International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

# Application of Empirical mode Decomposition with Wavelets Support Vector Machine in time Series Data

**A. Rafidah, Ani Shabri, Ernie Mazuin**

***Abstract*: This paper mainly discussed on the forecast of Thailand tourist visiting Malaysia. This paper proposed a three-stage technique in which the empirical mode decomposition (EMD) is combined with wavelet methods and support vector machine model. We used the proposed technique, EMD_WSVM to forecast two ASEAN country tourism timeseries. Detail experiments are conducted for the proposed method, in which there is a comparison between the EMD_WSVM, WSVM and SVM methods. The proposed EMD_WSVM model is determined to be dominant to the other methods in predicting the number of tourist arrivals.**

***Index Terms*: Forecasting, tourist arrivals, SVM model, WSVM model and EMD_WSVM model.**

## I. INTRODUCTION

Contemporary studies ascertained that tourist arrivals data typically follow the behavior of nonlinear and nonstationary. Hence, there is a difficulty in forecasting tourist arrivals by using the Box-Jenkins technique. Huang et al. [1] has studied the EMD as a tool powerful enough for tourist arrival quantitatively. It has also became the powerful modelling technique using statistical approach [2]. The nonlinear and nonstationary behaviors are capable to be handled by the EMD, consequently providing an alternative tool for researchers as well as practitioners. The explanation of time series data generation can be done by the EMD, through the splitting of signals of the time series. Based on scale separation, the splitting of signals is further divided into independent and intrinsic concrete implicational modes.

The decomposing of the components in this method is good and ensuring the stability of the components. Nevertheless, there is a clutter in the range of the frequency. This means that, the component which is having dissimilarity of amplitude and frequency existed in distinct IMFs. In addition, the orthogonality among IMFs is not good. Signals that are non-stationary and nonlinear can be dealt well with wavelet decomposition method [3].

**Revised Manuscript Received on January 30, 2020.**
***** Correspondence Author

**A.Rafidah***, Technical Foundation Department, University Kuala Lumpur, Pasir Gudang, Malaysia.

**Ani Shabri,** Mathematics Department, University Technology Malaysia, Johor Bahru, Malaysia.

**Ernie Mazuin**, Instrumentation & Control Engineering Department, University Kuala Lumpur, Pasir Gudang, Malaysia.



The procedure in processing the DWT is not autoregressive in nature. Furthermore, since the band-pass filters are chosen to decompose target signals, the band-pass filters really affect the accuracy of the decomposition. Another items that affect the results of the decomposition are the basics function of wavelet and also the decomposed layer [4]. For that reason, the accuracy of the DWT decomposition is quite lower than the EMD. However, the EMD model can be coupled with vector machine of discrete wavelet support to overcome these limitations [5].

The first segment of the paper presented the SVM model, Empirical mode decomposition theory and wavelet decomposition theory as the basic theories for the arrivals of tourist. The second segment of the paper proposed the prediction quality improvement by developing the hybrid of EMD_WSVM model. While the last segment of the paper discussed on the results of prediction together with comparison and analysis.

## II. METHODOLOGY

### SVM Model

The Support Vector Machines (SVM) is a learning algorithm that is statistical in nature. It is proposed by Vapnik. Vapnik developed this method with the basis of dimension theory of VC (Vapnik-Chervonenkis) and also the principle of SRM (Structural Risk Minimization) [5]. To obtain the best promotion ability, SVM looks for the best arrangement between the complexity of the model and its learning ability. The advantage of SVM is that it is strongly capable in treating nonlinear data. Finding a mapping function $\phi(x)$ that is nonlinear, then turning it into a linear inseparable data x in the low-dimensional feature space projected into the high-dimensional feature space H to make it a linear separable problem, is the basic working principle of support vector machines. The regression function of SVM in high dimensional feature space is:

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

Where $w$ is the weight vector of high dimensional feature Space, $w \in R^k$; and $b$ is the bias constant, $b \in R$.

