engineering Department, Universiti Kuala Lumpur, Bandar Seri Alam, Johor Bahru Malaysia.

Ani Shabri, Associate Prof: Mathematics Department, Universiti Technology Malaysia, Skudai Johor Bahru, Malaysia.

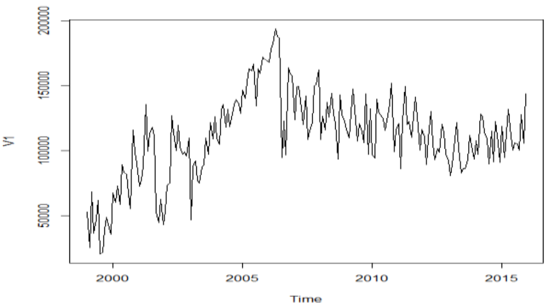

Fig. 2. Monthly Thailand tourist arrivals (Jan 1999 to Dec 2015)

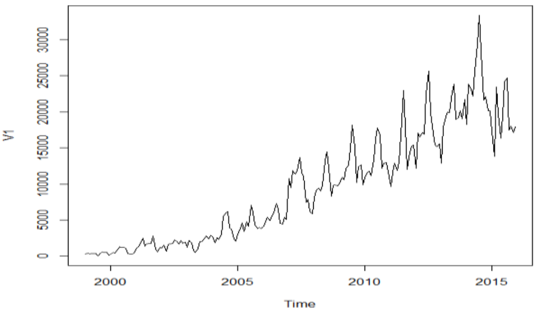

Fig. 3. Monthly Vietnam tourist arrivals (Jan 1999 to Dec 2015)

II. FORECASTING MODELS

A. Single Seasonal autoregressive integrated moving average (SARIMA) Model

The SARIMA model was first developed by Box-Jenkins approach in [15]. It is used when the time series data contain seasonal elements. The SARIMA model has been created from both seasonal and non-seasonal autoregressive (AR), the moving average (MA) and the combination of AR and MA models. It is usually represented by multiplicative model in the form of SARIMA (p,d,q)(P,D,Q)s, where p, P indicates the non-seasonal order of AR and q, Q denote the seasonal order of MA model with non-seasonal difference (d) and seasonal difference (D). The best fitted SARIMA model is selected using minimum Akaike Information Criterion (AIC). Mathematical formulation of AIC is defined as

$$AIC = \log\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right) + \frac{2p}{n} \qquad (1.1)$$

where p the number of parameters, the sum square error and n the periods of data (Akaike, 1974).

B. Single Support vector machine (SVM) Model

Support Vector Machine (SVM) is a new general machine learning method proposed by Vapnik (1995) [16]. SVM realize structural risk minimization of the data sample from the approximation accuracy and the approximating function. This method also has perfect theory in solving many real problems and can construct function in many kinds of function set [17]. Other than that, SVM is widely applied to the speech recognition, pattern recognition, analysis of time series, bioinformatics and economics, achieves some results in the aforesaid fields ([18]

According to Cortes and Vapnik‟s (1995), SVM represent the relationship between an output y and a set of inputs in the form:

$$y(x) = \sum_{i=1}^{n} w_i \phi(x) + b \qquad (1.2)$$

where $\phi(x)$ represent the higher dimensional feature space, which is nonlinearly mapped the input space *x*.

In SVM for function estimation, the estimation by minimizing regularized risk function:

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} L_\varepsilon(y_i) \qquad (1.3)$$

is an arbitrary penalty parameter called the regularization constant.

Basically, SVM penalize $f(x_i)$ when it departures from y_i by means of an ε-insensitive loss function:

$$L_\varepsilon(y_i) = \begin{cases} 0 & \text{if } |f(x_i) - y_i| < \varepsilon \\ |f(x_i) - y_i| - \varepsilon & \text{otherwise} \end{cases} \qquad (1.4)$$

The minimization of expression (2.3) is implemented by introducing the slack variable ξ_i^- and ξ_i^+ .Specifically, ε-Support Vector Regression (ε-SVR) solves the following quadratic programming problem:

$$\min_{\omega, b, \xi_i^-, \xi_i^+} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\left(\xi_i^- + \xi_i^+\right) \qquad (1.5)$$

