Bulletin of Electrical Engineering and Informatics
ISSN: 2302-9285
Vol. 6, No. 2, June 2017, pp. 149~158, DOI: 10.11591/eei.v6i2.648

Comparison of Tropical Thunderstorm Estimation between Multiple Linear Regression, Dvorak, and ANFIS

Wayan Suparta*, Wahyu Sasongko Putro
Space Science Centre (ANGKASA), Institute of Climate Change,
Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor Darul Ehsan, Malaysia
Tel: 603-89118485, Fax: 603-89118490
*Corresponding author, e-mail: wayan@ukm.edu.my

Abstract

Thunderstorms are dangerous and it has increased due to highly precipitation and cloud cover density in the Mesoscale Convective System area. Climate change is one of the causes to increasing the thunderstorm activity. The present studies aimed to estimate the thunderstorm activity at the Tawau area of Sabah, Malaysia based on the Multiple Linear Regression (MLR), Dvorak technique, and Adaptive Neuro-Fuzzy Inference System (ANFIS). A combination of up to six inputs of meteorological data such as Pressure (P), Temperature (T), Relative Humidity (H), Cloud (C), Precipitable Water Vapor (PWV), and Precipitation (Pr) on a daily basis in 2012 were examined in the training process to find the best configuration system. By using Jacobi algorithm, H and PWV were identified to be correlated well with thunderstorms. Based on the two inputs that have been identified, the Sugeno method was applied to develop a Fuzzy Inference System. The model demonstrated that the thunderstorm activities during intermonsoon are detected higher than the other seasons. This model is comparable to the thunderstorm data that was collected manually with percent error below 50%.

Keywords: *ANFIS, meteorology data, thunderstorm estimation*

1. Introduction

Thunderstorms are one of the greatest dangers to space activities such as in the commercial aircraft operation. The general effects can be found in the form of turbulence, heavy rain, and runway contamination. Hazen et al. [1] reported that near 75% of all space shuttle countdowns between 1981 and 1994 were delayed or scrubbed with about one-half of these due to bad weather. The primary causes of termination of space shuttle launch such are thunderstorm phenomena [2]. Thunderstorm phenomena are electrical discharges where the development of cumulonimbus cloud produced lightning during heavy rain and thunder during precipitation. Thunderstorms cover larger precipitation in the form of mesoscale convective systems (MCS). MCS tend to form near weather fronts and able to generate lightning. Accurate estimation of thunderstorm activity over MCS area is important to commercial space vehicle launch operation, navigation, agriculture, and so on.

Current studies have shown that thunderstorm activities can be estimated using rainfall radar over Jeddah, Saudi Arabia [3]. The rainfall radar based on the empirical relationship between reflectivity (Z) and rainfall rate (R) was used to estimate thunderstorm activity. However, the rainfall radar has a limitation in the limited area and minimum in signal detection. Velden et al. [4] have used a historical image treatment called *Dvorak* technique to improve this limitation. The statistical method e.g. Monte Carlo simulation can also be used to estimate thunderstorm activity as suggested by Balijepalli et al. [5]. However, the simulation is unable to evaluate the impact of a fault during an individual storm event. Recently, the Artificial Neural Network (ANN) method has attempted to forecast several thunderstorms [6-7]. However, another technique which more powerful to increase the convergence and to avoid overfitting was used the Adaptive Neuro-Fuzzy Inference System (ANFIS) [8-9]. This method is cost-effective, robust and with better accuracy for estimating of meteorological parameters. In this study, ANFIS is employed for estimating the frequency of thunderstorm. Tawau in Sabah,

Received February 28, 2017; Revised April 28, 2017; Accepted May 12, 2017

Malaysia of the tropical region is selected for the case study to plan for space activities development. In this area, the low cloud cover was higher during intermonsoon [10]. MCS activity is also increased the precipitation level during the winter season (December, January, and February) as reported by Suparta et al. [11].

2. Methodology

In this study, the estimation of thunderstorm frequency by ANFIS will be compared with *Dvorak* technique and Multiple Linear Regression (MLR). Six combination parameter inputs from meteorological parameter were constructed to develop a linear equation. The MLR with Jacobi method is an algorithm for determining the solutions of a diagonally dominant system of linear equations [12]. It was used to find the best iteration by using a linear equation over few combination parameters. While *Dvorak* technique is a widely system used to estimate the intensity of a tropical cyclone based on enhanced visible and infrared satellite images [4].

