USING OBSERVED HAIL TO QUALITY EXTREME HAIL CLIMATE IN VIET NAM. PART I: DATA QUALITY CONTROL AND CLIMATOLOGICAL CHARACTERISTICS OF HAIL

Dam Viet Bac, Nguyen Dang Mau(1), Bui Minh Tuan(2), Tran Thanh Thuy(1), Pham Thi Hai Yen(1), Tran Thi Tam(1), Do Van Man(1), Luong Xuan Thai(4), Nguyen Tung Anh(1), Tran Chan Nam(3), Hosam Ali(5), Muehlbauer Andreas(5), Osibanjo Olabosipo(5)

(1)Viet Nam Institute of Meteorology, Hydrology and Climate change, Hanoi, Viet Nam
(2)Hanoi University of Science, Viet Nam National University, Hanoi, Viet Nam
(3)Hanoi University of Natural Resources and Environment, Hanoi, Viet Nam
(4)Viet Nam Meteorological and Hydrological Administration, Hanoi, Viet Nam
(5)Center for Property Risk Solutions, Research Division, FM Global, USA

Received: 25 July 2022; Accepted: 18 August 2022

Abstract: *In this study, we focus on studying data and calculating climatic features of hail. In this study, a hail dataset was created by combining historic records from 186 meteorological stations for the period 1961 - 2021. The data are obtained from several sources including observations from the Viet Nam Meteorological and Hydrological Administration (VNMHA), annual reports of the Viet Nam Institute of Meteorology, Hydrology and Climate change (IMHEN) and the Viet Nam Disaster Management Authority (VDMA). The data used in this study is defined from the official data sources. Therefore, the data quality control shows that this data may be considered good available data of hail. Research results for 1961 - 2021 show that the highest number of hail events was observed in the Northwest (743 events), and in the Northeast (557 events) subregions; the lowest number of hail events was observed in the South Central (16 events) and the South (27 events) subregions. The higher frequency duration of hail occurs from March to May, with the peak in April. The largest hail diameter, recorded in Vietnam is 12 cm at Bac Ha station on March 27, 2013; the common largest hail diameter size in history at stations is around 3 - 5 cm at the Northern stations and 1 - 3 cm at the southern stations.*

Keywords: *Climatology, hail, hail size.*

1. Introduction

According to the Law on Natural Disaster Prevention, hailstorms are one of Viet Nam's most impactful natural hazards. Hailstorms usually occur during the transition seasons from warmer to colder (September - October) or from colder to warmer season (Mar - May) [8]. During these transition seasons, sudden deep convection would be formed due to the interaction of two air masses above the warm ground which maycause hail events. Especially in the mountainous areas of North Viet Nam.

Corresponding author: Nguyen Dang Mau
E-mail: mau.imhen@gmail.com

Large hail can cause significant damage to property and agriculture. However, little is known about the hazard posed by large hail. Large hail is the greatest contributor to insured losses from thunderstorms [3]. Hail accounts for the vast majority of severe weather-induced property loss in Viet Nam. However, our understanding of the geographic distribution and spatiotemporal variability of hail occurrence is limited due to numerous documented biases associated with hail reports.

There is a variety of factors that can introduce errors to hail reports, including (but not limited to) population density near a storm, concurrent weather hazards, and the use of reference objects to estimate the size [7, 1, 2, 11]. The largest and most quality-controlled annual hail report database produced by the Viet Nam Meteorological and Hydrological Administration (VNMHA) and Viet Nam Institute of Meteorology, Hydrology and Climate change (IMHEN) is also limited by the exclusion of small hail and an overall underrepresentation of severe hail frequency. These limitations prevent the testing of objective hail identification methods for the full range of possible hail sizes, which has motivated mutiple efforts to fill this data gap [9].

Despite these limitations, hail reports provide one of the most confident ground truths of hailfall. Thus, numerous studies have evaluated hail reports from different sources and time periods. Overall, there are no studies focusing on hail and its extreme size in Viet Nam.

Due to the limitations of above mentioned direct and indirect methods of hail identification, environmental parameters from model output have been leveraged as a proxy for hail occurrence in past and future climates. In this study, we focus on collecting all available data and defining several climatology characteristics of hail and its extreme size reported in Viet Nam.

