VIETNAM JOURNAL OF HYDROMETEOROLOGY

Research Article

Application of self-organizing maps and K–Means methods to classify summer heat wave weather patterns in Viet Nam

Mai Tran Thi Tuyet[1], Hoa Van Vo[1*], Tuan Le Danh[2]

[1] Regional Hydro–Meteorological Center of Red delta river basin; tuyetmai1295@gmail.com; vovanhoa80@yahoo.com
[2] Control Automation Production Institute of Technology (CAPIT); ledanhtuan@gmail.com

*Corresponding author: vovanhoa80@yahoo.com; Tel.: +84–912509932

Received: 23 March 2022; Accepted: 20 May 2022; Published: 25 June 2022

Abstract: The research applies self-organizing maps (SOM) technique in combination with K–Means method to objectively classify weather patterns that cause summer heat wave in Viet Nam based on the dataset from 1998 to 2018. The pressure of mean sea level (PMSL) and geopotential height at 500hpa (H500) of JRA25 reanalysis data are used. The heat wave is defined to occur if the daily maximum temperature of at least 2/3 of surface synoptic stations in research area was greater than 35°C. According to above mentioned criteria, 156 summer heat waves were subjectively found at Northern region in period of 1998–2018. In central and southern regions, the summer heat waves were respectively found 204 and 69. By applying SOM and K–Means, there were 4, 3 and 2 key weather patterns that caused summer heat waves in Northern, Central and Sothern region respectively. In fact, the weather pattern caused summer heat waves at research region is usually related to activities of the western hot depression pattern and Northwest Pacific Subtropical High Pressure. The combination of 2 weather patterns or more was usually found Northern and Central region. However, the number of heat wave detected by SOM is smaller than number of heat wave was subjectively determined by forecaster (there are 109, 171 and 62 heat waves detected by SOM for the Northern, Central and Sothern region respectively). The reason for this result is that SOM method has not been able to identify heat waves caused by the combination of many weather patterns or by small and meso–scale weather patterns.

Keywords: SOM; Classification; Summer heat waves; Heat wave weather patterns.

1. Introduction

In recent years, under the impact of climate change, heat wave has been occurring in Vietnam with an increasing trend in both frequency and intensity. Damage statistics show that heat waves are also a type of natural disaster that causes a lot of damage to people and property. Therefore, it is very necessary to increase the understanding of the mechanism that causes heat wave to improve forecast quality. In Vietnam, there have been many studies on summer heat waves in which refer to many aspects from the statistics of heat wave frequency based on past data sets [1–2], causes and patterns of heat wave [3–5], forecasting heat wave is based on statistical methods [6] and NWP models from short term to seasonal scale [7], predicting the change of heat in the future according to climate change scenarios, etc.

In the classification of weather patterns, the SOM method is widely applied, especially related to heavy rain problems [8–11]. The SOM method has been applied in the

classification of heat wave weather patterns, specifically, [12] used ERA–Interim reanalysis data of the ECMWF from 1979 to 2016 to classify patterns for heat waves where the daily maximum temperature on that day is greater than the 90th percentile of the data series. The results show that the synoptic patterns caused heat wave are classified into 6 clusters based on PMSL anomalies in East Asia. Recently, SOM method has been applied in the classification of abnormal heat wave weather patterns in winter for Northern part of Viet Nam [7].

According to studying the causes and weather patterns that cause heat wave, most studies have partially shown the causes and statistics of typical weather patterns. However, these studies are mainly based on synoptic analysis methods and are analyzed subjectively by forecasters, so the obtained results are still subjective and difficult to apply in operational prediction. To contribute additionally to the results of classification of weather patterns that cause heat wave in Vietnam in an objective way, the research applies the SOM (Self–Organizing Map) method in combination with JMA's JRA25 reanalysis data set to identify groups of weather patterns that cause large–scale heat wave events in some areas of Vietnam. The research mainly focuses on summer heat wave in Viet Nam that occurred in large scale and weather patterns that significantly caused these summer heat waves. The daily maximum temperature data at surface synoptic stations, pressure of mean sea level and geopotential height at 500hpa of JRA25 reanalysis data will be used. The next of paper will give out the dataset and methodology in SOM application. The results present in third part of paper. It finally is some conclusions and remarks.