2.1. Location and Data Processing

The Tawau station (TWU: 4.32°N, 118.12°E with elevation 17 m) is situated between Borneo (Indonesia) and the Celebes Sea. The climate condition in this area is relatively hot and wet with average shade temperature about 26°C to 29°C at noon and falling to around 23°C at night. The precipitation throughout the year, with a tendency for November, December and January to be the wettest months, and February and March become the driest months with mean rainfall varies from 1800 mm to 2500 mm [5]. However, in 2012, the weather condition over Tawau area increases with surface air pressure > 1008 mbar, relative humidity > 70%, and temperature < 32 °C at noon.

For this study, six parameter inputs such as Pressure (P), Temperature (T), Relative Humidity (H), and Cloud (C) were collected from the Malaysian Meteorological Department (MetMalaysia) to characterize the thunderstorm activity. Data was taken over one year (1 January~31 December 2012) for each parameter. Water vapor from radiosonde (PWV) and precipitation (Pr) data was taken from the University of Wyoming website (http://weather.uwyo.edu/upperair/sounding.html) and the NASA website (http://gdata1.sci.gsfc.nasa.gov), respectively.

After collecting six meteorological parameters, correlation and regression analysis between the two variables is conducted to identify which parameters to have a good correlation with thunderstorm data. The six combinations were proposed to create a linear equation. The value of correlation coefficient (r) is to indicate two variables correlated perfectly with a linear relationship. Furthermore, the best values of determination coefficient (R2) were used to design configuration input and output of observation data using MLR method. Since the thunderstorm data that was collected manually by MetMalaysia on a daily basis with thunderstorm occurred (recorded as 1) and no thunderstorm (recorded as 0), daily data were processed to filter and enhance the quality of raw data. A lot of linear equations are obtained and transposed into Jacobi formula. The training data to obtain estimation model were processed using Jacobi algorithm with 1,000 iterations. Three equation models were suggested to estimate the thunderstorm frequency based on the three seasons in 2012 which were summer, winter, and transition season (intermonsoon).

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

As introduced by Jang [13], ANFIS is the combination of fuzzy logic and an adaptive neural network. In this paradigm, a FIS is constructed in three steps: the rule base, fuzzy sets (membership function), and inference procedures (Takagi–Sugeno, Mamdani, and Tsukamoto types) [14]. The rule base section includes selection fuzzy If-Then rules, fuzzy set, and fuzzy reasoning which is the inference procedure rule base for the output target. One of the main issues when constructing a FIS is that there is no specific order for the shape selection of MFs (known as premise parameters) and rule procedures. The architecture and algorithm procedure for constructing the FIS with a fuzzy system from specified inputs and outputs is supervised by the ANN algorithm [15]. In this construction, the Takagi–Sugeno works with a linear technique [16], where membership function has two rule inputs (A and B), and consequently part f1 and f2 are the rule output. The premise part f1 (p1, q1, r1) and f2 (p2, q2, r2) are a linear parameter for the Takagi–Sugeno fuzzy model. Details of the MFs in this work can be referred to Suparta and Alhasa [9].

The architecture of ANFIS consists of five layers (see Figure 1), and the model is briefly explained as follows.

Layer 1: Generate membership grade for every node, such as and, with the output parameters as given by

$$O_{1,i} = \mu_{Ai}(x), \qquad i = 1,2 \tag{1}$$

$$O_{1,i} = \mu_{Bi-2}(y), \qquad i = 3,4 \tag{2}$$

In this study, all the inputs of the membership function are generalized by the Gaussian function:

$$\mu_{Ai}(x) = \frac{1}{1 + \left[\dfrac{x - c_i}{a_i}\right]^{2b}} \qquad \mu_{Bi-2}(y) = \frac{1}{1 + \left[\dfrac{x - c_i}{a_i}\right]^{2b}} \tag{3}$$

where *x* and *y* are inputs for the node; A_i or B_{i-2} is a fuzzy linguistic model associated with node; $O_{1,i}$ is a membership function grade of fuzzy set and $\{a_i, c_i\}$ is the parameter set of the membership function in the premise parameter with *b* is constant.

Layer 2: The AND operator (T-Norm operator) was used to find the output. Every type of node in this layer is fixed with the Π label, where each node output demonstrates the firing strength of a rule and expressed as:

$$O_{2,i} = w_i = \mu_{Ai}(x) \times \mu_{Bi}(y) \qquad i = 1,2 \tag{4}$$

where w_i is firing strength of each rule generated with the product method.

