Science on Natural Resources and Environment **54** (2024) 103-118

Science on Natural Resources and Environment
Journal homepage: tapchikhtnmt.hunre.edu.vn

RESEARCH ON 24-HOUR THUNDERSTORM FORECASTING IN SUMMER MONTHS FOR THE DIEN BIEN REGION USING THE STATISTICAL METHOD

Tran Chan Nam[1,*], Pham Van Quan[2], Pham Minh Tien[1]
[1]Hanoi University of Natural Resources and Environment, Vietnam
[2]Air Defense and Air Force Technical College, Vietnam

Received 30 September 2024; Accepted 23 December 2024

Abstract

Thunderstorms are one of the most dangerous weather phenomena, making them a constant subject of study for meteorologists. Weather forecasting in general, and thunderstorm forecasting in particular, have seen significant advancements in Vietnam in recent years. Forecast products from numerical models, radar data, and upper-air observations have improved thunderstorm forecasts. However, in the context of global climate change, natural disasters such as storms, heavy rain, floods, and especially small-scale phenomena like thunderstorms, squalls, tornadoes, and waterspouts have become increasingly complex and unpredictable. Therefore, quantitative forecasting methods are urgently needed to help forecasters make objective assessments to support thunderstorm forecasting operations effectively. Studying thunderstorm forecasting in the summer months for the Dien Bien region using statistical methods will provide us with a more accurate approach to forecasting this phenomenon.

Keywords: Thunderstorms; Dien Bien; Statistical methods.

[*]**Corresponding author, Email:** tcnam@hunre.edu.vn

DOI: http://doi.org/10.63064/khtnmt.2024.644

1. Introduction

Weather forecasting in general and thunderstorm forecasting in Vietnam have seen significant developments in recent years. Thunderstorm forecast products from numerical models, radar data, and satellite data have improved the quality of thunderstorm forecasts. However, in the context of global climate change, natural disasters such as storms, heavy rainfall, and floods, especially small-scale phenomena like thunderstorms, squalls, and tornadoes, are becoming increasingly complex and unpredictable. Therefore, there is a strong need for quantitative forecasting methods to help forecasters make objective assessments to effectively serve thunderstorm forecasting efforts.

In the Northwest region in general and Dien Bien province in particular, weather forecasting still primarily relies on traditional synoptic methods, resulting in forecasts that are heavily qualitative forecasts. Since thunderstorms are one of the dangerous weather phenomena, they have always been the subject of study for many meteorologists. A notable example is Phillip E. Shafer, who used logistic regression techniques to develop a thunderstorm forecasting equation in Florida [1]. In this study, the authors utilized thunderstorm data from an observation network and selected parameters from morning-sounding data to develop same-day afternoon thunderstorm forecast equations at 11 locations on the Florida peninsula. The results demonstrated that the forecast equations exhibited relatively good forecasting skills and operational forecasting potential.

Schmeits M.J et al., [2] employed various statistical methods to analyze output data from a non-hydrostatic model from April to September in 2006 and 2007, observational data from the Lightning Location Information System (LLIS), and radar data (TEPHI). The primary objective of the study was to utilize output from numerical forecast models to predict thunderstorms using binary logistic regression. Four indices were used: The K index, the Showalter index, the SWEAT index, and the total precipitable water. Results indicated that the best models could achieve a Probability of Detection (POD) of approximately 0.9, a False Alarm Ratio (FAR) of around 0.4, and a Critical Success Index (CSI) of about 0.6.

2. Data and research methodology

2.1. Data

The study utilized data collected from 4 - 9 over seven years (2013 - 2019). Specifically:

- Data from 4 observation periods at 3 meteorological stations: Dien Bien, Muong Lay, and Tuan Giao, including thunderstorms, temperature, dew point, humidity, and atmospheric pressure.

- Upper-air data from Dien Bien station at 7:00 AM: Including meteorological elements: Temperature, humidity, pressure, geopotential height, wind, and atmospheric instability index.

From the extracted data, the study identified the following predictive factors.

- Predictive factors from surface observations (Table 1 - primary factors).

- From the set of primary factors, a set of secondary factors can be created. For example, the variation of these factors can be calculated within 24 hours, 18 hours, 12 hours, and 6 hours, as well as calculating the dew point deficit from the observed data (Table 2 - secondary factors).

Table 1. Primary forecasting factors from surface observation data (16 factors)

Obs	**Factor and Symbol**			
	Temperature	**Dew point**	**Humidity**	**Atmospheric pressure**
1	T1	Td1	U1	P1
7	T7	Td7	U7	P7
13	T13	Td13	U13	P13
19	T19	Td19	U19	P19

Table 2. Secondary forecasting factors from surface observation data (68 factors)