2. Data and methodology

2.1. Data

a. Observation data

Observational data is obtained from 186 meteorological stations in Viet Nam (Figure 1) for the period 1961 - 2021.

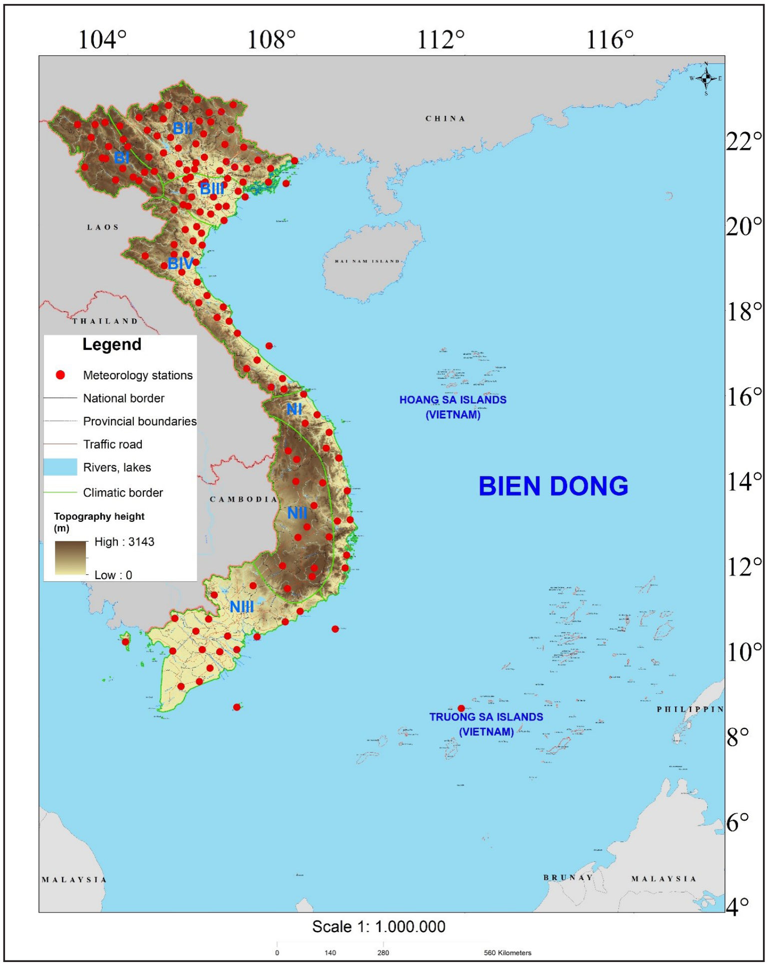


Figure 1. Distribution of meteorological stations

The meteorological stations are situated in seven climatic subregions including:

In the northern parts: (i) 22 meteorological stations in the Northwest subregion (BI); (ii) 52 meteorological stations in the Northeast subregion (BII); (iii) 14 meteorological stations in the Red River Delta subregion (BIII); (iv) 31 meteorological station in the North-Central Coast subregion (BIV);

In the Southern parts: (i) 20 meteorological stations in the South Central Coast subregion (NI); (ii) 18 meteorological stations in the Central Highlands subregion (NII); (iii) 30 meteorological stations in the South subregion (NIII).

b. Hail reports

The annual reports in Viet Nam were collected from websites of VNMHA (available at http://vnmha.gov.vn/tap-chi-kttv), IMHEN (available at http://imh.ac.vn), and VDMA (available at http://ccdpc.gov.vn/).

2.2. Methodology

a) Data combination:

Due to limitations in the number of data, the time series of hail data for calculations at the stations is established by the following approach: (i) The original data is the data at stations observed by VNMHA; (ii) The additional data for adding information into the original data of VNMHA is from the reports of IMHEN and VDMA.

b) Check outliers:

An outlier is a value or an observation that is distant from other observations, that is to say, a data point that differs significantly from other data points. Outliers are considered values that deviate so much from other observations that one might suppose a different underlying sampling mechanism. Sometimes, outliers can be extreme values, in some cases, they are mistakes. Although there is no strict or unique rule whether outliers should be removed or not from the dataset before doing statistical analyses, it is quite common to, at least, remove outliers that are due to an experimental or measurement error. Some statistical tests require the absence of outliers to draw sound conclusions, but removing outliers is not recommended in all cases and must be done with caution.