Layer 3: The *N* label was deployed in this layer to calculate the ratio of the i^{th} firing strength rule. The sum of all firing strength rules ($\overline{w}_i$) is expressed as:

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \qquad i = 1,2 \tag{5}$$

Layer 4: The total output from the contribution of the i^{th} rule from all nodes is obtained by following equation:

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \qquad i = 1,2 \tag{6}$$

where $\overline{w_i}$ is the normalized firing strength from the previous layer and $p_1x+q_1y+r_1$ are the parameter set on the first order of the Sugeno FIS model.

Layer 5: The Σ label was deployed in this node with a fixed node. The Σ symbol calculates the overall incoming signal input as follows:

$$O_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \qquad i = 1,2 \tag{7}$$

where f_i is the overall output of the signal and w_i is the premise parameter. The type of MFs for all the inputs was generalized from a Gaussian function. Jang (1993) has proposed a hybrid learning algorithm to optimize all the input parameters and combine the least squares estimation (LSE) to obtain the consequent parameter. The premise parameters (MFs) are assumed to be fixed-rate for the current cycle through the training set (Jang and Sun 1995), and the error rate is propagated backward. Furthermore, the gradient descent was used to update the premise parameter by minimizing the total average of the squared error in ANFIS.

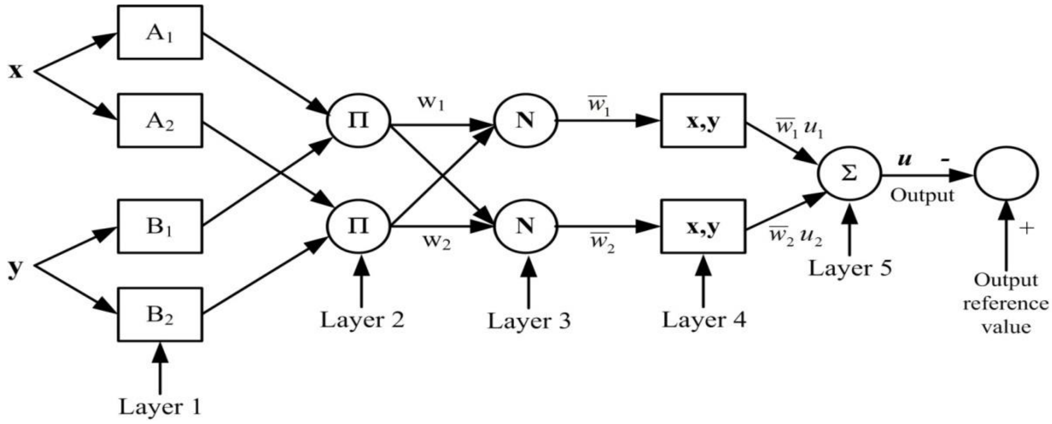

Figure 1. Adaptive neuro-fuzzy inference system structure adopted from Suparta and Alhasa [9]

To create rule base and MFs, the correlation analysis will determine the best configuration of FIS design which includes the initialization of input and output parameters. In this work, a grid partition method was applied to the FIS to obtain a total rule base with the equation expressed as follows:

$$\sum G_{rb} = 2^{n+1} \tag{8}$$

where ΣG_{rb} and n are the total numbers of rule bases and MFs, respectively. The design included determination of maximum and minimum values of the input parameter. The input parameter selected in this work is based on the target output designed to obtain a good estimation result with the highest accuracy. Three MFs for each input and output data are designed using the Gaussian function (µ).

3. Results and Analysis

3.1. Thunderstorm

Figure 2. shows the thunderstorm frequencies that occurred during transition season I (Jan–Feb) and II (Dec) in the Tawau area. From this figure, the thunderstorm can be detected if air pressure is < 1008 mbar, the temperature is < 26°C, relative humidity is > 80%, and cloud density is more than 6 oktas. Furthermore, PWV and precipitation have occurred over 40 mm and > 30 mm/day, respectively. However, there was a clear day in July during summer (June, July, and August) when the thunderstorm frequency was more than 10 events. The minimum precipitation and high temperature indicated that the clear day was obtained in the middle of July. In addition, the rainy day during the winter season occurred in January and February 2012 (winter I) and December (winter II). The temperature has decreased and precipitation has increased in the DJF month.