Subject to

$$y_i - (\omega'\phi(x_i) + b) \le \varepsilon + \xi_i^-$$

$$(\omega'\phi(x_i) + b) - y_i \le \varepsilon + \xi_i^+$$

$$\forall i, \xi_i^- \text{ and } \xi_i^+ \ge 0$$

The solution to this minimization problem is of the form

$$f(x) = \sum_{i=1}^{m}(\lambda_i - \lambda_i^*)K(x_i, x) + b \qquad (1.6)$$

Where λ_i and λ_i^* are the Langrage multipliers associated with the constrains $y_i - (\omega'\phi(x_i) + b) \le \varepsilon + \xi_i^-$ and $(\omega'\phi(x_i) + b) - y_i \le \varepsilon + \xi_i^+$ respectively.

There are number of kernels that can be used in Support Vector Machine models. In this study, the type of kernel that is focused on is the radial basis function (RBF). RFB kernel function is expressed as:

$$k(x_i, x_j) = \exp(\gamma\|x_i - x_j\|^2) \qquad (1.7)$$

RBF is chosen in this study because it ejects good performance [19]. This kernel function is by far the most popular choice of kernel types used in support vector machines. nonetheless, RBF is generally used kernel function because its nonlinearity mapped samples into a higher dimensional space, and it has less hyper parameters and less numerical computation. Durgesh and Lekha (2009) stated that in selecting parameter for the model, the search method is often used in cross validation to select the best parameter to the training data set and then obtain the classifier which is then used to classify the testing data set in order to get the accuracy.

The kernel function represents a dot product of input data points mapped into the higher dimensional feature space by transformation ϕ.

C. Wavelet Support Vector Machine (WSVM)

In recent years, wavelet transform has shown a very convincing result in multi resolution analysis as well as many other functions [20]. In this study, WSVM model is obtained by combining two methods, discrete wavelet transpose (DWT) and SVM [21]. The WSVM model is SVM model that uses sub time series components obtained using DWT on original data. For the WSVM model inputs, the original time series is decomposed into a certain number of sub time series components (D's) by Mallat DWT algorithm [22]. Each component plays a different role in the original time series and the behaviour of each sub time series is distinct [23].

The discrete wavelet transforms (DWT) can be derived by discretizing Equation (1.8), where a and b are two parameters given as follows:

$$a = a_a^m , \quad b = na_0^m b_0 \tag{1.8}$$

The variable n and m is an integers. Substituting a and b in Equation (1.9) results in:

$$W_x(m,n,\psi) = a_0^{-m/2}\int_{-\infty}^{+\infty} f(t)\psi(a_0^{-m}t - nb_0)\,d \tag{1.9}$$

Equation (1.9) is the mathematical relation of discrete wavelet transform (DWT).

In this study wavelet analysis is used to decompose the time series of tourist arrival data into various components. The original time series are decomposed into various details (Ds) and an approximation (As) at different resolution levels using DWT or simple format as:

$$W_x = A_M(t) + \sum_{m=1}^{M} D_m(t) \tag{1.10}$$

which $A_M(t)$ is called the approximation sub series or residual term at levels M and $D_m(t)$ where m=1, 2…, M are detail sub-series which can capture small features of interpretational value in data. Subsequently the decomposed components can therefore be utilized as inputs for the SVM model.

D. Hybrid Empirical Mode Decomposition (EMD) with WSVM (EMD_WSVM)

As an adaptive method used for signal analysis, [23] proposed EMD. An EMD is a form of an adaptive time series decomposition technique using Hilbert-Huang transform (HHT), which is a combination of EMD and Hilbert Transform (HT) and is specifically designed to analyses the nonlinear and non-stationary time series data. It does not require a priori knowledge about the data as it is a fully data driven [24].

A combination of EMD, wavelet and the SVM model will provides an effective way to improve the prediction accuracy for nonlinear and non-stationary time series. The algorithm flow of EMD is as follows:

$$x(t) = \sum_{i=1}^{n} c_j(t) + r_n(t) \tag{1.11}$$

where Thus, residue $r_n(t)$ is the mean trend of $x(t)$. The IMFs $c_1 \ldots c_n$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and they change with the variation of time series $x(t)$, while $r_n(t)$ represents the central tendency of time series $x(t)$.