No.	Factor	Symbol	No.	Factor	Symbol
1	The 6-hour temperature variation starting from the observation at 1 o'clock	1BT6h	35	The 6-hour humidity variation starting from the observation at 13 o'clock	13BU6h
2	The 6-hour temperature variation starting from the observation at 7 o'clock	7BT6h	36	The 6-hour humidity variation starting from the observation at 19 o'clock	19BU6h
3	The 6-hour temperature variation starting from the observation at 13 o'clock	13BT6h	37	The 12-hour humidity variation starting from the observation at 1 o'clock	1BU12h
4	The 6-hour temperature variation starting from the observation at 19 o'clock	19BT6h	38	The 12-hour humidity variation starting from the observation at 7 o'clock	7BU12h
5	The 12-hour temperature variation starting from the observation at 1 o'clock	1BT12h	39	The 12-hour humidity variation starting from the observation at 13 o'clock	13BU12h
6	The 12-hour temperature variation starting from the observation at 7 o'clock	7BT12h	40	The 12-hour humidity variation starting from the observation at 19 o'clock	19BU12h
7	The 12-hour temperature variation starting from the observation at 13 o'clock	13BT12h	41	The 18-hour humidity variation starting from the observation at 1 o'clock	1BU18h
8	The 12-hour temperature variation starting from the observation at 19 o'clock	19BT12h	42	The 18-hour humidity variation starting from the observation at 7 o'clock	7BU18h
9	The 18-hour temperature variation starting from the observation at 1 o'clock	1BT18h	43	The 18-hour humidity variation starting from the observation at 13 o'clock	13BU18h
10	The 18-hour temperature variation starting from the observation at 7 o'clock	7BT18h	44	The 18-hour humidity variation starting from the observation at 19 o'clock	19BU18h
11	The 18-hour temperature variation starting from the observation at 13 o'clock	13BT18h	45	The 24-hour humidity variation starting from the observation at 1 o'clock	1BU24h
12	The 18-hour temperature variation starting from the observation at 19 o'clock	19BT18h	46	The 24-hour humidity variation starting from the observation at 7 o'clock	7BU24h
13	The 24-hour temperature variation starting from the observation at 1 o'clock	1BT24h	47	The 24-hour humidity variation starting from the observation at 13 o'clock	13BU24h
14	The 24-hour temperature variation starting from the observation at 7 o'clock	7BT24h	48	The 24-hour humidity variation starting from the observation at 19 o'clock	19BU24h
15	The 24-hour temperature variation starting from the observation at 13 o'clock	13BT24h	49	The 6-hour atmospheric pressure variation starting from the observation at 1 o'clock	1BP6h

No.	Factor	Symbol	No.	Factor	Symbol
16	The 24-hour temperature variation starting from the observation at 19 o'clock	19BT24h	50	The 6-hour atmospheric pressure variation starting from the observation at 7 o'clock	7BP6h
17	The 6-hour Dew point temperature variation starting from the observation at 1 o'clock	1BTd6h	51	The 6-hour atmospheric pressure variation starting from the observation at 13 o'clock	13BP6h
18	The 6-hour Dew point temperature variation starting from the observation at 7 o'clock	7BTd6h	52	The 6-hour atmospheric pressure variation starting from the observation at 19 o'clock	19BP6h
19	The 6-hour Dew point temperature variation starting from the observation at 13 o'clock	13BTd6h	53	The 12-hour atmospheric pressure variation starting from the observation at 1 o'clock	1BP12h
20	The 6-hour Dew point temperature variation starting from the observation at 19 o'clock	19BTd6h	54	The 12-hour atmospheric pressure variation starting from the observation at 7 o'clock	7BP12h
21	The 12-hour Dew point temperature variation starting from the observation at 1 o'clock	1BTd12h	55	The 12-hour atmospheric pressure variation starting from the observation at 13 o'clock	13BP12h
22	The 12-hour Dew point temperature variation starting from the observation at 7 o'clock	7BTd12h	56	The 12-hour atmospheric pressure variation starting from the observation at 19 o'clock	19BP12h
23	The 12-hour Dew point temperature variation starting from the observation at 13 o'clock	13BTd12h	57	The 18-hour atmospheric pressure variation starting from the observation at 1 o'clock	1BP18h
24	The 12-hour Dew point temperature variation starting from the observation at 19 o'clock	19BTd12h	58	The 18-hour atmospheric pressure variation starting from the observation at 7 o'clock	7BP18h
25	The 18-hour Dew point temperature variation starting from the observation at 1 o'clock	1BTd18h	59	The 18-hour atmospheric pressure variation starting from the observation at 13 o'clock	13BP18h
26	The 18-hour Dew point temperature variation starting from the observation at 7 o'clock	7BTd18h	60	The 18-hour atmospheric pressure variation starting from the observation at 19 o'clock	19BP18h
27	The 18-hour Dew point temperature variation starting from the observation at 13 o'clock	13BTd18h	61	The 24-hour atmospheric pressure variation starting from the observation at 1 o'clock	1BP24h
28	The 18-hour Dew point temperature variation starting from the observation at 19 o'clock	19BTd18h	62	The 24-hour atmospheric pressure variation starting from the observation at 7 o'clock	7BP24h
29	The 24-hour Dew point temperature variation starting from the observation at 1 o'clock	1BTd24h	63	The 24-hour atmospheric pressure variation starting from the observation at 13 o'clock	13BP24h
30	The 24-hour Dew point temperature variation starting from the observation at 7 o'clock	7BTd24h	64	The 24-hour atmospheric pressure variation starting from the observation at 19 o'clock	19BP24h

No.	Factor	Symbol	No.	Factor	Symbol
31	The 24-hour Dew point temperature variation starting from the observation at 13 o'clock	13BTd24h	65	Dew point deficit at 1 o'clock observation	HTd1h
32	The 24-hour Dew point temperature variation starting from the observation at 19 o'clock	19BTd24h	66	Dew point deficit at 7 o'clock observation	HTd7h
33	The 6-hour humidity variation starting from the observation at 1 o'clock	1BU6h	67	Dew point deficit at 13 o'clock observation	HTd13h
34	The 6-hour humidity variation starting from the observation at 1 o'clock	7BU6h	68	Dew point deficit at 19 o'clock observation	HTd19h

Predictive factors from high non-primary data in Table 3

Table 3. Primary forecast factors from the atmospheric instability index (7 factors)

Obs	Factor and Symbol						
	SI index	**LI index**	**SWEAT index**	**K index**	**T index**	**CAPE index**	**CIN index**
7	7SI	7LI	7SWEAT	7K	7TT	7CAPE	7CIN