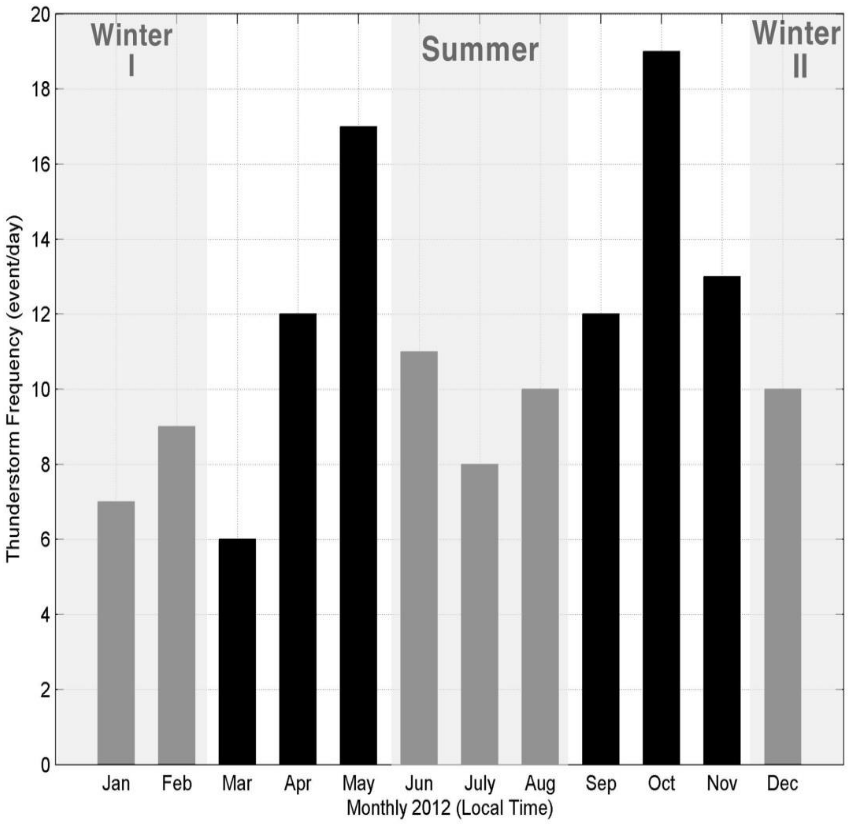

Figure 2. The variation of thunderstorm over the Tawau area over the year 2012. The solid bar shows the intermonsoon months

3.2. Relationship between Input and Output for Thunderstorm Estimation

A standard error regression (S) and R-squared (R^2) values are used to show the relationship between input and output in order to identify which meteorological parameter in good correlation with thunderstorm activity (see Table 1). From configuration developed showed that the error values for S and R^2 are obtained below 50% and 6%, respectively. Finally, results show that the configuration input with *H* and PWV obtained a good result as compared to another configuration. Note in the table that the configuration input with obtained error value of S which more than 50% is excluded. On the other hand, S value obtained above 47% is probably due to the predictor had an incomplete measurement data. After the best configuration input was obtained, the input with *H* and PWV are proposed to develop an MLR equation by using the observation data in 2012.

Table 1. The Relationship between Input and Output for the data of 2012 with the Output of Thunderstorm

Input	Configuration Input	S (%)	R^2 (%)
One Input	*P*	48.2	0.2
	T	47.8	2.0
	H	47.0	0.9
	C	48.1	0.5
	PWV	48.3	0
	Pr	47.1	4.9
Two Inputs	*T and H*	47.8	2.0
	T and C	47.8	2.1
	T and Pr	47.2	5.5
	T and PWV	47.8	2.1
	H and C	48.1	1.1
	H and Pr	47.1	5.1
	H and PWV	47.0	0.9
	C and Pr	47.1	4.9
	C and PWV	48.2	0.5
	Pr and PWV	47.1	5.0
	T and P	47.8	2.2
Three Inputs	*T, H, and C*	47.9	2.1
	T, H, and Pr	47.0	5.5
	T, H, and PWV	47.9	2.1
	T, C, and Pr	47.0	5.5
	T, C, and PWV	47.9	2.1
Four Inputs	*T, H, C, and Pr*	47.1	5.5
	H, C, Pr, and PWV	47.2	5.3
	T, C, Pr, and PWV	47.1	5.7
	T, H, Pr, and PWV	47.0	5.7
Five Inputs	*T, H, C, Pr, and PWV*	47.1	5.7
Six Inputs	*P, T, H, C, Pr, and PWV*	47.2	5.8

As mentioned in Section 2.1, selection configuration inputs with 1000 iterations was used to find the optimum input model for estimation process. The maximum iteration for x_1~x_2 is presented in two-three steps with two configurations for each season (see Table 2). The Python software was used to obtain the estimation model. On the other hand, the relative humidity (x_1) and PWV (x_2) can be determined to estimate the frequency of thunderstorm based on the Jacobi algorithm.