The procedures of EMD_WSVM model consist of three main steps which are (i) decomposition EMD stage, (ii) decomposition wavelet and individual forecasting stage, and (iii) ensemble forecasting stage.

III. FORECAST PERFORMANCE

The performance of each forecasting model for both training and testing sets is evaluated by using root mean square error (RMSE) and mean percentage error (MAPE) is used to analyses the performance of each proposed forecasting models. These means of evaluation are widely used in evaluating results of time series forecasting [25]. The definition of RMSE and MAPE are shown below:

RMSE is a quadratic scoring rule which measures the average magnitude of the error. The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a variable and not between variables, as it is scale-dependent. RMSE were written as;

$$\text{RMSE} = \sqrt{\frac{\sum_t^n e^2{}_t}{n}} \tag{2.1}$$

In statistical test MAPE is favoured and has gained popularity in the literature because it is not prone to change in the magnitude of time series data (Park, Yoon and Kim, 2000).

$$MAPE = \sum_{t=1}^{n}\left|\frac{e_t}{observed_t}\right| X \frac{10(}{n} \tag{2.2}$$

According to [1] "a rough scale for the accuracy of a model can be based on MAPE" following the suggestions gave in the Table I below.

TABLE I. ROUGH SCALE

MAPE	Forecasting accuracy
Less than 10%	Highly accurate
10-20%	Good
20-50%	Reasonable
Greater than 50%	Inaccurate

Source: (Baldigara, 2013)

IV. COMPARISON RESULTS

This section will discuss each of the result forecasting model SARIMA, SVM, WSVM and EMD_WSVM.

A. Estimated SARIMA Model

Based on these characteristics autocorrelation function (ACF) and partial autocorrelation function (PACF), Table 2 shows possible models that could satisfy this situation. The AIC value for each ASEAN country in SARIMA model are illustrated in the Table 2 for each data dataset.

TABLE II. RESULTS OF SARIMA MODEL

Country	SARIMA model	AIC value
	(3,1,3) (1,0,0)[12]	3146.39
Brunei	(3,1,3) (3,1,1)[12]	2908.52
	(2,1,0) (3,1,1)[12]	2911.33
	(2,1,1)(2,1,0)[12]	3004.84
Thailand	(3,2,3)(1,1,1)[12]	2984.6
	(4,2,3)(2,1,2)[12]	2987.88
	(1,1,1)(0,1,0)[12]	2419.91
Vietnam	(1,1,1)(1,1,1)[12]	2223.01
	(3,1,1)(3,1,1)[12]	2220.86

The best fitted model is selected based on minimum Akaike Information Criterion (AIC). Table 2 describes the SARIMA models for the forecast ASEAN tourist arrivals series. Further, from the Table II the best fitted model for Brunei, Thailand and Vietnam are (3,1,3)(1,0,0)[12], (3,2,3)(1,1,1)[12] and (3,1,1)(3,1,1)[12] respectively according to the minimum AIC criterion.

B. Estimated SVM Model

Forecasting using SVM model the first phase is needing to determine input lag model variable. According to [26] the input lag can determine by using PACF model. The Fig.4 represented the number of input lag variables chosen based on PACF value of three country ASEAN tourist arrival.

These model inputs also used in the cases of WSVM and EMD_WSVM models. Table III show the result three ASEAN countries using SVM model.

From the table we can see that the best model for Brunei is input lag 6 (SVM6) base on RMSE value but the MAPE value highest than input lag 3 (SVM3). Since the comparison slightly low, and this value MAPE indicate as good model. Therefore, input lag 6 was selected as the best model SVM for Brunei. As clearly show in the table that the best model for Thailand and Vietnam the best model is input lag 11 and 12 respectively according the lowest value of RMSE and MAPE.