Similarly, we can create a secondary factor set from Table 3 (Table 4 - secondary factors)

Table 4. Secondary predictive factors from data (7 factors)

No.	Factor	Symbol	No.	Factor	Symbol
1	24-hour SI variation at the 7 am observation	B7SI24h	5	24-hour TT variation at the 7 am observation	B7TT24h
2	24-hour LI variation at the 7 am observation	B7LI24h	6	24-hour CAPE variation at the 7 am observation	B7CAPE24h
3	24-hour SWEAT variation at the 7 am observation	B7SWEAT24h	7	24-hour CIN variation at the 7 am observation	B7CIN24h
4	24-hour K variation at the 7 am observation	B7K24h			

Thus, for Dien Bien, Muong Lay station, a total of 98 factors are involved in constructing the thunderstorm forecasting equation. For Tuan Giao station, due to the lack of pressure observation, factors related to atmospheric pressure are excluded from the study, resulting in a total of 78 factors involved in constructing the thunderstorm forecasting equation.

2.2. Methodology

2.2.1. Binary Logistic Regression

Binary logistic regression uses a binary dependent variable to estimate the probability of an event occurring given the information of independent variables [1].

For the forecasting problem of thunderstorm occurrence in this study, the event of having a thunderstorm is assigned a value of 1, and no thunderstorm is assigned a value of 0.

With binary logistic regression, the information we need to collect about the dependent variable is whether or not an event occurs. The dependent variable Y at this point has a value of 0 (none thunderstorm) and 1 (thunderstorm), and information about the independent variables X. From this binary dependent variable, a procedure is used to predict the probability of the event occurring according to the rule: If the predicted probability is greater than 0.5, the predicted result will be considered as "yes" the event occurs, otherwise, the predicted result will be "no" if the event occurs.

2.2.2. Forecast evaluation

For binary prediction problems such as thunderstorm forecasting, categorical statistics are commonly used to assess the frequency of occurrence of the event. Evaluation metrics are derived from a contingency table (Table 5).

Table 5. Contingency table for binary forecast

		Observation	
		Yes	**No**
Prediction	**Yes**	H	F
	No	M	CN

Or you can interpret the following letter in detail: (i) Hits (H) prediction + observation; (ii) Misses (M) = None prediction + Observation; (iii) False alarms (F) = Prediction + None observation: và (iv) Correct negatives (CN) = None prediction + None observation.

- Based on this frequency table and assuming there are N data records studied (H+M+F+CN=N), some evaluation indicators are set up to evaluate the forecasting skills of the forecasting models for the binary forecasting factor including the index

- PC/FC index (Percent Correct/ Fraction Correct): PC reflects the coincidence rate between the results of the model and the observation in both phases with and without the phenomenon. The value of the PC varies between 0 and 1.

- FBI index (BS or FBI- Bias score): is the score between the forecast zone and the observation zone. The FBI is the ratio between the number of times a phenomenon occurs according to the model and the observation. The FBI value varies between 0 and $+\infty$

- POD index (Probability of Detection): is the probability of a phenomenon, equal to the ratio between the number of matches between the model and the observation when the phenomenon occurs (hits) and the total number of occurrences of the phenomenon in reality.

- FAR index (False Alarm Ratio): This is the ratio of false alarms to the total number of forecasts indicating no event.

- MR/POFD index (Miss Rate/ Probability of False Detection): This is the ratio of misses to the total number of times the event did not occur.

- TS/CSI index (Threat Score/ Critical Success Index): This metric reflects the relationship between the number of forecasts indicating the event occurred and the number of times the event occurred.

3. Results

3.1. Development of forecast equations

Based on the presented methods and collected data, the article conducts calculations and selects a 24-hour thunderstorm forecast equation for 3 stations in Dien Bien province, relying on model fit, testing coefficients, and overall model fit. The results of developing thunderstorm forecast equations for the three stations in the Dien Bien area in April, May, June, July, August, and September are presented in Tables 6, 7, and 8.

a) For Dien Bien station

Table 6. Thunderstorm forecast equations at Dien Bien station from April to September

Month	Forecast equations
4	$P_i=\frac{e^{z(-171.815 + 0.508*T7 + 0.163*Td19 + 0.165*P7 - 0.179*1BTd12h - 0.194*7SI)}}{1+e^{z(-171.815 + 0.508*T7 + 0.163*Td19 + 0.165*P7 - 0.179*1BTd12h - 0.194*7SI)}}$
5	$P_i=\frac{e^{z(-63.719 + 7.182*Td1 - 4.197*U1 - 0.987*7BTd6h + 0.012*7CIN + 0.383*7TT)}}{1+e^{z(-63.719 + 7.182*Td1 - 4.197*U1 - 0.987*7BTd6h + 0.012*7CIN + 0.383*7TT)}}$
6	$P_i=\frac{e^{z(-33 + 0.757*Td1 + 0.567*P7 - 0.55*P13 - 0.283*1BU24h - 0.012*7SWEAT)}}{1+e^{z(-33 + 0.757*Td1 + 0.567*P7 - 0.55*P13 - 0.283*1BU24h - 0.012*7SWEAT)}}$
7	$P_i=\frac{e^{z(-0.1 - 0.346*U7 + 0.319*U13 - 0.19*1BT6h - 0.736*1BT24h + 0.872*1BTd24h)}}{1+e^{z(-0.1 - 0.346*U7 + 0.319*U13 - 0.19*1BT6h - 0.736*1BT24h + 0.872*1BTd24h)}}$
8	$P_i=\frac{e^{z(-19.640 + 0.747*T1 + 0.538*7BT24h + 0.162*13BU18h + 0.752*7BP6h - 0.286*19BP12h)}}{1+e^{z(-19.640 + 0.747*T1 + 0.538*7BT24h + 0.162*13BU18h + 0.752*7BP6h - 0.286*19BP12h)}}$
9	$P_i=\frac{e^{z(-28.598 - 1.426*T1 + 1.345*Td1 + 1.103*U7 + 0.663*7LI + 0.223*B7TT24H)}}{1+e^{z(-28.598 - 1.426*T1 + 1.345*Td1 + 1.103*U7 + 0.663*7LI + 0.223*B7TT24H)}}$