Table 2. The Estimation of *H* and PWV using Jacobi Algorithm for the Season in 2012

Season	The Jacobi Equation	Solution
Winter I	$3.48= 0.0462\ x_1$	$x_1= 75.324$
(JF) month	$3.33=0.0514\ x_1-0.0104\ x_2$	$x_2= 52.08$
Summer	$-2.64=-0.0294\ x_1$	$x_1= 89.795$
(JJA) month	$-2.58=-0.0294\ x_1 + 0.00092\ x_2$	$x_2= 65.19$
Winter II	$3.49=0.0470\ x_1$	$x_1= 74.255$
(Dec) month	$3.90 =0.0465\ x_1 + 0.0080\ x_2$	$x_2= 55.89$
Transition I	$1.60=0.0248\ x_1$	$x_1= 64.516$
(MAM) month	$1.61=0.0264\ x_1-0.00217\ x_2$	$x_2= 42.96$
Transition II	$2.12=0.0325\ x_1$	$x_1= 65.230$
(SON) month	$2.06=0.0371\ x_1-0.00736\ x_2$	$x_2= 48.92$

Note that x_1 and x_2 stand for *H* and PWV, respectively

From Table 2, three equations representing of each season in Tawau, Sabah for the data in 2012 with the output of thunderstorm can be presented as follows.

Summer: $$y_s = 2.58 - 0.0294\,a + 0.00092\,b \quad (9)$$

Intermonsoon: $$y_i = -2.96 + 0.03175\,a + 0.004765\,b \quad (10)$$

Winter: $$y_w = -3.41 + 0.04895\,a - 0.0092\,b \quad (11)$$

where *a* and *b* represent for *H* and PWV, and *y* is the output (thunderstorm).

Figure 3 shows the frequency of thunderstorm over Tawau between estimation and the observation data. As can be seen from the figure, five equations were used in the Jacobi methods which were winter I, a transition I, summer, transition II, and winter II. Based on the investigation, the minimum and maximum *H* and PWV are 40~100% and 38~65 mm, respectively. The estimation result of thunderstorm frequency for each season was obtained by using rounding number technique. For example, if the output reached ≤ 0.5, it would mean that there is no event (value=0) while output at > 0.5 indicate that an event may occurred (value=1).