Boxplots are also useful to detect potential outliers. A boxplot helps to visualize a quantitative variable by displaying five common statistical parameters of a dataset (minimum, maximum, median, first and third quartiles) and any observation that was classified as a suspected outlier using the interquartile range (IQR) criterion.

The median is the middle observation in a ranked dataset (or mean of the two middle observations for an even numbered dataset) and is a measure of the central tendency of the data. The median is equivalent to the 50th percentile in a percentile rank analysis with the same number of observations below as above the median. An advantage of the median is its resistance against outlying values for $n \geq 3$, where n is the number of observations. Whereas the mean can be skewed by an extreme outlying observation, especially for relatively small datasets, the median is less affected and therefore remains more robust.

The interquartile range: The box represents the middle 50% of the ranked data and is drawn from the lower quartile value to the upper quartile value (i.e., the 25th to 75th percentile). The lower (upper) quartile is computed by taking the median of the lower (upper) half of the ranked data. The difference between the upper and lower quartile values is referred to as the interquartile range (IQR), and the height of the box is proportional to the statistical disparity or spread of the inner 50% of the ranked data.

The IQR criterion means that all observations above:

$$Q_3 + 1.5 \times IQR$$

or below:

$$Q_1 - 1.5 \times IQR$$

(where Q_1 and Q_3 correspond to the first and third quartile respectively, and *IQR* is the difference between the third and first quartile) are considered potential outliers. In other words, all observations outside of the following interval will be considered potential outliers:

$$[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$$

Where *IQR* is calculated as the difference between Q_3 and Q_1. The outliers then will be checked by the paper charts.

c) Defining Climatological characteristics of hail:

Because hail is a very rare phenomenon, the data series at meteorological stations is not long enough, and several hail also occurred in a day, meanwhile the climatological characteristics of hail are not defined by statistical daily calculations. Therefore, to clarify the Climatological characteristics of hail, we focus on calculating statistical climatic characteristics for 1961 - 2021 as follows:

- A total number of hail phenomena (event): (i) Calculate the total number of hail that occurred at the stations; (ii) Calculate the total number of hail that occurred in the climatic subregions from the total number of hail at the stations.

- Ratio of monthly hail (%) in contribution to the total annual hail: Percentage of monthly hail events accounting for a total number of hailstorms in a year.

- Average size of hail diameter: Average size of observed hail diameter over the observed history.

- The largest hail diameter in the historical observation at stations: The absolute largest diameter in the history of observation at the station.

3. Results

3.1. Data quality control

Calculation results in Table 1 show that the outliers of hail diameter are determined at monitoring stations in 5/7 climatic subregions (Northwest, Northeast, North Central, South Central and Central Highlands). In which, outliers were identified at 10/22 stations in the Northwest, 17/52 stations in the Northeast, 3/31 stations in the North Central Coast, 1/20 stations in the South Central Coast, and 2/18 stations in the Central Highlands. The number of outliers is small because the sample size observed is very small (hail rarely occurs). After these outliers were identified, these values were compared and normalized according to the original data of the stations (stored by paper documents).

In the Northwest, the median of hail diameter ranges from 1.1 cm (Sin Ho station) to 1.35 cm (Muong Te station). Outlier of hail diameter ranges from 2 cm to 3.5 cm. In which, most stations have only one outlier; Only at Muong Te station and Son station, there are two outliers (Table 1).

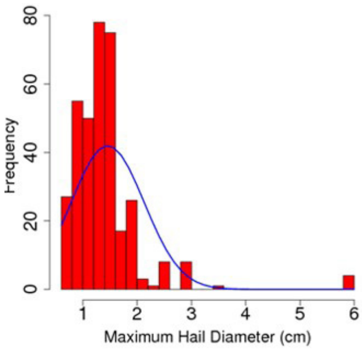
In the Northeast, the median of hail diameter ranges from 1.0 cm to 1.5 cm. However, the value of outlier fluctuates in a very large range, from 2.0 to 12.0 cm (largest at Bac Ha station). At most stations, the number of outliers is commonly defined as one outlier; some stations with 2 outliers observed at Bac Ha, Lao Cai, Tuyen Quang, Cao Bang; three outliers observed at Mu Cang Chai station. In particular, the outlier value is very large at Bac Ha station with the value reaches to 12 cm and 7 cm; and at Chiem Hoa station is 10 cm (Table 1).