From Table 6, it can be seen that the factors involved in the thunderstorm forecast equation for the Dien Bien area in April include T7, Td19, P7, 1BTd12h, and 7SI; in May: Td1, U1, 7BTd6h, 7CIN, and 7TT; in June: Td1, P7, P13, 1BU24h, and 7SWEAT; in July: U7, U13, 1BT6h, 1BT24h, and 1BTd24h; in August: T1, 7BT24h, 13BU18h, 7BP6h, and 19BP12h; And in September: T1, Td1, U7, 7LI, and B7TT24H. The factors involved in the forecast equation include both surface and upper-air meteorological factors.

b) For Muong Lay station

Table 7. Thunderstorm forecast equations at Muong Lay station from April to September

Month	Forecast equations
4	$P_i=\frac{e^{z(-15.028 + 0.329*Td7 - 0.431*1BT18h + 0.387*1BT24h + 0.228*1BP24h + 0.194*7TT)}}{1+e^{z(-15.028 + 0.329*Td7 - 0.431*1BT18h + 0.387*1BT24h + 0.228*1BP24h + 0.194*7TT)}}$
5	$P_i=\frac{e^{z(-3.266 - 0.879*Td7 + 0.383*U1 + 0.344*7K - 0.007*B7SWEAT24H - 0.086*B7TT24H)}}{1+e^{z(-3.266 - 0.879*Td7 + 0.383*U1 + 0.344*7K - 0.007*B7SWEAT24H - 0.086*B7TT24H)}}$
6	$Pi=\frac{e^{z(1.325 - 0.33*19BU16h - 0.732*HTd1h + 2.189*HTd7h - 0.014*7SWEAT + 0.001*7CAPE)}}{1+e^{z(1.325 - 0.33*19BU16h - 0.732*HTd1h + 2.189*HTd7h - 0.014*7SWEAT + 0.001*7CAPE)}}$
7	$P_i=\frac{e^{z(-0.897 + 0.651*1BT24h - 1.521*B7SI24H - 1.822*B7LI24H - 1.573*B7TT24H - 0.003*B7CAPE24H)}}{1+e^{z(-0.897 + 0.651*1BT24h - 1.521*B7SI24H - 1.822*B7LI24H - 1.573*B7TT24H - 0.003*B7CAPE24H)}}$
8	$P_i=\frac{e^{z(-97.175 + 10.722*Td1 + 0.28*U19 - 12.525*1BTd18h + 6.77*1BU18h - 5.659*U1)}}{1+e^{z(-97.175 + 10.722*Td1 + 0.28*U19 - 12.525*1BTd18h + 6.77*1BU18h - 5.659*U1)}}$
9	$P_i=\frac{e^{z(-368.192 + 1.392*T7 + 0.34*P1 - 16.509*1BTd6h + 9.797*1BU6h + 0.002*7CAPE)}}{1+e^{z(-368.192 + 1.392*T7 + 0.34*P1 - 16.509*1BTd6h + 9.797*1BU6h + 0.002*7CAPE)}}$

From Table 7, it can be seen that the factors involved in the thunderstorm forecast equation for the Muong Lay area in April include Td7, 1BT18h, 1BT24h, 1BP24h, and 7TT; In May: Td7, U1, 7K, B7SWEAT24H, and B7TT24H; In June: 19BU16h, HTd1h, HTd7h, 7SWEAT, and 7CAPE; In July: 1BT24h, B7SI24H, B7LI24H, B7TT24H, and B7CAPE24H; In August: Td1, U19, 1BTd18h, 1BU18h, and U1; And in September: T7, P1, 1BTd6h, 1BU6h, and 7CAPE. The factors involved in the forecast equation include both surface and upper-air observations.

c) For Tuan Giao station

Table 8. Presents the equations used to forecast thunderstorms at Tuan Giao station from April to September

Month	Forecast equations
4	$P_i=\frac{e^{z(-17.157 + 0.261*T1 + 0.184*Td19 + 3.089*1BTd18h - 2.321*1BU18h + 0.164*7TT)}}{1+e^{z(-17.157 + 0.261*T1 + 0.184*Td19 + 3.089*1BTd18h - 2.321*1BU18h + 0.164*7TT)}}$
5	$P_i=\frac{e^{z(-39.913 + 0.961*7SI + 0.276*7K + 0.680*7TT -0.182*B7K24H + 0.278*1BT18h)}}{1+e^{z(-39.913 + 0.961*7SI + 0.276*7K + 0.680*7TT -0.182*B7K24H + 0.278*1BT18h)}}$
6	$Pi=\frac{e^{z(-22.015 + 0.508*T13 - 0.348*13BT6h + 0.519*13BTd6h - 0.661*HTd1h + 0.243*7K)}}{1+e^{z(-22.015 + 0.508*T13 - 0.348*13BT6h + 0.519*13BTd6h - 0.661*HTd1h + 0.243*7K)}}$
7	$P_i=\frac{e^{z(-29.622 + 1.154*Td19 - 0.939*19BTd12h + 1.681*7SI - 1.805*7LI - 0.61*B7SI24H)}}{1+e^{z(-29.622 + 1.154*Td19 - 0.939*19BTd12h + 1.681*7SI - 1.805*7LI - 0.61*B7SI24H)}}$
8	$P_i=\frac{e^{z(20.397 + 0.219*T13 - 0.188*Td1 - 4.906*Td19 + 0.378*U13 + 2.797*U19)}}{1+e^{z(20.397 + 0.219*T13 - 0.188*Td1 - 4.906*Td19 + 0.378*U13 + 2.797*U19)}}$
9	$P_i=\frac{e^{z(-24.653 + 1.089*Td19 + 0.985*1BTd6h + 1.506*B7SI24H + 0.856*B7TT24H + 0.002*B7CAPE24H)}}{1+e^{z(-24.653 + 1.089*Td19 + 0.985*1BTd6h + 1.506*B7SI24H + 0.856*B7TT24H + 0.002*B7CAPE24H)}}$