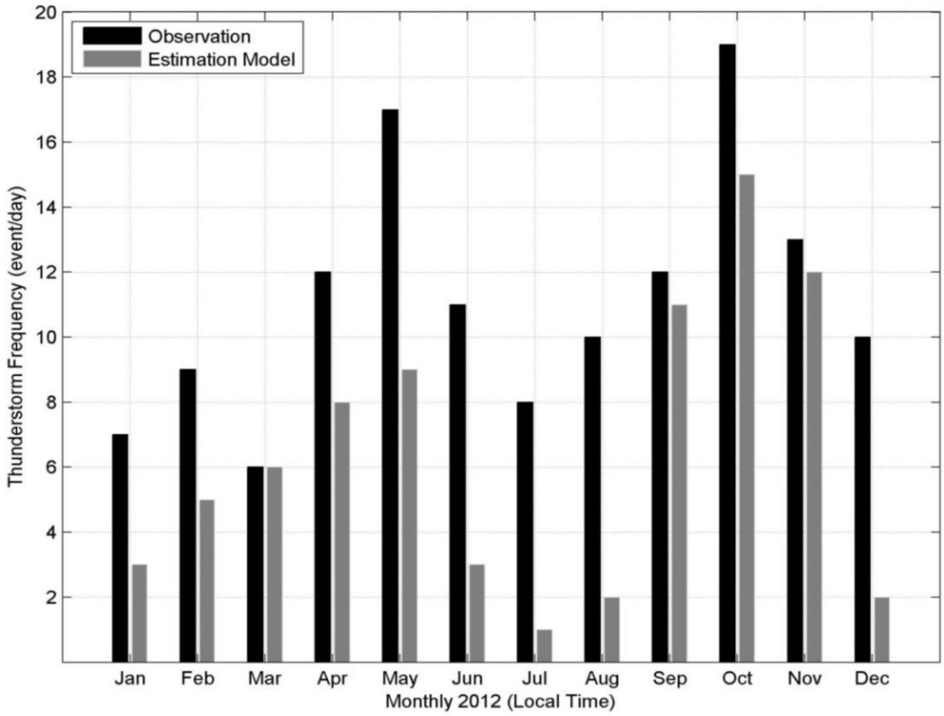

Figure 3. The estimation model of thunderstorm activity over Tawau station

3.3. Comparison between MLR model, Dvorak Technique and ANFIS FCM

In order to evaluate the performance of the MLR model in estimating thunderstorm activity, the MLR model was compared with *Dvorak* technique and *ANFIS* with *Fuzzy Clustering Method* or Fuzzy C-means (*FCM*). FCM is a method of clustering which allows each data point to belong to two or more clusters. Membership grades are assigned to each of the data points. The *Dvorak* techniques showing the three level of evolution cloud which is weakest, stronger and strongest. However, in this study, the level of the storm was determined as 1 (strongest) and 0 (low–mid level or weakest-stronger) based on one-year data (1 January~31 December 2012). For weakest, stronger, and strongest levels, the density of cloud index is reached 0~3 oktas, 3~5 oktas and > 5 oktas, respectively. After one-year data processed for TWU, the strongest thunderstorm event captured by MTSAT satellite-Japan was obtained in October 2012 (see Figure 4). More than 16 events per month were captured, particularly in intermonsoon season. In addition, the thunderstorm frequency has decreased in summer and winter seasons, respectively.

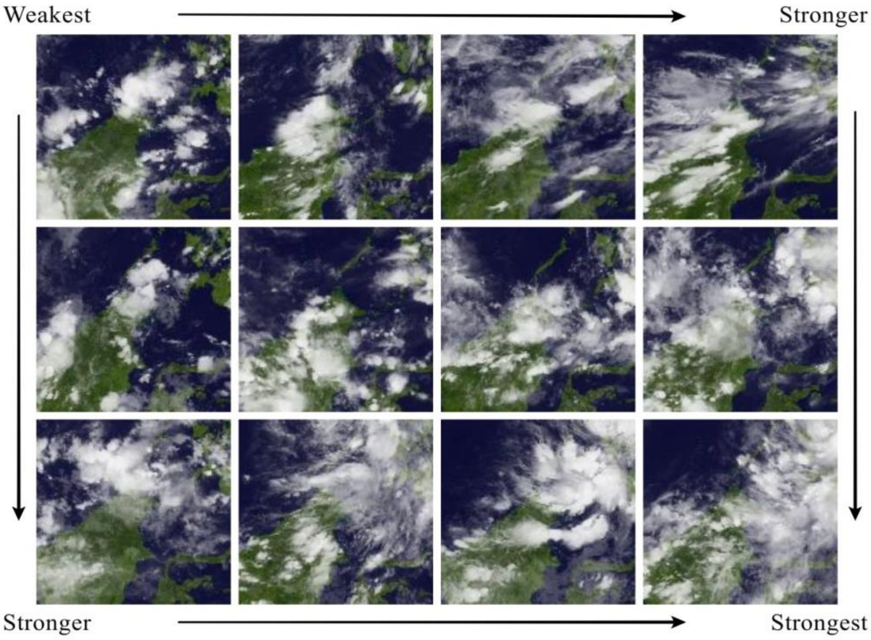

Figure 4. The evaluation of cloud level during the tropical storm 2012 using *Dvorak* technique. The image is taken from MTSAT satellite with resolution 140×160 (http://weather.is.kochi-u.ac.jp/archive-e.html)

Furthermore, Figure 5 shows the comparison of the estimation result of ANFIS FCM model with two configurations inputs (*H* and PWV) and one output with MLR model and ANFIS FCM model and *Dvorak* technique. The model used 1,000 epoch and five layers (fixed layer ANFIS) and found with a maximum error below 1%. From the figure, it can be seen that the three models compared, it provides a strong relationship (> 0.75 at the 99% confidence level), where ANFIS FCM reached the highest correlation followed by the *Dvorak* technique and the MLR model. In addition, all three methods reached minimum error (see Table 3) as indicated by root mean square error (RMSE), mean absolute error (MAE), and percent error (PE). Based on the comparison above, ANFIS is suggested in the near future for construction of a prediction model for monitoring thunderstorm activity.