According to the principle of structural risk minimization, Equation (1) can be converted into:

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

Subject to

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

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

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

The solution to this minimization problem is of the form

$$f(x) = \sum_{i=1}^{m}\left(\lambda_i - \lambda_i^*\right)K(x_i, x) + b \quad (3)$$

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

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

And respectively.In this paper, the radial basis function is used as the kernel
function which is given by:

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (4)$$

Where σ is the width of the radial basis function

**The discrete wavelet transforms method (WSVM)**

The theory of wavelet analysis was founded on the Fourier analysis [6]. The model is obtained by combining two methods, DWT and SVM. The WSVM model is an SVM model, which uses sub-time series components obtained using DWT on original data [7]. For WSVM model inputs, the original time series data are decomposed into a certain number of sub-time series components (Ds) [8].

Each component plays different role in the original time series and the behavior of each sub-time series is distinct. Each component plays a different role in the original time series and the behaviour of each sub time series is distinct [9]. To select the number of decomposition level, the formula M=log(n) was used where n is length of the time series and M is decomposition level [10].

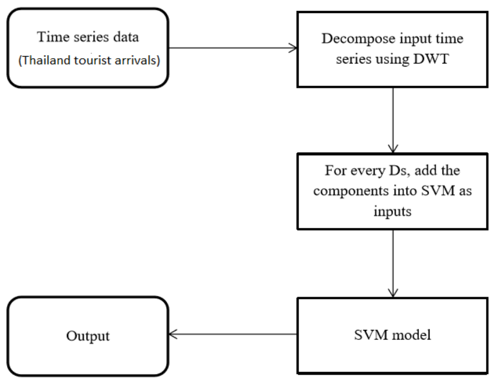


**Fig. 1 The WSVM model struct**

**Empirical Mode Decomposition (EMD)**

The function EMD was to capture the nonlinear and non-stationary signals. By using EMD, any complex datasets can be decomposed into a finite and small number of components called IMF [11]. EMD has shown its applicability in signal processing.

In this study, EMD was used to decompose the number tourist arrivals datasets into sub datasets, also known as IMFs. As EMD is a proven method for capturing nonlinear and non-stationary data, the idea of using EMD is that EMD uses a non-stationary filter.

The main idea of EMD is to decompose the original data into a finite and small number of oscillatory modes based on the local characteristic time scale itself. An Intrinsic Mode Function (IMF) expresses each oscillatory mode, which is similar to a harmonic function [12].

The basic principle of EMD is to decompose a time series into a sum of oscillatory functions, namely, intrinsic mode functions (IMFs). In the EMD, the IMFs must satisfy two conditions:

(i)The number of extrema (sum of maxima and minima) and Comparative Performance of All Models (ii) The local average is zero. The condition that the local average is zero implies that envelope mean of the upper envelope and lower envelope is equal to zero.

## III. DATA

Monthly data of tourist arrivals for Thailand countries consist of 204 observations which ranged from January 1999 till December 2015.

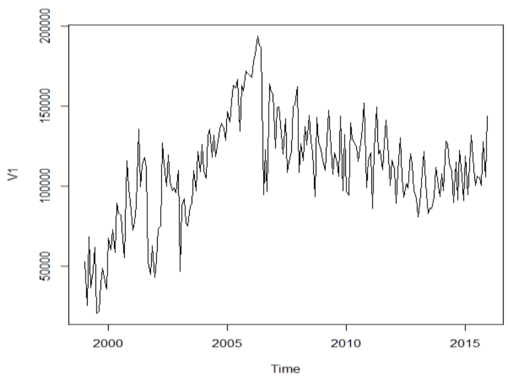


**Fig. 2 Monthly Thailand tourist arrivals (Jan 1999 to Dec 2015)**

## IV. RESULTS AND DISCUSSION

In this study, number tourist arrivals time series data of Thailand countries are used to present the forecasting accuracy of the EMD_WSVM method. Three forecasting methods are used to validate the forecasting performance of EMD_WSVM. Table 1 shows three error measurements with their formula. These measurements will be utilized to evaluate the forecasting accuracy for each method.