2. Materials and Methods

2.1. Dataset

To be able to find out statistics of heat waves occurring over Vietnam in the 21 years from 1998 to 2018, we collected maximum temperature data (Tx) at 183 surface synoptic stations. As is known, the heat wave cannot be directly observed, but it is determined from the observed quantities based on the given criteria. In the operational forecast, according to the intensity of heat wave, it can be divided into 3 types including heat wave ($35^oC \leq Tx < 37^oC$), strong heat wave ($37^oC \leq Tx < 39^oC$) and extreme heat wave ($Tx \geq 39^oC$). According to influenced area, heat wave can be divided into large–scale heat wave and local heat wave. To simplify the determination and ensure that there is enough sample size for SOM method, in this study we use the criterion $Tx \geq 35^oC$ and have at least 1/2 of the surface synoptic stations in the study area satisfy the condition $Tx \geq 35^oC$ at the same time. In addition, the heat wave in this study is mainly considered under the concept of a "*wave*", which satisfies the condition that at least 1 day occurs or there are 2 or more days as sequency. In case if the satisfactory days are interleaved with the unsatisfactory days, it is also defined as a heat wave.

Heat wave occurs every year in Vietnam and there are differences between climate in terms of origin, causes, intensity, scope, etc. The research aims at a large–scale heat wave as mentioned above and only aims to identify the dominant weather patterns, so there will be many heat waves occurring in many regions at the same time. Therefore, to ensure that the classification by SOM is objective, not duplicated and clearly shows the dominant weather patterns, we divide the study area across the country into 3 regions including the North region, the Central Region and the South region. Basing on the criteria for determining the heat wave, we have subjectively determined 156 heat waves in the northern region for period of 1998–2018. For Central Region and the South region, number of found heat waves is respectively 204 and 69 (Table 1).

To have data on the grid as input for the SOM method, we collected the JRA25 reanalysis data of JMA corresponding to the period of the aforementioned heat wave dataset. NOAA's NNRP2 and ECMWF's ERA–Interim reanalysis data sources were not collected due to

JRA25 has the higher resolution. Because it was developed by JMA, the quality of the JRA25 data source has higher accuracy in Asia area. Since the atmospheric fields are correlated, using all fields in the identification is unnecessary and may give undesirable results. Therefore, the selection of characteristic quantities will limit the amount of work and computation time. Based on the knowledge of weather patterns and previous studies, in this study, we only collect pressure of mean sea level (PMSL) and geopotential height at 500hpa (H500). The PMSL used to characterize the surface hot depression and the H500 will characterize the Northwest Pacific Subtropical High Pressure.

Table 1. Subjective statistics of heat waves for each of research are in Viet Nam by using criteria (Tx ≥ 35ºC and have at least 1/2 of the surface synoptic stations in the study area satisfy the condition Tx ≥ 35ºC at the same time) for period of 1998–2018.

Year	Northern Area	Central Area	Southern Area	Viet Nam
1998	5	6	1	12
1999	4	8	1	13
2000	3	5	0	8
2001	2	7	0	9
2002	1	4	2	7
2003	8	11	4	23
2004	5	9	2	16
2005	4	9	1	14
2006	8	10	2	20
2007	6	7	5	18
2008	8	10	2	20
2009	8	9	2	19
2010	15	14	7	36
2011	6	8	2	16
2012	11	14	4	29
2013	7	12	7	26
2014	11	11	6	28
2015	16	11	5	32
2016	13	12	11	36
2017	9	15	2	26
2018	6	12	3	21
Sum	**156**	**204**	**69**	**429**