In the North Central, the number of stations has outliers is 3 stations, and only one outlier observed. In which, median value ranges from 1.1 cm to 1.15 cm; outlier values ranged from 3 to 5 cm (Table 1). In the South Central, outlier was detected only at Tra My station, with a value of 5 cm. In the Central Highlands, outlier was determined at Kon Tum and Da Lat stations, with both values being 3 cm (Table 1).

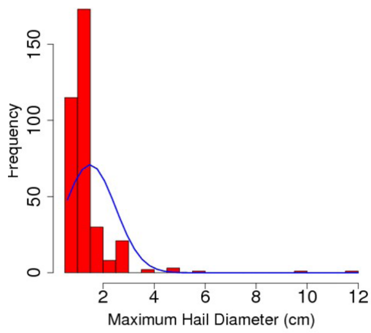
Table 1. The statistical values for checking hail data

No.	Station name	Statistical values of hail's diameter (cm)				Climatic subregion
		Lower row (Minimum Score) Q_1-1.5*(Q_3-Q_1)	Median	Upper row (Maximum Score) Q_2+1.5*(Q_3-Q_1)	Outlier	
1.	Muong Te	0.5	1.35	2.5	3; 3.5	Northwest
2.	Sin Ho	0.5	1.1	2.2	3	
3.	Muong Lay	0.5	1.2	2	5	
4.	Bac Yen	0.8	1.4	2	3	
5.	Co Noi	1	1.4	1.8	2	
6.	Phu Yen	0.7	1.3	2	2.5	
7.	Moc Chau	0.8	1.4	2	2.5	
8.	Son La	0.7	1.2	1.5	2; 2.5	
9.	Mai Chau	0.8	1.4	1.5	3	
10.	Kim Boi	1	1.3	1.5	2	
11.	Bac Ha	0.8	1.5	3	7; 12	Northeast
12.	Mu Cang Chai	0.6	1.2	1.5	2; 2.5; 3	
13.	Lao Cai	1.2	1.5	3	4; 5	
14.	Sa Pa	0.6	1.1	2.2	3	
15.	Ha Giang	0.6	1.05	2	3	
16.	Chiem Hoa	0.8	1.4	5	10	
17.	Tuyen Quang	0.8	1.2	1.5	2; 2.5	
18.	Bac Can	0.8	1.2	3	6	
19.	Cho Ra	0.8	1.15	1.7	3	
20.	Minh Dai	0.8	1.05	1.2	2	
21.	Phu Ho	0.8	1.1	1.5	2.5	
22.	Viet Tri	1	1.35	1.62	4	
23.	Dinh Hoa	0.8	1.3	1.7	3	
24.	Bao Lac	0.8	1.25	1.5	3	
25.	Cao Bang	0.7	1.2	1.5	2.5; 3; 5	
26.	Huu Lung	0.8	1.25	1.4	3	
27.	Bac Giang	1	1.15	1.72	3	
28.	Bai Thuong	0.8	1.1	1.4	3	North Central Coast
29.	Hoi Xuan	1	1.15	2	4	
30.	Quy Chau	0.7	1.2	1.5	5	South Central Coast
31.	Tra My	1	1.35	2.3	5	
32.	Kon Tum	0.8	1.3	2	3	Central Highlands
33.	Da Lat	0.8	1.3	2	3	

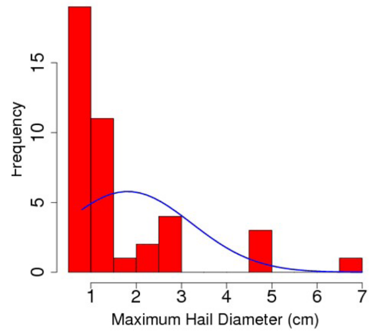
A histogram is a visual representation of the distribution of a maximum hail diameter dataset. As such, the shape of a histogram shows where a relatively large amount of the data is situated and where there is very little data to be found. In other words, a histogram presents where the center of the distribution is located, how the data scatter around this center, and where possible outliers are found. In short, the histogram consists of an x-axis, a y-axis, and various bars of different heights. The y-axis shows how frequently the values on the x-axis occur in the data, while the bars group ranges of values or continuous categories on the x-axis. The first step is to divide the entire range of values into a series of intervals. The second step is to count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent and are often (but not required to be) of equal size (Figure 2).