**Table. 1 Error measures are used in study**

| Name of measure error | Formula of measure error |
|---|---|
| Root Mean Squared Error | $RMSE = \sqrt{\frac{\sum_t^n e_t^2}{n}}$ |
| Mean Absolute Percentage Error | $MAPE = \sum_{t=1}^{n} \left\lvert \frac{e_t}{observed_t} \right\rvert X \frac{100}{n}$ |

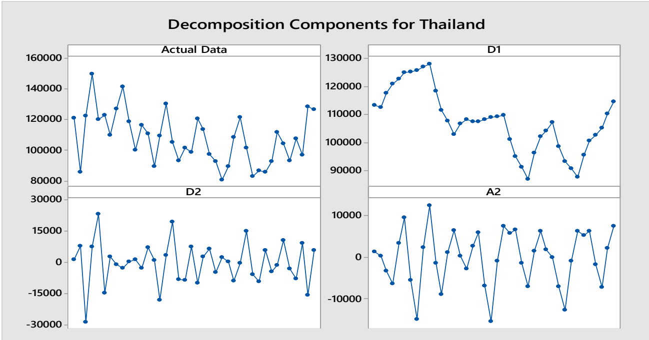


**Fig. 3 Decomposed wavelet sub-series components (Ds) of tourist arrival data for Thailand**

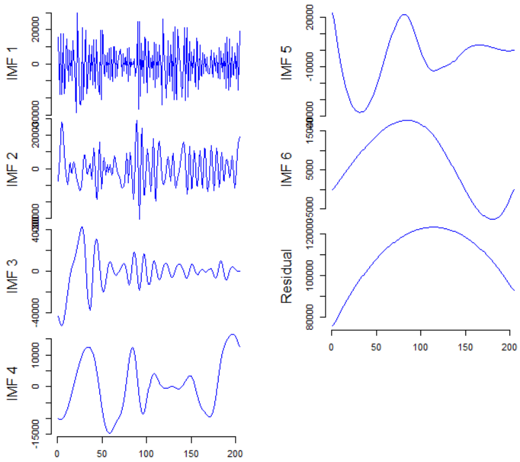


**Fig. 4 The IMFs and Residue Components of tourist arrival data for Thailand**

For this study, two decomposition levels (D1 and D2) and one approximation (A2) were used for dataset. The decomposition components and approximation components then acted as input for SVM forecasting. **Error! Reference source not found.** show the original tourist arrivals data time and their Ds decomposition using wavelet method for each of the country. From the observation on the figures 3, D2 and A2 displayed stationarity plot since the data randomly distributed around zero, uniform movement cannot be found to be present.

The monthly time series for the Thailand tourist arrivals were decomposed using EMD. Figures 4 show the decomposition components of IMFs and residue for case studies. It can be observed that EMD had decomposed the datasets into six IMFs and one residue for data tourist arrivals data set.

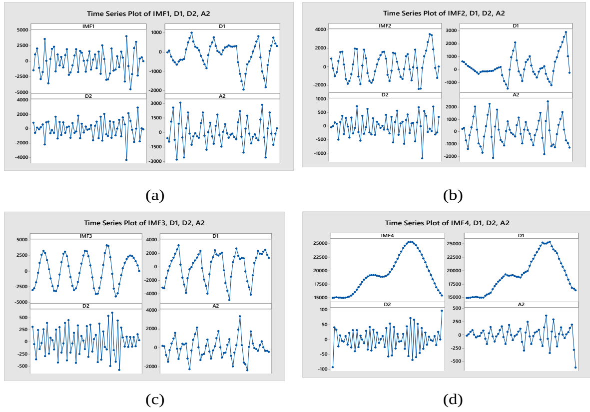


**Fig. 5 Decomposition IMFs Dataset (a) IMF1 (b) IMF 2 (c) IMF 3 (d) IMF4**

From the Figures 5, all the non-stationary data in IMFs were transposed into stationary dataset and from the observation plot, the decomposition level (D1) produced similar pattern to its IMFs'.