2.2. Methodology

The research flowchart is shown on Figure 1. Specifically, basing on known heat waves that are subjectively determined according to the given criteria, PMSL and H500 data of JRA25 will be normalized and selected principal components based on PCA analysis. These components are then fed into the SOM for analysis and the creation of a U–Matrix map. Next, the K–Means method is applied to find the data clusters. Finally, from the given data clusters, weather pattern maps for each cluster are used to find out the main weather patterns that cause the heat wave. Because the meteorological elements in the JRA25 dataset vary in dimension as well as range of variation. Therefore, the predictors must be normalized before training the SOM network. The set of predictors will be normalized to a new set of factors according to the following formula:

$$x' = \frac{x - \mu_X}{\sigma_X} \quad (1)$$

where x' is normalized variable of x predictor, μ_X and σ_X respectively are simple average and standard deviation of x predictor that is calculated based on past dataset. After normalizing, x' variable is non–dimension.

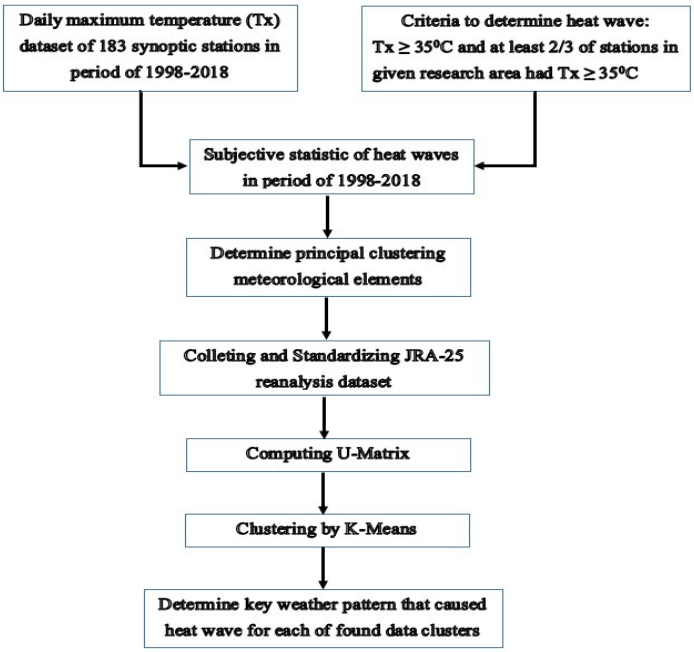

Figure 1. Flowchart of using SOM and K–Means methods to classify weather patterns that cause summer heat wave in Viet Nam based on JRA25 reanalysis dataset.

After normalizing the PMSL and H500 data, in order to reduce data redundancy and computing cost, we will not directly use the PMSL and H500 data on the grid but will analyze them into a series of principal components by applying PCA method. The input data for the PCA method is also normalized according to formula (1) as mentioned above. As a result, instead of including 2 × 40 × 29 variables in clustering (40 and 29 is number of JRA25 grid point in meridional and zonal direction), we only have to cluster the number of 15–20 variables depending on the region.

Basing on principal components are selected, SOM training is implemented according to following step by step:

Step 1: Initializing weight vector $w_j^{(0)}$ with j = 1,2, …, d^* by random select in input dataset (D).

Step 2: Iterating solve

Step 3: Assign x in D with given probability value

Step 4: Finding best fit noron i(x) in Kohonen class basing on Euclidean distance between vector w_j and x: $i(x) = \arg\min_{1 \le j \le d^*} \left\| x - w_j^{(s)} \right\|$

Step 5: Updating weight for all norons in out layer as following formular: $w_j^{(s+1)} = w_j^{(s)} + \eta(s) h_{j,i(x)}(s)(x - w_j^{(s)})$

Step 6: Iterating Step 2 if there is no significant change of SOM feature map (reduce radius of topological neighborhood at specified time)

