PREDICTION OF THUNDERSTORM OCCURRENCES IN TROPICAL AREAS USING A NUMERICAL MODEL

Wayan SUPARTA1, Wahyu Sasonoko PUTRO2, T. Ken DARMASTONO3

DOI: 10.21163/GT_2021.161.0

ABSTRACT:
The prediction of thunderstorm activity is not only significant for weather forecasting but also for the standardization of risk assessment as in the aviation industry or emergency unit purposes. This paper aimed to develop a prediction of thunderstorm occurrences using a nonlinear model. For this work, the data used for a case study is one-year (1 January 2012 to 31 December 2012) located in a tropical area. The Jacobi algorithm has been employed to construct a prediction model with six combinations of input and one output (target). The predicted target is thunderstorm occurrence. The parameter input is surface pressure, air temperature, relative humidity, clouds, precipitable water vapor, and precipitation. The result obtained a better fit prediction model with four optimum parameters and estimation errors of 5.73%. May and October are the highest occurrences of thunderstorms where prediction errors were found high during the intermonsoon season.

Keywords: Numerical model development, Meteorological data, Thunderstorm prediction, Tropical areas

1. INTRODUCTION

Thunderstorm activities in the atmosphere are of subject concern such as to air transportation business as well as security strategies. This phenomenon of thunderstorms can be categorized as the most dangerous natural disaster because it is one of the main causes of spacecraft being constrained in a launch (Kuk et al., 2012). Thunderstorms are an electrical storm where thunder and lightning are generated by cumulonimbus clouds during heavy rain and precipitation. Lightning is also possible at any stage of cumulonimbus development regardless of precipitation. Subsequently, when the cumulonimbus clouds occur and form a cluster, they develop into a thunderstorm ensemble with a high rainfall index of up to about 100 km or more in one last horizontal direction, and then they are defined as convective mesoscale systems (MCS) (Houze, 2018).

The MCS formed to an altitude of 8 ~ 14 km from small to long-lived with an average maximum area of $160 \times 10^3$ km$^2$ and can be active for up to 13 hours (e.g., Maddox, Rodgers, & Howard, 1982) are a major concern in aviation safety. Therefore, the prediction of thunderstorm occurrence over MCS areas is important to commercial space vehicle launch operation and power utilities. Various methods of estimating thunderstorms occurrence have been conducted by several researchers such as Velden et al. (2006) using Dvorak techniques, and Wilson et al. (1998) using mathematics modeling. Maddox (1980) has been identified the MCSs using early infrared satellite imagery. One of the limitations that appeared of this imagery is difficulty in distinguishing cumulonimbus clouds besides equivalently the occurrence of lightning at the same time. However, in recent years, geostationary satellite imagers, for example, Himawari-8 satellite operated by the Japan Meteorological Agency (JMA) has high-resolution to improve thunderstorm monitoring (Lee et al., 2019). On the other hand, a numerical weather prediction (NWP) system is one of the tools that statistically enhance

1 Institut Teknologi Nasional Yogyakarta, Department of Electrical Engineering, Jl. Babarsari, Depok, Sleman, Yogyakarta 55281, Indonesia, Corresponding author: wayan@itny.ac.id, drwaynesparta@gmail.com
2 Golden Autotech Corp, Department of Research and Innovation, Sidoarjo 61254, East Java, Indonesia, wahyu@gautotech.com
3 Sekolah Tinggi Teknologi Adisutjipto, Department of Informatics, Bantul, Yogyakarta 55198, Indonesia, tkendarmastono@stta.ac.id
thunderstorm forecasts (Simon et al., 2018). This also showed a logistic regression (Gijben et al., 2017) or artificial neural networks (Collins and Tissot, 2015) are capable of accommodating nonlinear relationships between latent variables. Thus, Somesan and Makkai (2016), and Ahmed et al. (2019) studied the thunderstorm frequency distribution based on the analysis of convective available potential energy (CAPE) and convective inhibition (CIN) focused in Transylvania and over Pakistan, respectively. Although trends in CAPE do not explain the variability of thunderstorm over the whole study area explicitly, numerical models will be explored in-depth in this recent work for the prediction of thunderstorm occurrence.