Data set	Input Lag	SVM Input Structure
Brunei	1	$y_t = f(x_{t-1})$
	2	$y_t = f(x_{t-1}, x_{t-2})$
	3	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3})$
	6	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6})$
	9	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9})$
	12	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}, x_{t-11}, x_{t-12})$
Thailand	1	$y_t = f(x_{t-1})$
	2	$y_t = f(x_{t-1}, x_{t-2})$
	12	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}, x_{t-11})$
Vietnam	1	$y_t = f(x_{t-1})$
	3	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3})$
	7	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7})$
	11	$y_t = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}, x_{t-11})$

Fig. 4. Input Lag SVM, WSVM and EMD_WSVM

TABLE III. RESULTS OF SVM MODEL

Country	Input Lag	Measures	
		RMSE	MAPE (%)
Brunei	SVM1	27913.5047	18.74
	SVM2	24489.4751	15.28
	SVM3	24509.2058	14.65
	SVM6	23608.6639	15.47
	SVM9	23236.0573	18.16
	SVM12	25432.0357	19.00
Thailand	SVM1	17074.8282	13.70
	SVM2	17508.7745	13.35
	SVM12	15388.7186	11.87
Vietnam	SVM1	10726.182	47.37
	SVM3	10535.7521	46.76
	SVM7	10161.6451	44.74
	SVM11	10094.8003	44.30

C. Estimated WSVM Model

Wavelet decomposition decomposes the data into approximate scale and details scale. The approximate scale is also known as the high scale (low frequency) and the details scale are known as the low scale (high frequency). The number of details components depend on the resolution levels implemented.

For this study, two decomposition levels (D1 and D2) and one approximation (A2) are used for each ASEAN tourist arrival dataset. The decomposition is successful when filtering process able to decompose the input variable into approximation and decomposition component. Fig. 5-7 show the original tourist arrivals data time and their Ds decomposition using wavelet method each ASEAN country.

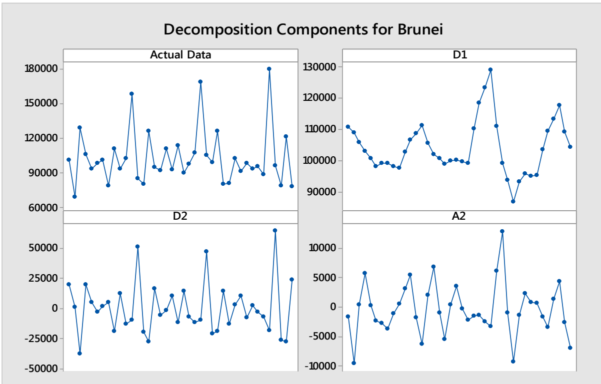

Fig. 5. Decomposed wavelet sub-series components (Ds) of tourist arrival data of Brunei

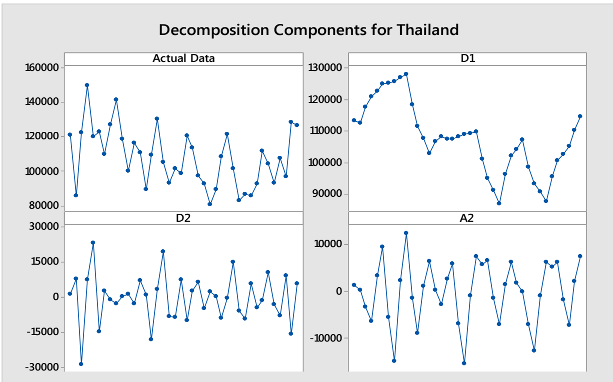

Fig. 6. Decomposed wavelet sub-series components (Ds) of tourist arrival data of Thailand

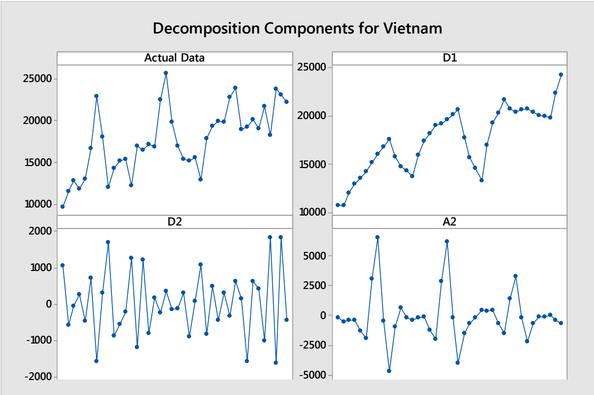

Fig. 7. Decomposed wavelet sub-series components (Ds) of tourist arrival data of Vietnam

The input variables on model performance is same as Fig.4 (in SVM model). The result of WSVM for each ASEAN tourist arrivals in the test period present in Table IV. From the table, the selection of WSVM model based on smallest RMSE, MAE and MAPE with highest correlation value.