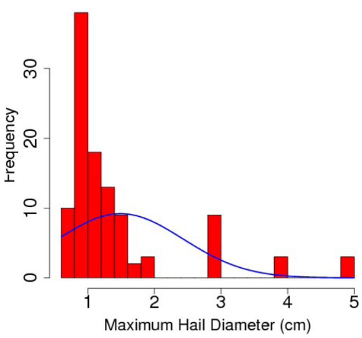
(a) BI



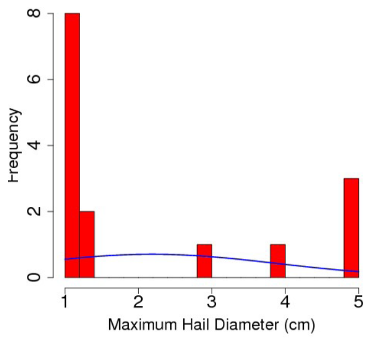
(b) BII



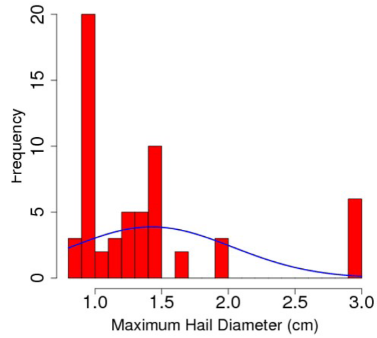
(c) BIII



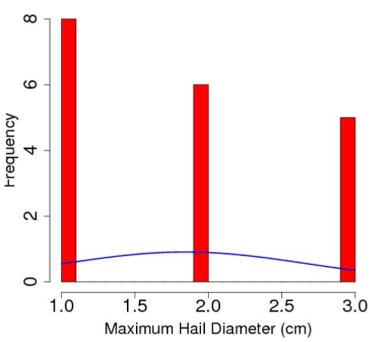
(d) BIV



(e) NI



(f) NII



(g) NIII

Figure 2. Histogram of maximum hail diameter for the seven climatic regions of Viet Nam

The general hail size of each climatic region is shown in Figure 3 for individual stations. It can be seen from the results that there are significant differences between hail diameter distribution patterns in the seven regions in Viet Nam. In general, the combined maximum hail diameter exhibits skewness in Northwest, Northeast, Red River Delta, North Central, and South Central while the patterns are not clear in the Central Highlands and Southern Plain. The hail diameter distribution is close to the log-normal distribution in the Northeast and North Central. In contrast, hail diameter is distributed in a wide range in the Central Highlands and Southern Plain. The combined data in the Northwest, Northeast, Red River Delta, North Central, and Central Highlands represent the skewed left pattern, reflecting that the mean of the maximum hail diameter in these regions is smaller than the median (Figure 2).

3.2. Climatological characteristics of hail in Viet Nam

3.2.1. Climatology frequency of hail

A number of hail:

According to the 1961 - 2021 statistical data in Table 1, hail has been observed in all seven climatic subregions of Viet Nam. However, the number of times observed is significantly different; higher in northern climates and much lower in southern climatic subregions. During 1961 - 2021, there are 743 times of hail in the Northwest, and 557 times of hail in the Northeast. However, there are only 40 times in the Red River Delta, 16 times in the South Central, and 27 times in the South observed (Table 1). These results show that hail mostly occurs in the northern mountainous areas; and is much rare in the deltas of Viet Nam. The causes of the higher frequency of hail in the northern mountainous areas may be driven by the interaction of two specific air masses above the complicated terrain [1]. Calculation results in Table 2 also show that the major hail season in the northern areas is in March-April, and in April-May in the southern areas. In the summer and winter months, hail is very rare or not occurring.