Based on Table 8, the factors influencing the thunderstorm forecasting equation for the Tuan Giao area vary by month. For April, these factors include T1, Td19, 1BT18h, 1BU18h, and 7TT. In May, the influential factors are 7SI, 7K, 7TT, B7K24H, and 1BT18h. For June, the equation considers T13, 13BT6h, 13BTd6h, HTd1h, and 7K. July's equation involves Td19, 19BTd12h, 7SI, 7LI, and B7SI24H. August's equation uses T13, Td1, Td19, U13, and U19. Finally, September's equation incorporates Td19, 1BTd6h, B7SI24H, B7TT24H, and B7CAPE24H.

3.2. Forecast testing

The forecasting capability of these equations is evaluated through the presented evaluation criteria. The dependent data series used for evaluation was taken from 2013 to 2018, while the independent data series used for calculation was taken in 2019.

The evaluation of forecast quality for these equations is presented as follows:

a) Dien Bien station

The results of the forecast quality evaluation at Dien Bien station on the dependent data series are presented in Table 9.

Table 9. Thunderstorm forecast results on the data chain dependent on Dien Bien station

Station	Month	Frequency				Total	Evaluation criteria						
		H	F	M	CN		PC	% Prediction	FBI	POD	FAR	MR	TS
Dien Bien	4	41	39	15	46	141	0.62	62	1.43	0.73	0.49	0.46	0.43
	5	28	12	15	53	108	0.75	75	0.93	0.65	0.3	0.18	0.51
	6	4	2	39	99	144	0.72	72	0.14	0.09	0.33	0.02	0.09
	7	23	33	25	105	186	0.69	69	1.17	0.48	0.59	0.24	0.28
	8	28	19	37	102	186	0.7	70	0.72	0.44	0.40	0.16	0.33
	9	14	21	7	84	126	0.78	78	1.67	0.67	0.6	0.2	0.33

Based on Table 9, for Dien Bien station, the PC index of the forecast equations is relatively high, with values ranging from 0.62 to 0.78, and the highest PC value is for the forecast equation in September. This indicates that the coincidence rate between the model results and observations in both phases with and without the occurrence of the phenomenon from the constructed forecasts is relatively large.

The FBI index of the forecast equations in the Dien Bien area for the dependent data series in May, June, and August is less than 1, indicating that the equations produce many missed forecasts, especially for the forecast equation in June. Conversely, the forecast equations in April, July, and September are greater than 1, indicating that the equations produce many false alarms, especially the thunderstorm forecast equation for September. The forecast equations in May and July have FBI values very close to 1, meaning that the forecast regions of these models are relatively consistent with the observation regions.

The POD of the forecast equations in April, May, and September is relatively higher than the other months, with corresponding values of 0.73, 0.65, and 0.67. Conversely, the POD of the forecast equations in June, July, and August are relatively lower compared to the other months, indicating that for the calculation data series, the success rate of these equations is not high. However, it should also be noted that POD is only sensitive to undetected phenomena, not to false detections. Therefore, to comprehensively assess the equation quality, it is necessary to evaluate two indices: FAR and MR.

The FAR of the thunderstorm forecast equations in April, May, June, and August is relatively low, with the lowest being in May; The FAR value of the forecast equation in April is 0.49, indicating that the equation produces false alarms that are equivalent to cases where the occurrence of thunderstorms is correctly predicted (the number of false alarms is only slightly smaller than the number of correct forecasts). The thunderstorm forecast equations for July and September are greater than 0.5, indicating that these equations produce more false alarms than correct forecasts of thunderstorm occurrence.

The MR index of the forecast equations is quite small, ranging from 0.02 to 0.46, and it can be seen that the MR of the forecast equation in June is 0.02, which is very close to the optimal value, indicating that the impact of false alarms from this equation is almost zero for the calculation data series (in this case, it is necessary to pay attention to the F phase in the data table, the equation only gives 2 false alarms (forecast yes, observation no) out of a total of 144 forecasts), however, the miss rate of this equation is relatively high, leading to a low success score (TS) and detection probability.

The TS index values of the forecast equations in April and May are relatively higher than the other months. This indicates that, for the "occurrence of the phenomenon" phase, the forecast equation in April and May performs better than the equations in other months, and the best is the forecast equation in May. However, it should also be emphasized that for this index, the cases of "no occurrence of the phenomenon" have been ignored, and it only considers the ratio of correct forecasts to the total number of observations and forecasts with the occurrence of the phenomenon.

In summary, when comparing the calculated indices from the forecast equations for the dependent data series in the Dien Bien area, it can be seen that the thunderstorm forecast equations in the considered months can all forecast thunderstorms with a certain degree of accuracy. However, for a more objective evaluation, it is necessary to evaluate the equations with the independent data set in 2019 to draw the most objective conclusions. The evaluation results of the forecast models are presented in Table 10.