The performance results tourist arrivals from Vietnam the best model is input lag 1(WSVM1). Meanwhile, Brunei and Thailand the best model is input lag 2(WSVM2) with the least value RMSE (valued at 23205.05 and 11891.62), MAE (valued at 16062.49 and 9135.07), MAPE (valued at 14.71% and 8.58%) and highest correlation (valued at 0.7391 and 0.7391) respectively.

TABLE IV. RESULTS OF WSVM MODEL

Country	Input Lag	Measures	
		RMSE	MAPE (%)
Brunei	WSVM1	25817.6414	16.01
	WSVM2	23205.0548	14.71
	WSVM3	28626.8633	19.12
	WSVM6	27220.9942	18.68
	WSVM9	25344.6132	17.97
	WSVM12	27570.4409	18.80
Thailand	WSVM1	13584.7586	10.14
	WSVM2	11891.6177	8.58
	WSVM12	13070.2048	10.12
Vietnam	WSVM1	10303.529	41.53
	WSVM3	10233.1059	42.65
	WSVM7	10699.6945	46.74
	WSVM11	10249.9278	44.44

D. Estimated EMD_WSVM Model

The model EMD_WSVM is combination between EMD with WSVM. In this model the first stage, the input of data reconstructs into IMFs components and approximation component. Figure 8-10 demonstrate the IMFs and residue components of ASEAN tourist arrival data time series.

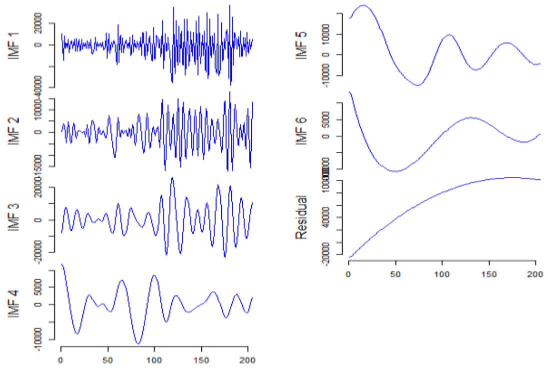

Fig. 8. The IMFs and Residue Components of tourist arrival data of Brunei

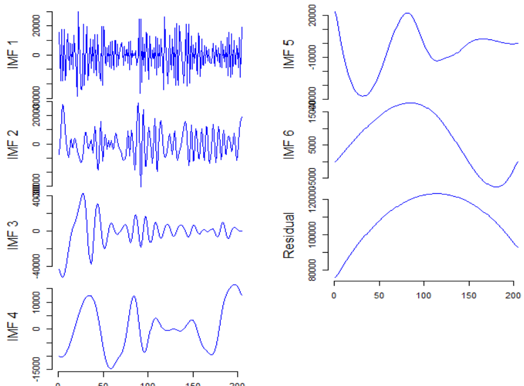

Fig. 9. The IMFs and Residue Components of tourist arrival data of Thailand

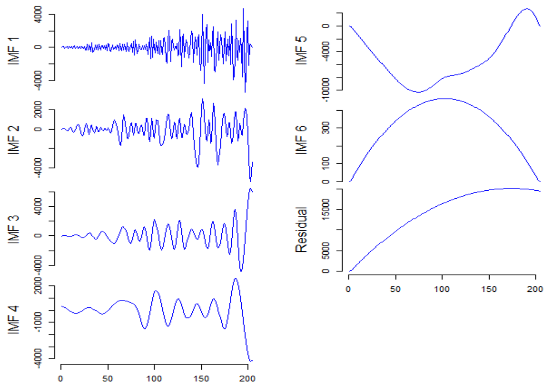

Fig. 10. The IMFs and Residue Components of tourist arrival data of Vietnam

Second stage, each of the IMFs component with residual will be decompose into wavelet method. then the last stage, forecasted values all extracted SVM model are summed together to produce the final forecasting for the original time series. Table 6 represented the result EMD_WSVM. The input lag of each model same as SVM input lag in Fig 4.