In the Northwest (BI), the total number of observed hail events is 743 times, concentrated from January to May. The highest in April with the number of hail occurrences is 361 times (corresponding to a frequency of 5.9 times/month). In addition, hail was also observed in October-December and January, with a very much lower number from 2 times in October to 14 times in January; not observed in August-September (Table 2, Table 3).

In the Northeast (BII), the total number of observed hail event is 557 times, concentrated from January to May. The major season of hail is from March to April, with the highest number of hail events being 220 times in April. In the other months, hail rarely occurs (Table 2, Table 3).

In the Red River Delta (BIII), hail rarely occurs with an observed frequency of about 40 times in the period 1961 - 2021, much lower than in the Northwest and Northeast. Hail in this area mainly occurs in April, with a total number of events being 12 times. During the summer, autumn, and winter months, hail is almost absent in this area (Table 2, Table 3).

In the North Central (BIII), the number of hail events in the period 1961 - 2021 is 122 times, higher than in the Red River Delta. In this area, hail mainly occurs in March-May. In the other months, hail almost is not observed (Table 1, Table 2).

In the South Central (BIV), the total number of hail events in this area is the lowest in Viet Nam, with 16 hail times observed in the period 1961 - 2021. In general, hail occurs very rarely, and the biggest number of occurrences is 6 times in May (Table 2, Table 3).

In the Central Highlands, the total number of hail events in this area is 113 times, with the major hail season in March-May, and the highest in April with 52 observed hail times. In the other months, hail occurs very rarely in this area (Table 2, Table 3).

In the Southern, hail is very rare in this area. In the period 1961 - 2021, the total number of observed hail events is 27 times. The period of hail occurrence mainly concentrates in April-May and July (Table 2, Table 3).

Table 2. The 1961 - 2021 total number of hail (event) combined from stations in seven climatic subregions in Viet Nam

Climatic subregions	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
Northwest (BI)	14	103	179	361	66	4	1	0	0	2	7	6	743
Northeast (BII)	34	75	139	220	40	6	1	3	5	1	25	8	557
Red River Delta (BIII)	5	4	5	12	6	2	1	0	1	0	4	0	40
North Central Coast (BIV)		8	18	64	22	5	1	0	0	0	0	0	122
South Central Coast (NI)	0	0	1	5	6	1	1	0	1	0	0	1	16
Central Highlands (NII)	0	6	18	52	16	5	9	4	2	1	0	0	113
South (NIII)	0	0	0	8	6	1	8	2	1	1	0	0	27

Table 3. The 1961 - 2021 percent contribution (%) to the total number of hail events of the mean total number of hail events observed in seven climatic subregions in Viet Nam

Climatic subregions	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Northwest (BI)	3.0	11.7	23.2	48.3	11.4	0.5	0.2	0.0	0.0	0.2	0.6	0.9
Northeast (BII)	6.7	18.4	21.1	38.4	5.8	1.1	1.2	0.4	0.4	0.6	5.6	0.5
Red River Delta (BIII)	19.0	8.8	10.6	29.7	11.2	5.8	0.9	0.0	1.7	0.0	12.3	0.0
North Central Coast (BIV)	4.7	3.9	22.0	46.2	20.1	2.7	0.4	0.0	0.0	0.0	0.0	0.0
South Central Coast (NI)	0.0	0.0	4.2	18.1	50.0	2.8	2.8	0.0	5.6	0.0	0.0	16.7
Central Highlands (NII)	0.0	2.3	8.1	46.2	19.4	1.4	20.9	1.2	0.4	0.2	0.0	0.0
South (NIII)	0.0	0.0	0.0	45.5	29.5	0.9	7.1	15.2	0.9	0.9	0.0	0.0

Hail diameter:

Figure 3 presents the results of calculating the average size of hail diameter at the stations. The results show that the commonly observed mean diameter is less than 1.0 cm. At several observation stations in the Northwest mountains, the size of hail diameter is significantly larger, such as about 2.28 cm in Sapa (Lao Cai province), 1.51 cm in Sin Ho (Lai Chau province), 1.37 cm in Lao Cai (Lao Cai province) and 1.03 cm in Muong Lay (Lao Cai province).