The cold surge and its modification such as vortex formation in the equator region especially located over the Borneo region such as in Tawau, Malaysia is important due to the thunderstorm level are dangerous. The transition of the zonal circulation into maritime tropical circulation will generate synoptic processes (Furtuna, Haidu, & Maier, 2018). This is one of the motivations why Tawau located in the Borneo region near the sea waters is one option. Also, with limited funds for purchasing satellite imagery to improve rainfall data, the thunderstorm study in the Tawau area is carried out by numerical analysis. This study focused on exploring numerical modeling to predict thunderstorms. With high accuracy in predicting the occurrence of thunderstorms, this will facilitate future mitigation.

2. METHODS

2.1 Data and Location

For the case study, six input parameters such as Pressure (P), Temperature (T), Relative Humidity (H), Cloud (C), precipitation (Pr), and precipitable water vapor from Radiosonde (RSPWV) for one year period (1 January ~ 31 December 2012) measurement is employed to predict thunderstorm occurrence. P, T, and H are the basic weather parameters that fundamentally affecting daily life and environments, while precipitation and water vapor are the most significant hydrological parameter that controls water cycle variability in the atmosphere. The selection of these six parameters as input was meant to characterize which parameter is most fit for predicting the occurrence of thunderstorms in this area. The first four parameters (P, T, H, and C) were collected from the Malaysian Meteorological Department (MetMalaysia). RSPWV and Precipitation were taken from the University of Wyoming website (http://weather.uwyo.edu/upperair/sounding.html) and the NASA (Tropical Rainfall Measuring Mission, TRMM) website (http://gdata1.sci.gsfc.nasa.gov), respectively. Parameters for the input data are made daily because the thunderstorm occurrence based on observations is only available on a daily basis.

![Fig. 1. The location of meteorological data at Tawau.](image-url)
Figure 1 shows the location of the case study where the weather station is located in the Tawau airport (TWU: 4.32°N, 118.12°E with elevation 17 m) in Borneo Island. The location of the observation station is closed to the Celebes Sea. The climate conditions in this area are similar to other regions in the equatorial where precipitation throughout the year tends to heavy during November, December, and January. This area is probably experienced high precipitation along the Celebes Sea that has been affected by variations of land-ocean distribution and solar activities (Bian et al., 2015).

2.2 Data Processing

The surface meteorological data on a daily basis is processed. To filter and check the quality of the observation data used, the missing data in their time series is replaced by Not a Number (NaN). Figure 2 shows the flowchart of data processing which applied six parameters for each season. The six combinations parameter is proposed to generate linear equations. Furthermore, the linear equation is transposed to a nonlinear equation using the Jacobi algorithm.

![Flowchart of data processing and identification](image)

Fig. 2. Flowchart of data processing and identification.

Equatorial region has only two seasons (summer or dry and winter or wet), and another one is called intermonsoon. In this study, five seasons are defined as summer (JJA), winter I (JF), winter II (December), transition I (MAM), and transition II (SON) rather than only three seasons. This categorization is to minimize prediction errors because throughout the year there are incomplete daily data for certain parameters. On the other hand, if only 2 or 3 selected seasons for this case, the criteria defined for the thunderstorm occurrence do not match the data observations. From five categorized seasons and six combination parameters, we have 30 linear equations to be proposed to predict thunderstorms. To achieve a stable numerical model, 1,000 iterations are carried out. Validation is carried out in 20% of the raw data. When the error in the training and checking data is below 10% or very small, the designed model can be used to predict the occurrence of thunderstorms.
2.3 Linear Equations