From the Table V illustrated that the best model for Brunei is input lag 2 (RMSE=20486.05, MAPE=11.32%), Thailand is input lag 12 (RMSE=10711.56, MAPE=7.96%) and Vietnam is input lag 1(RMSE=3523.25, MAPE=13.85%). Base on MAPE value of Thailand country indicates that this model highly accurate, meanwhile another two country the model category as good model.

TABLE V. RESULTS OF EMD_WSVM

Country	Input Lag	Measures	
		RMSE	MAPE (%)
Brunei	1	23666.92	14.51
	2	20486.05	11.32
	3	21495.9	12.58
	6	21153.91	12.53
	9	20998.05	12.60
	12	20356.05	12.12
Thailand	1	13296	9.62
	2	12438.18	9.38
	12	10711.56	7.96
Vietnam	1	3152.845	13.37
	3	3344.864	14.07
	7	3616.818	15.43
	11	3713.171	15.99

V. DISCUSSION AND RESULTS

For further analysis, the error statistics of SARIMA, SVM, WSVM, and EMD_WSVM, were compared to each other in order to find the best model for river flow forecasting. Table VI compares the testing results among the four approaches and based on two statistical measurements, which are RMSE and MAPE

The result obtained for Brunei shows that the lowest RMSE and MAPE were acquired from EMD_WSVM model. Therefore, EMD_WSVM is declared as the best model representing Brunei, followed by WSVM, SVM and SARIMA. These models can be defined as second, third and fourth best approaches for Brunei. Table VI also indicates that the hybrid model EMD_WSVM and WSVM have outperformed the single SARIMA and SVM model.

Fig. 11 shows comparison of the observed and the best model predicted values of monthly tourist arrivals for Brunei. Most of the points in Fig. 11 are close to the observation data indicating good prediction capability.

Consequently, the result obtained for Thailand shows that the lowest RMSE and MAPE were observed from EMD_WSVM model. With these results, EMD_WSVM is declared as the best model representing Thailand. Meanwhile, the other models which are WSVM, SVM and SARIMA, can be defined as second, third, and fourth best approaches for Thailand. Fig. 12 show comparison of the observed and the predicted values of monthly tourist arrival for Thailand. Most of the points in Fig. 12 are close to the observation data indicating good prediction capability.

As for the case study in Vietnam, the results indicate that the lowest RMSE and MAPE, were obtained from SARIMA model. As produces the best results for almost all statistical measurements, therefore SARIMA is declared as the best model representing Vietnam. Meanwhile, the other models, which are EMD_WSVM, SVM, and WSVM models, can be defined as second, third and fourth best approaches for Vietnam. Fig. 13 shows the comparison of the observed and the predicted values of monthly river flow for Vietnam.

TABLE VI. FORECASTING RESULTS

Country	Model	Measures	
		RMSE	MAPE (%)
Brunei	SARIMA	1252.1026	21.62
	SVM	23608.6639	15.47
	WSVM	23205.0548	14.71
	EMD_WSVM	20486.05	11.32
Thailand	SARIMA	42564.5561	29.80
	SVM	15388.7186	11.87
	WSVM	11891.6177	8.58
	EMD_WSVM	10711.56	7.96
Vietnam	SARIMA	3034.7668	13.01
	SVM	10094.8003	44.30
	WSVM	10233.1059	42.65
	EMD_WSVM	3152.845	13.37