Figure 4 shows that the largest hail diameter in the period 1961 - 2021 is 15 cm, the common value is in the range of 1.0 - 5.0 cm. In the Northern areas, the largest hail diameter is about 3 - 5 cm; In the South, the common largest diameter size is less than 3 cm. The largest hail diameter size ever observed in Viet Nam is 12.0 cm in Bac Ha (Lao Cai) on March 27, 2013. In addition, several stations with a significant hail diameter ever observed as 5 cm at Muong La (Dien Bien province), Tam Duong (Lao Cai province), Cao Bang (Cao Bang province), Son Tay (Hanoi), Nhu Xuan (Thanh Hoa province); 6 cm at Bac Can station (Bac Can province), Quy Chau and Tuong Duong (Nghe An province), Tam Ky and Tra My (Quang Nam province); 7 cm at Ha Dong station (Hanoi); 10 cm at Chiem Hoa station (Tuyen Quang province).

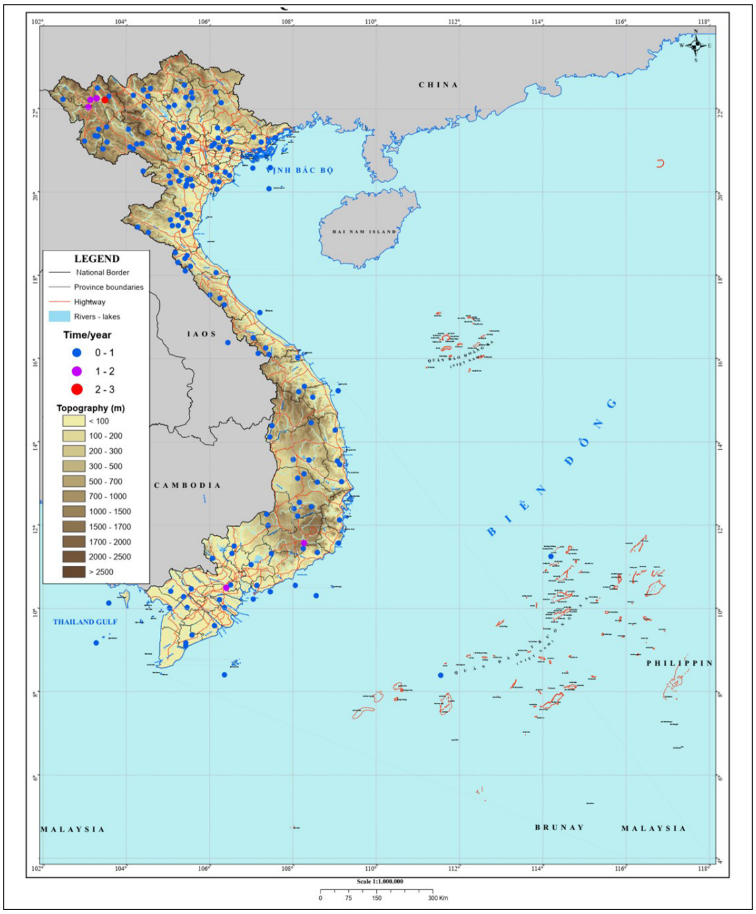


Figure 4. The 1961 - 2021 historical maximum hail diameter (cm)

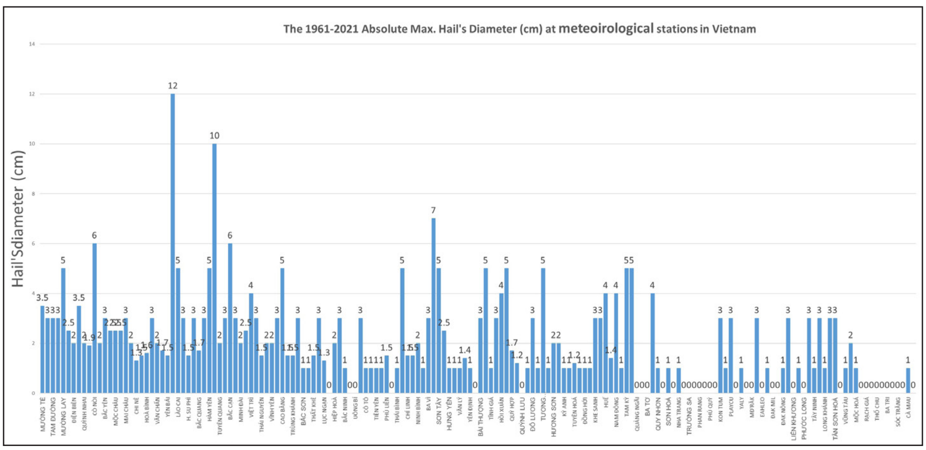


Figure 3. The distribution of the 1961 - 2021 annual mean number of the hail event