Table 3. Statistical Comparison of Thunderstorm Activity for Tawau in 2012 Based on Three

Method	Correlation (R^2)	RMSE	MAE (%)	PE (%)
MLR Model	0.8911	4.7697	4.4167	39.7288
Dvorak technique	0.9352	3.8944	3.6667	36.4322
ANFIS FCM	0.9812	1.8930	1.7500	17.2097

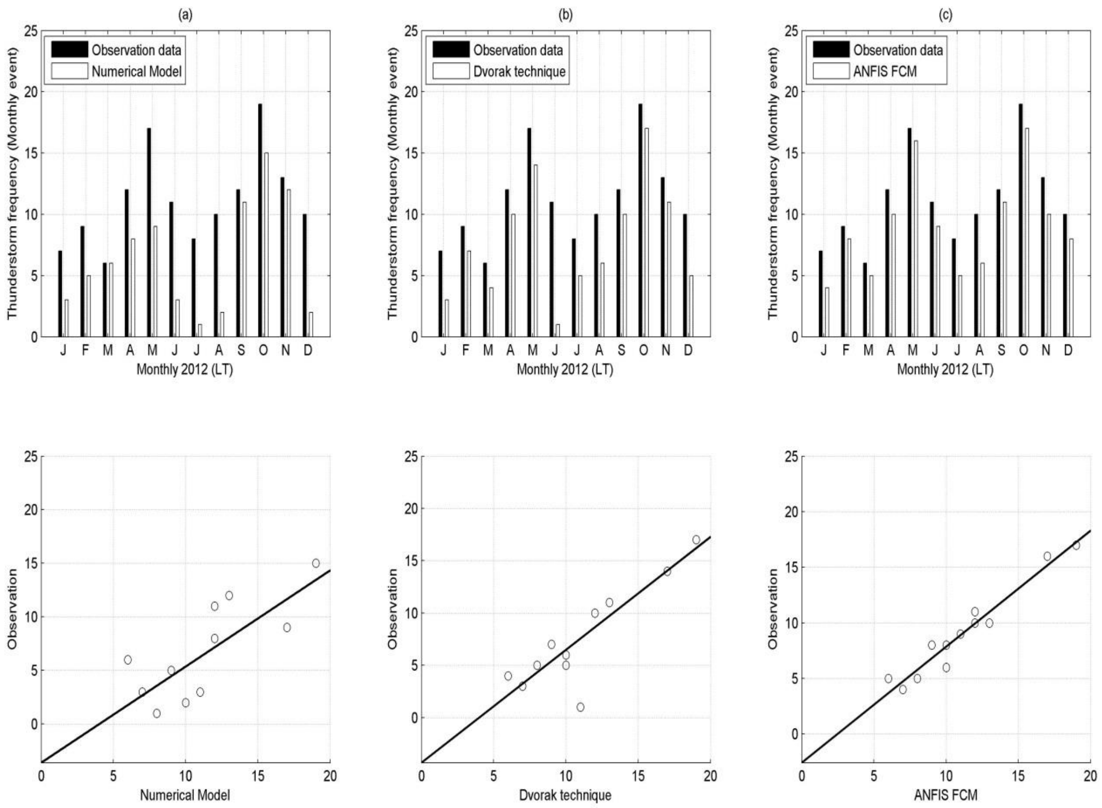

Figure 5. Comparison of observation data between (a) MLR model, (b) *Dvorak* technique, and (c) ANFIS FCM model

4. Conclusion and Recommendation for Future Work

The MLR Jacobi, Dvorak technique, and ANFIS FCM were carried out to estimate thunderstorm activity over Tawau Area for the year of 2012. Based on the six meteorological data examined, a model has been successfully developed based on two configuration inputs (*H* and PWV) to estimate thunderstorm frequency. Results showed that the thunderstorm activity in Tawau was found highest in the intermonsoon season in March~May (winter to summer) and September~November (summer to winter), respectively. Comparison between the three models has shown that MLR model gives percentage error 40%, which probably unsuitable employed for estimation of a complex thunderstorm. The ANFIS model is more advantages in constructing the estimation of thunderstorm activity, and therefore, it is suggested as a predictive model for the next studies. Finally, from a mathematic model that has been obtained, it can also be applied in other locations to estimate thunderstorm activity as long as meteorological data are provided.