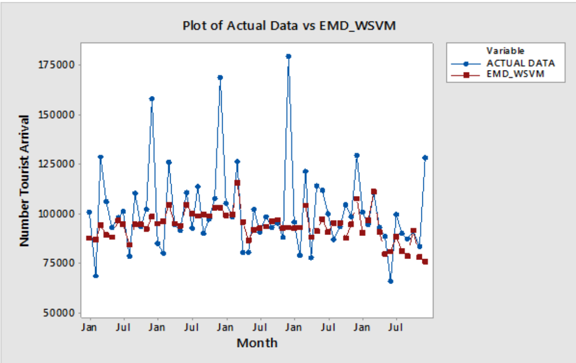

Fig. 11. Plot Actual Data vs EMD_WSVM model of Brunei dataset

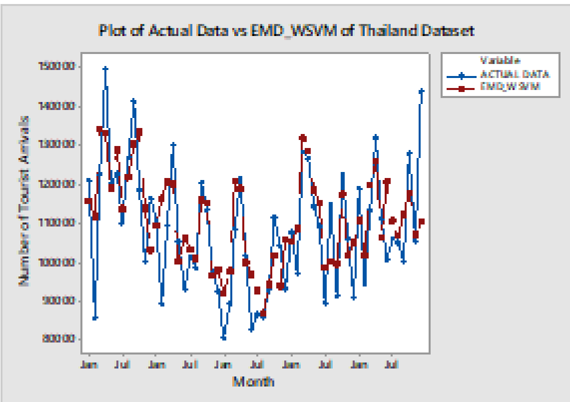

Fig. 12. Plot Actual Data vs EMD_WSVM model of Thailand dataset

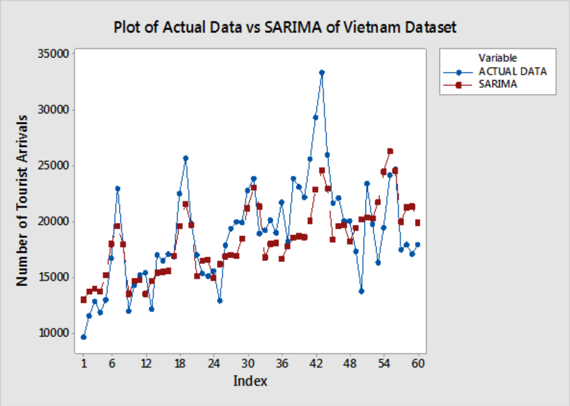

Fig. 13. Plot Actual Data vs SARIMA model of Vietnam dataset

VI. Conclusion

A correct and accurate tourist arrivals prediction plays a vital role in the making of successful development policy for the government. The current and future trend of tourist arrivals specifies the future of ASEAN tourism demand. Therefore, this research proposed a continuing future demand of foreign tourist arrivals in Malaysia. It can be summarized that the EMD_WSVM model is the best for most of the tourist arrivals except for Vietnam. However, SVM and WSVM models also show a promising method because the differences in the performance measurements between EMD_WSVM and SARIMA are relatively small

Acknowledgment

This research has been funded by (Grant HoRI) Universiti Kuala Lumpur, Malaysia.