The linear equation is a first order form and is very useful to model any phenomena based on mathematics method. In the general, linear equations can be written in the form of matrix in \( n \times n \) size or \( Ax = b \) (Collins, 2003; Yang and Mittal, 2014). To solve the first until the \( n_{th} \) equations, the multiple linear regression can be formulated in general iteration scheme with the form

\[
x_i^{(k)} = \frac{1}{a_{ii}} \left[ b_i - \sum_{i \neq j}^{n} (a_{ij}x_j^{(k-1)}) \right]
\]

The equation (1) can be solved by the Gauss-Seidel iteration methods. To improve the iteration rate after the value of \( x_i \) is obtained, we can write the iteration scheme as

\[
x_i^{(k)} = \frac{1}{a_{ii}} \left[ b_i - \sum_{j=1}^{i-1} (a_{ij}x_j^{(k)}) - \sum_{j=i+1}^{n} (a_{ij}x_j^{(k-1)}) \right], \quad \text{for } i = 1, 2, 3, \ldots, n
\]

The iteration in equation (2) can produce a faster convergence rate, but there may be problems in the stability of the convergence speed (Collins, 2003). To solve this drawback, the Jacobi iteration method is proposed.

2.4 Jacobi Algorithm

The Jacobi method or Jacobi iterative method can be described as an algorithm for determining solutions of a diagonally dominant system of linear equations (Yang and Mittal, 2014). The method is used for solving systems of multi-dimensional linear equations. If the linear equations in a matrix form split into \( A \), it will obtain as follows.

\[
A = \begin{bmatrix}
a_{11} & 0 & \ldots & 0 \\
0 & a_{22} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & a_{nn}
\end{bmatrix} - \begin{bmatrix}
0 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
-\alpha_{n1} & \ldots & -\alpha_{nn}\end{bmatrix} = D - L - U
\]

If \( b = (D - L - U)x \) and \( D \) is transposed into \( D^T \), the solution of \( x \) from \( Ax = b \) is given by

\[
x = D^{-1} (L + U) x + D^{-1} b
\]

Equation (4) is a matrix form of Jacobi and in the iterative method, it can be expressed as

\[
x^{(k)} = D^{-1} (L + U) x^{(k-1)} + D^{-1} b, \quad \text{for } k = 1, 2, 3, \ldots
\]

Assume that \( T = D^T (L + U) \) and \( c = D^T b \), Jacobi iteration method can also be written as

\[
x^{(k)} = T x^{(k-1)} + c, \quad k = 1, 2, 3, \ldots,
\]

If the initialization initial value \( x_0 = 0 \), the iteration will end if \( x_n^{(k-1)} \approx x_n^k \). With \( n \) iterations, the predicted value obtained will be very stable, and therefore, this method is similar to the neural network algorithm.
3. RESULT AND DISCUSSION

3.1 Meteorological Variation

Figure 3 showed the variation of surface meteorology (P, T and H), Cloud (oktas), RSPWV, and precipitation for the period of 1 January 2012 ~ 31 December 2012 for a case study. The value of P, T, and H is reached ≥ 1010 mbar during April, reached ≥ 28°C during June, August and September, and reached ≥ 85% during January, February, and December, respectively. Furthermore, the extreme value of cloud density (oktas), RSPWV, and precipitation are reached ≥ 7 oktas in June and July, reached ≥ 80 mm in February and December, and reached ≥ 55 mm/day in May and October, respectively.

Fig. 3. The variation of meteorological parameters over the Tawau area. The shaded area represented the categorized seasons. The data used is from Suparta and Putro (2017).

As in Figure 3, a thunderstorm is probably undetectable if the surface pressure reached ≥ 1000 mbar with a temperature ≥ 27°C, the relative humidity value < 80% and cloud density is less than four oktas, although the RSPWV and precipitation occurred more than 40 mm and ≥ 30 mm/day, respectively. However, the precipitation is a minimum while the temperature is higher as marked by clear days in the summer season (JJA). On the other hand, the rainy days over the winter season occurred in January and February (winter I) and December (winter II), where the temperature has decreased while precipitation has increased.