Acknowledgements

This study is partially funded by the Ministry of Science, Technology and Innovation (MOSTI) under Science Fund 01-01-02-SF1100 and 01-01-02-SF1241 grants. We would like to thank the Malaysian Meteorological Department (MET Malaysia) for providing the pressure, temperature, relative humidity, cloud, and thunderstorm data, NASA website for providing the precipitation data, and the Department of Atmospheric Science, The University of Wyoming for providing the Radiosonde PWV data.

References

[1]. Hazen DS, Roeder WP, Lorens BF, Wilde TL. *Weather Impacts on Launch Operations at the Eastern Range and Kennedy Space Center*; Preprints, Sixth Conf. On Aviation Weather Systems, Dallas, TX, Amer. Meteor. Soc. 1995: 270-275.
[2]. Kuk B, Kim H, Ha J, Lee H, Lee G. A Fuzzy Logic Method for Lightning Prediction Using Thermodynamic and Kinematic Parameters from Radio Sounding Observations in South Korea. *Wea. Forecasting*. 2012; 27: 205–217.
[3]. Albar A, AL-Khalaf A K, Abdel-Basset H. Radar Rainfall Estimation of a Severe Thunderstorm over Jeddah. *Atmospheric and Climate Sciences*. 2015; 5: 302-316.
[4]. Velden C, Harper B, Wells F, Beven II J L, Zehr R, Olander T, Mayfield M, “Chip” Guard C, Lander M, Edson R, Avila L, Burton A, Turk M, Kikuchi A, Christian A, Caroff P, Mccrone P. The DVORAK Tropical Cyclone Intensity Estimation Technique, A Satellite-Based Method that Has Endured for over 30 Years. *Bulletin of American Meteorological Society*. 2006; 87: 1195-1210.
[5]. Balijepalli N, Venkata S, Richter Jr C, Christie R, Longo V. Distribution System Reliability Assessment Due To Lightning Storms. *IEEE Trans. Power Delivery*. 2005; 20: 2153-2159.
[6]. Suykens J A K. Nonlinear Modeling and Support Vector Machine. *Instrumentation and Measurement Technology Conference 2001 (IMTC 2001).* Budapest, 2001: 287-294.
[7]. Litta A J, Idicula S M, Mohanty U C. Artificial Neural Network Model in Prediction of Meteorological Parameters during Premonsoon Thunderstorms. *International Journal of Atmospheric Sciences*. 2013: 1-14, doi: 10.1155/2013/525383.
[8]. Chaudhuri S, Das D, Middey A. An Investigation on the Predictability of Thunderstorms over Kolkata, India Using Fuzzy Inference System and Graph Connectivity. *Natural Hazards*. 2015; 76(1): 63-81.
[9]. Suparta W, Alhasa K M. Modeling of Tropospheric Delays Using ANFIS. *Springer Briefs in Meteorology*. 2016: 109, doi: 10.1007/978-3-319-28437-8.
[10]. Engel-Cox J, Nair N, Ford J. Evaluation of Solar and Meteorological Data Relevant To Solar Energy Technology Performance in Malaysia. *Journal of Sustainable Energy & Environment*. 2012; 3: 115-124.
[11]. Suparta W, Putro W S, Singh M S J, Asillam M F. Characterization of GPS and Meteorological Parameters for Mesoscale Convective Systems Model over Tawau, Malaysia. *Advanced Science Letters*. 2015; 21(2): 203-206.
[12]. Wilson J M, Crook N A, Mueller C K, Sun J, Dixon M, Nowcasting Thunderstorms: A Status Report. *B. Am. Meteorol. Soc*. 1998; 79: 2079–2099.
[13]. Jang J S R. ANFIS: Adaptive Network-Based Fuzzy Inference Systems. *IEEE Transactions on Systems Man and Cybernetics*. 1993; 23: 665–685.
[14]. Lin J Y, Cheng C T, Sun Y G, Chau, K. Long-term Prediction Of Discharges In Manwan Hydropower Using Adaptive-Network-Based Fuzzy Inference Systems Models. *Lecture Notes Computer Science*. 2005; 3612: 1152–1161.
[15]. Jang J S R, Sun C T, Mizutani E. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. New Jersey: Prentice Hall. 1997; (3th Ed.).
[16]. Takagi T, Sugeno M. Fuzzy Identification of System and Its Application to Modeling and Control. *IEEE Transactions on Systems Man and Cybernetics*. SMC-15, 1985: 116-132.