Table 10. Thunderstorm forecast results on the independent data series of Dien Bien station (in 2019)

Station	Month	Frequency				Total	Evaluation criteria						
		H	F	M	CN		PC	% Prediction	FBI	POD	FAR	MR	TS
Dien Bien	4	9	9	3	8	29	0.59	59	1.5	0.75	0.5	0.53	0.43
	5	2	9	4	16	31	0.58	58	1.83	0.33	0.82	0.36	0.13
	6	1	5	7	16	29	0.59	59	0.75	0.13	0.83	0.24	0.08
	7	4	5	5	17	31	0.68	68	1	0.44	0.56	0.23	0.29
	8	7	9	2	13	31	0.64	64	1.78	0.78	0.56	0.41	0.39
	9	2	4	3	19	28	0.75	75	1.2	0.4	0.67	0.17	0.22

As shown in Table 10, the calculated evaluation metrics based on the independent data series exhibit significant differences in the performance of thunderstorm forecast equations for the Dien Bien region across different months.

For the April thunderstorm forecast equation, the PC index is higher only compared to the May equation, but the POD index ranks second (only after the POD of the August equation) and the TS index is the highest among all equations. The FBI index being greater than 1 indicates a tendency of the equation to produce false alarms, however, the FAR of this equation is the lowest among other equations, suggesting that the rate of producing false alarms is the lowest. The MR of this equation is relatively high, reaching 0.53.

The May thunderstorm forecast equation has the lowest PC accuracy compared to the other equations, and the highest FBI, indicating that this equation tends to produce false alarms. This is also evident in the relatively high FAR of the equation, reaching 0.82, second only to the FAR of the June forecast equation.

The POD accuracy and success score (TS) of the equation are also not high. Therefore, this equation can be excluded from the selection process.

The FBI of the June thunderstorm forecast equation is less than 1, indicating that the equation tends to miss events. The POD and TS indices of the equation are the lowest compared to other equations. The highest FAR indicates that the rate of producing false alarms of this equation is much higher than the correct forecasts of thunderstorm occurrence. The MR of the equation is relatively low, but this value may be due to the tendency of the equation to miss events.

The July and September thunderstorm forecast equations have relatively high PC indices, and the FBI values of both equations are close to 1, indicating that the forecast area of the equation and the observed area are relatively consistent. The FAR and MR of these equations are also relatively low. However, the probability of detection and success score of these two equations are not significantly higher than other options. When considering the contingency table, it can be seen that the number of false alarms and missed events of these equations is much larger than the correct forecasts of thunderstorm occurrence.

Although the August thunderstorm forecast equation has an average PC accuracy compared to other equations, the probability of detection (POD) of this equation is the highest, and the TS is also relatively high, second only to the April thunderstorm forecast equation. The FAR of this equation is also relatively low. The MR of both the April and August forecast equations is quite high due to the relatively high rate of false alarms compared to the correct forecasts of no thunderstorm occurrence.

In conclusion, based on the above evaluation analysis, the April and August thunderstorm forecast equations can be selected for forecasting thunderstorms in the Dien Bien station area. For May and June, other forecasting methods should be used to obtain the most accurate forecast results. For the July and September forecast equations, further studies with longer data series are needed to accurately assess their forecasting capability.

b) Muong Lay station

The results of the evaluation of forecast quality at Muong Lay station on the dependent data series are presented in Table 11.

Table 11. Thunderstorm forecast results on the data chain dependent on Muong Lay station

Station	**Month**	**Frequency**				**Total**	**Evaluation criteria**						
		H	**F**	**M**	**CN**		**PC**	**% Prediction**	**FBI**	**POD**	**FAR**	**MR**	**TS**
Muong Lay	4	40	20	21	60	141	0.71	71	0.98	0.66	0.33	0.25	0.49
	5	50	21	19	50	140	0.71	71	1.03	0.72	0.3	0.3	0.56
	6	32	15	17	50	114	0.72	72	0.96	0.65	0.32	0.23	0.5
	7	9	3	19	56	87	0.75	75	0.43	0.32	0.25	0.05	0.29
	8	32	20	40	94	186	0.68	68	0.72	0.44	0.38	0.18	0.35
	9	14	15	6	102	137	0.85	85	1.45	0.7	0.52	0.13	0.4

From Table 11, we can see that the PC index of thunderstorm forecasting equations for the Muong Lay area is quite high, with the highest value for the September thunderstorm forecasting equation, reaching 0.85, very close to the optimal value. The lowest value is in August, with a PC index of 0.68.

The FBI index of thunderstorm forecasting equations in May and September is slightly greater than 1, indicating that these equations produce a relatively large number of false alarms. The forecasting equations for April, June, July, and August produce more misses, hence the FBI of these equations is less than 1; In particular, July has the lowest FBI as the equation yields the highest proportion of misses. In April, May, and June, the forecasting equations have FBI values very close to 1, indicating that the forecast regions of the equations are relatively consistent with the observed regions.

The POD of thunderstorm forecasting equations in May and September is higher than in other months due to a much larger number of correctly detected cases compared to the number of missed cases. Conversely, in July and August, the forecasting equations have lower POD values than other months, as these equations produce relatively many misses.

The FAR of the equations is relatively small, with the highest FAR being that of the forecasting equation in September and the lowest for the July thunderstorm forecasting equation. This indicates that for the dependent data series, the September thunderstorm forecasting equation produces relatively many false alarms, while the forecasting equations for other months produce fewer false alarms compared to the correct detections of thunderstorms.

The MR of the forecasting equations is also relatively small, ranging from about 0.05 to 0.3. The lowest MR is for the July thunderstorm forecasting equation, with an MR of 0.05, very close to the optimal value. However, it should also be noted that the low MR of the July thunderstorm forecasting equation is due to the equation's tendency to produce misses, hence the indices such as POD and TS are also very low.