3.2 Thunderstorm Prediction

Based on the meteorological data depicted in Figure 3, a set of linear equation can be generated from six combinations of input: (a) P, (b) P and T, (c) P, T, H, and C, (d) P, T, H, and C, (e) P, T, H, C, and Pr, and (f) P, T, H, C, Pr, and RSPWV. The output target (thunderstorm) is defined as C7 while C1 ~ C6 are predictors represented P, T, H, C, Pr, and RSPWV, respectively. The first predictor selected is P with the assumption that the atmosphere in low pressure can develop a synoptic or mesoscale system.

Table 1 shows the Jacobi equation in the matrix form and the solution obtained. The first matrix with a single column (left side) is regression constants (intercept) before transformed to the Jacobi equation. The second matrix in the middle is the Jacobi equation where the first row is a linear
equation for configuration (a), while the second row is for configuration (b). The third row until the sixth row corresponds to the configuration (c) until configuration (f), respectively. The first to six columns in this matrix are regression coefficients for $x_1 \sim x_6$, respectively. The last a single matrix contained $x_1 \sim x_6$ is predictors. The Jacobi method can also be used to find the best iteration for the six parameter combinations. The maximum iteration for $x_1 \sim x_6$ is presented in five steps with six combinations for each season. To run the iteration, the Python software version 2.7 was used to obtain the prediction model. The best configuration is with a minimum value, both in iterations and the root mean square error (RMSE).

The Jacobi transformation equation for five categorized seasons in 2012

| Season in 2012 | The Jacobi Equation (in a matrix form) | Solution |
|---|---|---|
| Winter I (JF) | \begin{bmatrix} -149.9 \\ -130.0 \\ -128.2 \\ -101.9 \\ -117.8 \\ -105.3 \end{bmatrix} \begin{bmatrix} -0.1484 & 0 & 0 & 0 & 0 & 0 \\ -0.1234 & -0.199 & 0 & 0 & 0 & 0 \\ -0.1229 & -0.177 & 0 & 0 & 0 & 0 \\ -0.0937 & -0.218 & -0.0133 & -0.356 & 0 & 0 \\ -0.1092 & -0.195 & -0.0119 & -0.450 & -0.0112 & 0 \\ -0.0969 & -0.163 & -0.0161 & -0.539 & -0.0122 & -0.0052 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} | \begin{bmatrix} x_1 = 1010.11 \\ x_2 = 26.86 \\ x_3 = 74.05 \\ x_4 = 6.69 \\ x_5 = 6.57 \\ x_6 = -105 \end{bmatrix} |
The optimum input value from six configuration equations for each categorized season found that $X_5$ (RSPWV) or $X_6$ (precipitation) or both are poor results as marked by gray. It is clear from the table that during Winter (DJF) and Transition (MAM and SON), the RSPWV and/or precipitation values are out-of-range. Figure 3 can clarify that the minimum and maximum values of RSPWV and precipitation are $38 \sim 82$ mm and $0 \sim 76$ mm/day, respectively. Furthermore, these two parameters are not used to predict thunderstorms. Therefore, the meteorological parameter to be selected as input for predicting thunderstorm events are $x_1 \sim x_4$. In this model, clouds also play roles in the formation of the MCS, especially on the sea boundary during the late afternoon, where the cloud shield reaches its maximum size in the early morning until the next day.

Figure 4 shows the predicted thunderstorms for each season using the Jacobi method with four input parameters based on the five categorized seasons. The figure shows that the estimation model and the observational data are well corresponded. This can be seen in May and October which are the highest of the occurrence of thunderstorms where this month is categorized as the intermonsoon season. The correlation coefficient ($r$) obtained is 0.75, which is significant at the 95% confidence level. The RMSE obtained was 5.56% and the average percent error was around 44.94%. This moderate relationship appears to be affected by the low accuracy of predictions over the summer (JJA), and this opens up new options for other input combinations for future work. The estimation error was < 10% indicates that the model is acceptable for the prediction of thunderstorms.