**Table. 2 Comparative Performances of All Models**

| Method | RMSE | MAPE(%) |
|---|---|---|
| SVM | 15388.7186 | 11.87 |
| WSVM | 11891.6177 | 8.58 |
| **EMD_WSVM** | **10711.56** | **7.96** |

For further analysis, the error statistics of SVM, WSVM and EMD_WSVM were compared to each other to find the best model for Thailand tourist arrivals forecasting. Table 2 compares the testing results among the current hybrid model WSVM approaches and single model SVM based on two statistical measurements, which were RMSE and MAPE values.

The results obtained for Thailand indicated that the lowest RMSE and MAPE values were acquired from EMD_WSVM model. Therefore, EMD_WSVM was declared to be the best model representing tourist arrivals from Thailand, followed by WSVM and SVM.

## V. CONCLUSION

The approximation ability of a function by using intelligent algorithm based on methods of prediction and application is powerful. This is because, in dealing with nonlinear systems that are complex, these methods are able to fit the expression of a function of a system that is unknown and the methods are effective too. However, this method cannot fully identify and extract the internal characteristics of complex nonlinear and non-stationary timeseries [8].

Despite the fact, the prediction methods with decomposition algorithm can effectively identify and extract the internal features and laws of nonlinear non-stationary time series. The combination the two method can improve the prediction accuracy of nonlinear non-stationary timeseries [5].

## VI. ACKNOWLEDGMENT

This work has been supported by Research & Innovation, Universiti Kuala Lumpur, Malaysian Institute of Industrial Technology (MITEC).

## REFERENCES

1. 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.
2. A. Shabri, "A Hybrid of EEMD and LSSVM-PSO model for Tourist Demand Forecasting," Indian J. Sci. Technol., vol. 9, no. 36, pp. 1–6, 2016.
3. 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.
4. L.-Y. Geng, "Forecast of Stock Index Volatility Using Wavelet Support Vector Machines," Adv. Manag. Sci., pp. 19–22, 2012.
5. J. Li, S. Zhu, Q. Wu, and P. Zhang, "A Hybrid Forecasting Model Based on EMD-GASVM-RBFNN for Power Grid Investment Demand," vol. 2018, 2018.
6. N. G. . SadeghpourHajia.M, Mirbagherib.S.A, Javidc A. H., Khezria.M, "A Wavelet Support Vector Machine Combination Model for Daily Suspended International Journal of Engineering A Wavelet Support Vector Machine Combination Model for Daily Suspended," Int. J. Eng., vol. Vol. 27, no. June 2014, pp. 855–864, 2014.
7. O. Kisi and M. Cimen, "A wavelet-support vector machine conjunction model for monthly stream flow forecasting," J. Hydrol., vol. 399, no. 1–2, pp. 132–140, 2011.
8. A. Rafidah, A. Shabri, A. NurulhudaAbd, 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.
9. 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.
10. S. M. Pandhiani and A. Bin Shabri, "Time Series Forecasting by Using Hybrid Models for Monthly Stream flow Data," Appl. Math. Sci., vol. 9, no. 57, pp. 2809–2829, 2015.
11. A. M. Jaber, M. T. Ismail, and A. M. Altaher, "Application of empirical mode decomposition with local linear quantile regression in financial time series forecasting," Sci. World J., vol. 2014, 2014.
12. S. Abadan and A. Shabri, "Hybrid Empirical Mode Decomposition-ARIMA for Forecasting Price of Rice," Appl. Math. Sci., vol. 8, no. 63, pp. 3133–3143, 2014.

## AUTHOR PROFILE

**A.Rafidah**, Technical Foundation Department, University Kuala Lumpur, Pasir Gudang, Malaysia.

**Ani Shabri,** Mathematics Department, University Technology Malaysia, Johor Bahru, Malaysia.

**Ernie Mazuin**, Instrumentation & Control Engineering Department, University Kuala Lumpur, Pasir Gudang, Malaysia.