The TS of the forecasting equations in April, May, and June is relatively higher than in other months. This indicates that the forecasting equations in April, May, and June have better forecasting skills compared to the forecasting equations in July, August, and September for the "occurrence" phase.

To assess more objectively the forecasting capability of the thunderstorm forecasting equations for Muong Lay station, testing on the independent data series in 2019 is needed, and the results are shown in Table 12.

The May thunderstorm forecasting equation has a moderate PC accuracy. The FBI index is very high, reaching 2.13, indicating that the equation mainly produces false alarms. Although the POD of the equation is relatively high, the FAR and MR of the equation are also very high, indicating that the high probability of detection of the equation is due to the equation's tendency to produce false alarms, which leads to a relatively low TS score and PC accuracy despite the high POD.

The June thunderstorm forecasting equation has a rather low PC accuracy. The POD and TS indices of this equation are also relatively low, and the FAR and MR indices are not superior to those of forecasting equations in other months. Therefore, this equation can be eliminated during the selection of a forecasting option for the Muong Lay area.

Table 12. Thunderstorm forecast results on the independent data series of Muong Lay station (in 2019)

Station	**Month**	**Frequency**				**Total**	**Evaluation criteria**						
		H	**F**	**M**	**CN**		**PC**	**% Prediction**	**FBI**	**POD**	**FAR**	**MR**	**TS**
Muong Lay	4	7	7	3	12	29	0.66	66	1.4	0.7	0.5	0.37	0.41
	5	5	12	3	10	30	0.5	50	2.13	0.63	0.71	0.55	0.25
	6	4	5	10	10	29	0.48	48	0.64	0.29	0.56	0.33	0.21
	7	1	4	8	17	30	0.6	60	0.56	0.11	0.8	0.19	0.08
	8	7	8	6	10	31	0.55	55	1.15	0.54	0.53	0.44	0.33
	9	1	4	5	19	29	0.69	69	0.83	0.17	0.8	0.17	0.1

The forecasting equations for July and September have relatively high PC indices, however, the probability of detection (POD) and success index (TS) of these two equations are quite low compared to other equations. The false alarm ratio (FAR) of these two equations is very high, reaching 0.8, indicating that these equations produce significantly more false alarms than correct forecasts. The miss ratio (MR) of these two equations is relatively low, suggesting that the equations perform better in predicting the absence of thunderstorms. However, the relatively good MR score might be due to the tendency of these equations to miss forecasts, as shown by the bias in the failure rate (FBI) which is less than 1, and the frequency table showing more missed forecasts than correct forecasts of thunderstorms.

The forecasting equations for April and August have moderate accuracy but their probability of detection (POD) and success index (TS) are relatively higher compared to other equations. Meanwhile, the false alarm ratio (FAR) and miss ratio (MR) of these two equations are also at a moderate level.

Therefore, the forecasting equation for April and August can be selected for thunderstorm forecasting in the Muong Lay station area. For other months, other approaches should be considered to make the most accurate forecasts

c) Tuan Giao station

The results of the assessment of forecast quality at Tuan Giao station on the dependent data series are presented in Table 13.

Table 13. Thunderstorm forecast results on the data chain dependent on Tuan Giao station

Station	Month	Frequency				Total	Evaluation criteria						
		H	F	M	CN		PC	% Prediction	FBI	POD	FAR	MR	TS
Tuan Giao	4	16	15	32	78	141	0.67	67	0.65	0.33	0.48	0.16	0.25
	5	25	16	29	70	140	0.68	68	0.76	0.46	0.39	0.19	0.36
	6	35	24	25	60	144	0.66	66	0.98	0.58	0.41	0.29	0.42
	7	17	6	8	56	87	0.84	84	0.92	0.68	0.26	0.1	0.55
	8	29	15	39	103	186	0.71	71	0.65	0.43	0.34	0.13	0.35
	9	20	37	2	67	126	0.69	69	2.59	0.91	0.65	0.36	0.34

The PC index of the forecast equations in the Tuan Giao area is also relatively high when calculated with the dependent dataset, with the lowest value of 0.66 for the June forecast equation and the highest of 0.84 for the July forecast equation. Also for the July forecast equation, the TS index is the highest compared to other forecast equations; The POD index is only smaller than the September thunderstorm forecast equation; The FAR and MR indices are the smallest compared to other equations and close to the optimal value. When considering the contingency table, it can be seen that the number of false alarms and missed forecasts is very small.

The FBI index of most forecast equations is less than 1, indicating that the equations tend to produce missed forecasts. However, it can also be seen that some forecast equations have FBI values very close to 1, such as the June and July thunderstorm forecast equations, indicating that the forecast area of these equations almost coincides with the observed area. The FBI index of the September thunderstorm forecast equation reaches 2.57, indicating that the equation produces a lot of false alarms.

As mentioned above, the probability of detection (POD) of the July and September forecast equations is relatively higher compared to other forecast equations. The reason is that these equations produce very few missed forecasts compared to correct forecasts of thunderstorm occurrence. However, for the September thunderstorm forecast equation, the low POD is due to the equation's tendency to produce false alarms, leading to a very high POD but also the highest FAR and MR indices compared to other equations, and the TS index is relatively low.

The FAR and MR values of the forecast equations are also quite low, except for the September thunderstorm forecast equation. As mentioned above, the lowest MR and FAR are for the July thunderstorm forecast equation and the highest for the September thunderstorm forecast equation.

Similar to the Muong Lay and Dien Bien stations, after calculating the dependent data series at Tuan Giao station, it is necessary to calculate the independent data series in 2019 to examine the forecasting capability of the forecast equations, and then select the optimal forecast equation for this area.