![Graph showing thunderstorm occurrence per month](image)

**Fig. 4.** The prediction of thunderstorm occurrence for each season using the Jacobi method and the standard deviation error. The data used is from Suparta and Putro (2017).

To predict thunderstorm occurrence in Figure 4, the rounding number technique was applied. For example, if its output reached $\leq 0.50$ would mean that there is no thunderstorm event (rounded value = 0, or No), while the output with $> 0.50$ indicated that there is a thunderstorm event (rounded value = 1, or Yes). With satisfactory results obtained from configuration (d) with 776 iterations, now the performance of thunderstorm occurrence can be determined by using linear regression equation and expressed as

$$y(x) = 22.17 - 0.0197x_1 - 0.00954x_2 + 0.0055x_3 - 0.0099x_4$$  (7)

where $y(x)$ is the prediction of thunderstorm occurrence. For example, if $P (x_1) = 995$ mbar, $T (x_2) = 26^\circ C$, $H (x_3) = 99\%$, and Clouds ($x_4$) = 5 oktas, then the prediction will be 0.6035 $\approx 1$, which means that a thunderstorm event is occurred.
5. CONCLUSIONS

The prediction of thunderstorm occurrence was successful by using a numerical model from six parameter inputs such as Pressure, Temperature, Relative Humidity, Cloud, RSPWV, precipitation, and one parameter output of thunderstorm. The model generated using the Jacobi method was completely working in predicting the thunderstorm data.

In the process of predicting the occurrence of thunderstorms, as a case study, a regression equation from 2012 observation data in the Tawau area has been generated. There are five equations for each season, such as winter I, transition I, summer, transition II, and winter II, which are processed to find the best input parameter values for the prediction model. Furthermore, this equation is transformed into a nonlinear equation to find the optimum input parameter value using the Jacobi method. Of the six input parameters, RSPWV and precipitation are found out-of-range. The two peaks from March to May (winter to summer) are referred to as the transition I, and September to November (summer to winter) is called the transition II, which are overabundant thunderstorm events. With validation data of 20% (November and December), the estimation model is obtained with an error of 5.73%.

The numerical model with the Jacobi method is advantageous in constructing the prediction of thunderstorm occurrences. In this region, May and October pronounced the highest month of thunderstorm events. For future studies, the numerical model can be examined again with the combination of other inputs to achieve accurate and cost-effective thunderstorm predictions and employ different years of data for comparison. If too many input parameters are required, the model is high-cost. Estimation models that have been formulated only apply to local climates. Different climates and topography of the region will provide different weather parameters for input in predicting thunderstorms. Thus, examining the same parameters elsewhere is necessary, especially if this model can be compared in performance with other models such as soft computing techniques.

Acknowledgements

The author appreciates the support of previous research from the Ministry of Science, Technology, and Innovation (MOSTI) Malaysia under grant 01-01-02-SF1100. The first author was a former Associate Professor at Universiti Kebangsaan Malaysia. We would like to thank the Malaysian Meteorological Department (MetMalaysia) for providing the surface meteorological data, the NASA website for providing the precipitation data, and also the University of Wyoming for providing the Radiosonde PWV data. We appreciate the reviewer and editors comments that substantially improved the manuscript quality.

REFERENCES

Ahmed, R., Latif, M., Adnan, S., & Abusar, M. K. (2019) Thunderstorm frequency distribution and associated convective mechanisms over Pakistan. *Theor. Appl. Climatol.*, 137, 755–773. https://doi.org/10.1007/s00704-018-2619-x.

Bian, Y., Li, J., Jian, Z., Chu, F., Chu, Z., Kuhnt, W., & Ye, L. (2015) Vegetation and climate changes around Celebes Sea during Holocene. *Journal of China University of Geosciences*, 40(5), 870-880, doi: 10.3799/dqkx.2015.071.