References

[1] T. Baldigara, "Forecasting Tourism Demand in Croatia: A Comparison of Different Extrapolative Methods," J. Bus. Adm. Res., vol. 2, no. 1, 2013.
[2] H. Song, G. Li, S. F. Witt, and G. Athanasopoulos, "Forecasting tourist arrivals using time-varying parameter structural time series models," Int. J. Forecast., vol. 27, no. 3, pp. 855–869, 2011.
[3] N. A. Yahya, R. Samsudin, and A. Shabri, "Tourism forecasting using hybrid modified empirical mode decomposition and neural network," Int. J. Adv. Soft Comput. its Appl., vol. 9, no. 1, pp. 14–31, 2017.
[4] A. Rafidah, A. Shabri, A. Nurulhuda, and Y. Suhaila, "A Wavelet Support Vector Machine Combination Model for Singapore Tourist Arrival to Malaysia," in IOP Conference Series: Materials Science and Engineering, 2017, vol. 226, no. 1.
[5] A. Rafidah, A. Shabri, P. S. Ilmu, B. S. Alam, and S. Faculty, "HYBRID APPROACH OF AN EMPIRICAL MODE DECOMPOSITION AND," vol. 7, no. Xi, 2018.
[6] F. L. Chu, "Forecasting tourism demand in Asian-Pacific countries," Ann. Tour. Res., vol. 25, no. 3, pp. 597–615, 1998.
[7] N. Kulendran and K. K. F. Wong, "Modeling seasonality in tourism forecasting," J. Travel Res., vol. 44, no. 2, pp. 163–170, 2005.
[8] J. G. Brida and W. A. Risso, "Research note: Tourism demand forecasting with sarima models - The case of south tyrol," Tour. Econ., vol. 17, no. 1, pp. 209–221, 2011.
[9] L. Mei, "Tourism Demand Forecasting by Improved SVR Model," vol. 8, no. 5, pp. 403–412, 2015.
[10] O. Claveria, S. Torra, and E. Monte, "Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection," Rev. Econ. Apl., vol. 24, no. 72, pp. 109–132, 2016.
[11] L. Nanthakumar, T. Subramaniam, and M. Kogid, "Is 'Malaysia Truly Asia'? Forecasting tourism demand from ASEAN using SARIMA approach," Tourismos, vol. 7, no. 1, pp. 367–381, 2012.
[12] X. Liu, L. Bo, and H. Luo, "Bearing faults diagnostics based on hybrid LS-SVM and EMD method," Meas. J. Int. Meas. Confed., vol. 59, pp. 145–166, 2015.
[13] V. Bajaj and R. B. Pachori, "EEG signal classification using empirical mode decomposition and support vector machine," in Advances in Intelligent and Soft Computing, 2012, vol. 131 AISC, no. VOL. 2, pp. 623–635.
[14] I. Ebtehaj, H. Bonakdari, S. Shamshirband, and K. Mohammadi, "A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer," Flow Meas. Instrum., vol. 47, pp. 19–27, 2016.
[15] D. Asteriou and S. G. Hall, "ARIMA Models and the Box–Jenkins Methodology," Appl. Econom. (Second ed.). Palgrave MacMillan., pp. 265–286, 2011.
[16] V. V N, "Vapnik V.N. The nature of statistical learning theory.pdf," Journal of Contaminant Hydrology. 2010.
[17] H. Wang and D. Hu, "Comparison of SVM and LS-SVM for Regression," in Neural Networks and Brain, 2005. ICNN&B '05. International Conference on, 2005, vol. 1, pp. 279–283.
[18] L. Lai and J. Liu, "Support Vector Machine and Least Square Support Vector Machine Stock Forecasting Models," Comput. Sci. Inf. Technol., vol. 2, no. 1, pp. 30–39, 2014.
[19] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert-Huang transform and support vector machine," Appl. Acoust., vol. 75, no. 1, pp. 1–9, 2014.
[20] T. M. Choi, Y. Yu, and K. F. Au, "A hybrid SARIMA wavelet transform method for sales forecasting," Decis. Support Syst., vol. 51, no. 1, pp. 130–140, 2011.
[21] O. Kisi and M. Cimen, "A wavelet-support vector machine conjunction model for monthly streamflow forecasting," J. Hydrol., vol. 399, no. 1–2, pp. 132–140, 2011.
[22] J. Morlet, G. Arens, E. Fourgeau, and D. Glard, "Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media," Geophysics, vol. 47, no. 2, pp. 203–221, 1982.
[23] W. Wang and J. Ding, "Wavelet Network Model and Its Application to the Prediction of Hydrology," Forecast, vol. 1, no. 1, pp. 67–71, 2003.
[24] N. E. Huang et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. A Math. Phys. Eng. Sci., vol. 454, no. 1971, pp. 903–995, 1998.
[25] L. Olatomiwa, S. Mekhilef, S. Shamshirband, K. Mohammadi, D. Petković, and C. Sudheer, "A support vector machine-firefly algorithm-based model for global solar radiation prediction," Sol. Energy, vol. 115, pp. 632–644, 2015.
[26] A. Kang, Q. Tan, X. Yuan, X. Lei, and Y. Yuan, "Short-Term Wind Speed Prediction Using EEMD-LSSVM Model," Adv. Meteorol., vol. 2017, 2017.

Authors Profile

A. Rafidah, Lecturer: Technical Foundation Department, Universiti Kuala Lumpur, Bandar Seri Alam, johor Bahru Malaysia. rafidahali@unikl.edu.my.

Ernie Mazuin, Lecturer: Instrumentation and Control Engineering Department, Universiti Kuala Lumpur, Bandar Seri Alam, Johor Bahru Malaysia.

Ani Shabri, Associate Prof: Mathematics Department, Universiti Technology Malaysia, Skudai Johor Bahru, Malaysia.