Table 14. Thunderstorm forecast results on the independent data series of Tuan Giao station (in 2019)

Station	Month	Frequency				Total	Evaluation criteria						
		H	F	M	CN		PC	% Prediction	FBI	POD	FAR	MR	TS
Tuần Giáo	4	2	4	7	16	29	0.62	62	0.67	0.22	0.67	0.2	0.15
	5	3	3	7	17	30	0.67	67	0.6	0.3	0.5	0.15	0.23
	6	3	9	6	11	29	0.48	48	1.33	0.33	0.75	0.45	0.17
	7	1	5	6	18	30	0.63	63	0.86	0.14	0.83	0.22	0.08
	8	3	6	9	13	31	0.52	52	0.75	0.25	0.67	0.32	0.17
	9	1	5	2	20	28	0.75	75	2	0.33	0.83	0.2	0.13

Based on the calculations using the independent dataset, the thunderstorm forecast equation for May is the most suitable for selecting as the thunderstorm forecast scheme for the Tuan Giao area. This is because this equation has a relatively high accuracy, with a PC of 0.67, second only to the forecast equation for September (the September thunderstorm forecast equation has a fairly good skill in forecasting the phase where the phenomenon does not occur, but for the phase where the phenomenon occurs, this equation shows a relatively poor forecasting skill). The POD of this equation is at a moderate level; the TS is the largest compared to other forecast equations; and at the same time, the FAR and MR indices are the smallest. The other equations have a relatively low success score, the probability of detecting the phenomenon is not significantly higher, and the FAR index is too high, indicating that false alarms account for a much higher proportion than correct forecasts of the occurrence of thunderstorms. When examining the frequency table, it can be seen that the equations produce many more false alarms and missed forecasts than correct forecasts of the occurrence of thunderstorms. Therefore, this thesis selects the May forecast equation as the thunderstorm forecast scheme for the Tuan Giao station area.

3.3. Results of selection of thunderstorm forecasting equation for Dien Bien area

From the above results, the thesis selects two suitable equations to forecast thunderstorms for the Dien Bien area and its surroundings, including:

a) Dien Bien station

** April*

$$P_i=\frac{e^{z(-171.815 + 0.508*T7 + 0.163*Td19 + 0.165*P7 - 0.179*1BTd12h - 0.194*7SI)}}{1+e^{z(-171.815 + 0.508*T7 + 0.163*Td19 + 0.165*P7 - 0.179*1BTd12h - 0.194*7SI)}}$$

a2) August:

$$P_i=\frac{e^{z(-19.640 + 0.747*T1 + 0.538*7BT24h + 0.162*13BU18h + 0.752*7BP6h - 0.286*19BP12h)}}{1+e^{z(-19.640 + 0.747*T1 + 0.538*7BT24h + 0.162*13BU18h + 0.752*7BP6h - 0.286*19BP12h)}}$$

b) Muong Lay station

** April*

$$P_i=\frac{e^{z(-15.028 + 0.329*Td7 - 0.431*1BT18h + 0.387*1BT24h + 0.228*1BP24h + 0.194*7TT)}}{1+e^{z(-15.028 + 0.329*Td7 - 0.431*1BT18h + 0.387*1BT24h + 0.228*1BP24h + 0.194*7TT)}}$$

b2) August:

$$P_i=\frac{e^{z(-97.175 + 10.722*Td1 + 0.28*U19 - 12.525*1BTd18h + 6.77*1BU18h - 5.659*U1)}}{1+e^{z(-97.175 + 10.722*Td1 + 0.28*U19 - 12.525*1BTd18h + 6.77*1BU18h - 5.659*U1)}}$$

c) Tuan Giao station

** May*

$$P_i = \frac{e^{z(-39.913 + 0.961*7SI + 0.276*7K + 0.680*7TT - 0.182*B7K24H + 0.278*1BT18h)}}{1 + e^{z(-39.913 + 0.961*7SI + 0.276*7K + 0.680*7TT - 0.182*B7K24H + 0.278*1BT18h)}}$$

4. Conclusion

From the research, it can be seen that the developed equations have an acceptable level of accuracy and can be applied in practical forecasting activities. Specifically, the thunderstorm forecasting equations for April and August in the Dien Bien area have accuracies of approximately 62 % and 70 %, respectively; The thunderstorm forecasting equations for April and August in the Muong Lay area have accuracies of approximately 71 % and 68 %, respectively; And the May thunderstorm forecasting equation in the Tuan Giao area has an accuracy of approximately 68 %. Most of the equations involve both observed factors and instability indices.

REFERENCES

[1]. Phillip e. Shafer and Henry e. Fuelberg (2005). *A statistical procedure to forecast warm season lightning over portions of the Florida peninsula*. Shafer and Fuelberg, pp. 851 - 868.

[2]. Schmeits M.J, Kok K.J, Vogelezang D.H.P., (2008). *Probabilistic forecasts of (Severe) thunderstorms for the purpose of issuing a weather alarm in the Netherlands.* Weather Forecast, 23: 1253 - 1267.

[3]. Wilks D. S., (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press, Second Edition, 649 pp.

[4]. Schmeits, Maurice J. et al., (2005). *Probabilistic Forecasting of (Severe) Thunderstorms in the Netherlands Using Model Output Statistics.* Weather and Forecasting, Vol. 20, No2., pp. 134-148.

[5]. Stephen P. W. Lau et al., (2008). *Application of high-resolution NWP Model Output in Thunderstorm Forecasts*. Report. Department of Physics, The Chinese University of Hong Kong, link: http://www.phy.cuhk.edu.hk/hko/08/stephen.pdf