Collins, G. W. (2003) Fundamental numerical methods and data analysis. http://ads.harvard.edu/books/1990fmd.book/ [Accessed 13 November 2020]

Collins, W. & Tissot, P. (2015) An artificial neural network model to predict thunderstorms within 400 km² South Texas domains. *Meteor. Appl.*, 22, 650–665, doi:10.1002/met.1499.

Furtuna, P., Haidu, I., & Maier, N. (2018) Synoptic processes generating windthrows. a case study in the Apuseni Mountains (Romania). *Geographia Technica*, 13(2), 52–61, doi: 10.21163/GT_2018.132.04.

Gijben, M., Dyson, L. L., & Loots, M. T. (2017) A statistical scheme to forecast the daily lightning threat over southern Africa using the Unified Model. *Atmos. Res.*, 194, 78–88, doi:10.1016/j.atmosres.2017.04.022.
Houze, R. A. Jr. (2018) 100 years of research on mesoscale convective systems. *Meteor. Monogr.*, 59, 17.1–17.54, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0001.1

Kuk, B., Lim, H., Ha, J., Lee, H., & Lee, G. (2012) A Fuzzy Logic Method for Lightning Prediction Using Thermodynamic and Kinematic Parameters from Radio Sounding Observations in South Korea. *Wea. Forecasting*, 27, 205–217.

Lee, K., Kim, H. S., & Choi, Y. S. (2019) Effects of high-resolution geostationary satellite imagery on the predictability of tropical thunderstorms over Southeast Asia. *Nat. Hazards Earth Syst. Sci.*, 19, 2241–2248, 2019, doi://10.5194/nhess-19-2241-2019.

Maddox, R. A. (1980). Mesoscale convective complexes *Bull. Amer. Meteor. Soc.*, 61, 1374–1387, doi: 10.1175/1520-0477(1980)061,1374:MCC.2.0.CO;2.

Maddox, R. A., Rodgers, D. M., & Howard, K. M. (1982) Mesoscale convective complexes over the United States in 1981: Annual summary. *Mon. Wea. Rev.*, 110, 1501–1514.

Simon, T., Fabsic, P., Mayr, G. J., Umlauf, N., & Zeileis, A. (2018) Probabilistic forecasting of thunderstorms in the Eastern Alps. *Mon. Wea. Rev.*, 146, 2999 - 3009, doi: 10.1175/MWR-D-17-0366.1.

Somesan, R. A. & Makkai, G. (2016) Medium and large hail in Transylvania, locality of band, 24th May 2015. *Geographia Technica*, 11(1), 84 – 91, doi: 10.21163/GT_2016.111.09.

Suparta, W. & Putro, W. S. (2017). Using multiple linear regression model to estimate thunderstorm activity. *IOP Conf. Series: Materials Science and Engineering*, 185, 012023, doi:10.1088/1757-899X/185/1/012023.

Velden, C., B. Harper, F. Wells, J. L. Beven II, R. Zehr, T. Olander, M. Mayfield, C. “Chip” Guard, M. Lander, R. Edson, L. Avila, A. Burton, M. Turk, A. Kikuchi, A. Christian, P. Caroff, P. Mcrone., (2006) The dvorak tropical cyclone intensity estimation technique a satellite-based method that has endured for over 30 years. *Am. Meteorol. Soc.*, 87, 1195-1210.

Wilson, J. M., Crook, N. A., Mueller, C. K., Sun, J., & Dixon, M. (1998) Nowcasting thunderstorms: A status report. *B. Am. Meteorol. Soc.*, 79, 2079–2099.

Yang, X. I. A. & Mittal, R. (2014) Acceleration of the Jacobi iterative method by factors exceeding 100 using scheduled relaxation. *J. Comput. Phys.*, 274, 695–708.