4. Conclusion and discussion

This study has investigated the characteristics of hail in Viet Nam for the period 1961 - 2021 based on observed data at meteorological stations (collected from VNMHA), annual reports of IMHEN, and VDMA. The 1961 - 2021 time series data of hail events at stations are established by using the original data of VNMHA combined with data from IMHEN and VDMA. The data of the study is expected as the good available hail data in Viet Nam developed by combining three official data. The data quality control results show that this data can be used to calculate the hail's characteristics in Viet Nam.

According to the data for the period 1961 - 2021, hail event occurs mostly in the Northwest and Northeast; rarely occurs in the Northern Delta, the South Central, and the South. In Viet Nam, the major hail season is defined from March to May, with the higher frequency of the season in April. In other months, hail may occur, however, the frequency is much lower than in the hail season. The common mean hail diameter size is about 1 cm across stations. The common largest hail diameter size is about 3 - 5 cm in the northern areas and about 1 - 3 cm in the southern areas. The largest hail diameter size ever observed in Viet Nam is 12 cm at Bac Ha station on March 27, 2013.

Hail is a relatively rare hydrometeorological disaster in Viet Nam. In recent years, many large-scale and large-size hail events have occurred, causing serious damage to people and property, especially in the northern mountains. Moreover, studies on hail, especially hail forecasts and warnings need to be be conducted in the future.

Acknowledgment: *This study was supported by the project "Calculating return periods of hail size in Viet Nam" funded by FM Global and the project "Characteristics of Trade Wind and its Influences on Weather and Climate of Viet Nam", Code: TNMT.2021.562.05.*

References

1. Allen JT, and Tippett MK, (2015), "*The characteristics of United States hail reports: 1955 - 2014. Electron", J. Severe Storms Meteor*, 10 (3), https://ejssm.org/ojs/index.php/ejssm/article/viewArticle/149
2. Allen JT et al. (2017), *"An extreme value model for U.S. hail size". Mon. Wea. Rev*, 145, 4501-4519, 10.1175/MWR-D-17-0119.1
3. Gunturi, P., and M. K. Tippett, (2017), *Managing severe thunderstorm risk: Impact of ENSO on U.S. tornado and hail frequencies*. WillisRe Tech. Rep., 5 pp.
4. https://programmathically.com/chi-square-distribution-table.
5. IMHEN, 2011-2021: *Annual climate in Viet nam (from 2011 report to 2021 report).*
6. John T. Allen et al. (2017), *"An Extreme Value Model for U.S. Hail Size", AMS*. DOI: 10.1175/MWR-D-17-0119.1
7. Kelly et al. (1985), *"Climatology of nontornadic severe thunderstorm events in the United States", Mon. Wea.* Rev, 113, 1997-2014, 10.1175/1520-493(1985)113<1997:CONSTE>2.0.CO;2
8. Nguyen D.N and Nguyen T.H, (2004), *Climate and climatic resources in Viet Nam*. Agricultural publisher.
9. *"NIST/SEMATECH e-Handbook of Statistical Methods"*. www.itl.nist.gov. doi:10.18434/m32189. Retrieved 2021-12-14.
10. Ortega, L. et al. (2009), *"The Severe Hazards Analysis and Verification Experiment". Bull. Amer. Meteor.* Soc., 90, 1519–1530, doi:10.1175/2009BAMS2815.1.
11. Witt A et al. (2018), *"Rapid-scan radar observations of an Oklahoma tornadic hailstorm producing giant hail", Wea. Forecasting*, 33, 1263–1282, 10.1175/WAF-D-18-0003.1.
12. Xiang Ni et al. (2020), *"A Climatology and Extreme Value Analysis of Large Hail in China". AMS*. DOI: 10.1175/MWR-D-19-0276.1.