Step 7: Finishing and give out final SOM feature map

The result of training the SOM network is to create two–dimensional matrix of Kohonen neurons in which each main neuron is a vector whose size is equal to the number of input neurons. Next, to create cluster boundaries in the study, we conduct use the U–Matrix technique combined with the K–Means algorithm. Specifically, basing on the U–Matrix map,

the K–Means method is applied to classify the number of possible heat wave clusters based on the SOM characteristics. After determining the number of possible heat wave clusters, the past heat wave data that subjectively determined as mentioned above will be used to classify each heat wave to which given clusters. After this classification step, the process of determining the dominant weather pattern is performed by displaying atmospheric field maps from the JRA25 data of each heat wave in the given cluster. The all steps to find dominant weather pattern that caused heat wave based on the SOM and K–Means method are shown in Figure 1.

3. Results and Discussion

Figures 2 to 4 show the origin and clustered maps of the U–matrix by applying the K–Means method for the Northern, Central and Southern regions, respectively. Specifically, for the Northern region, the results have 4 clusters found corresponding to 4 groups of weather patterns (Figure 3). However, the degree of clustering (separation between groups) is not really clear. For the Central region, there are 3 data clusters found after applying the K–Means method. For the Southern region, only 2 data clusters were found. Regarding the degree of distinction between clusters, the Southern region shows more clearly than the Northern and Central regions. Figures 5 to 7 respectively show typical weather patterns for each element of the matrix U with size 8 × 8 when applied to PMSL and H500 fields for the North, Central and Southern region, respectively. From these weather pattern maps it is clearer to see the degree of separation in the ground and upper–air weather pattern for each molecule of the U–matrix. Subjectively, the use of the weather pattern matrix as in the Figures 5 to 7 can also analyze and make judgments about the number of groups of weather patterns included in the dataset used to put into SOM analysis.

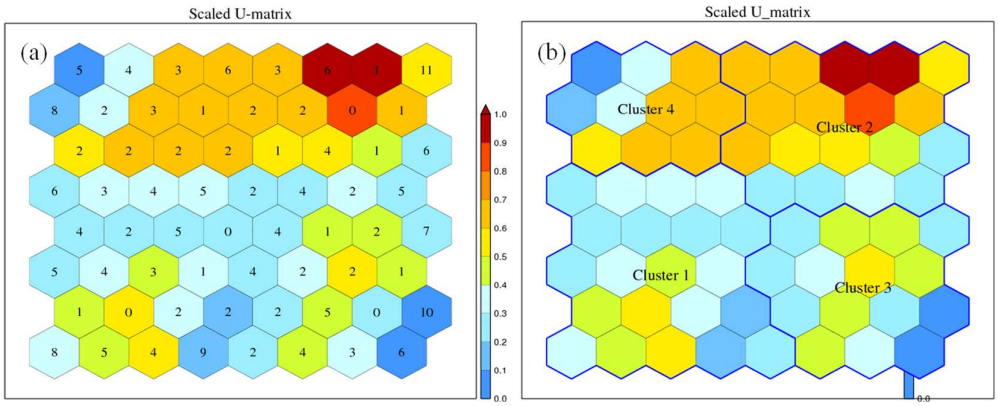

Figure 2. The origin U–Matrix (left) and cluttered U–Matrix by applying K–Means method (right, the cluster boundary line is bold blue) for Northern region basing on JRA25 dataset in period of 1998–2018 (the number means the cases belong the given U–Matrix element, the color palette means the distance of wight vector for each U–Matrix element).

Table 2 gives the results of determining the number of heat waves in each year in the period 1998–2018 for each study region and compares it with the number of heat waves subjectively determined basing on the above criteria. The number of heat waves identified from SOM is less than that determined by forecaster. Specifically, for the Northern region, the number of heat waves detected by the SOM method accounts for about 70% of the subjective determination. Meanwhile in the Central and Southern regions, it is 84 and 90% respectively. Thus, the determination of the number of heat waves by SOM method in the Central and Southern regions is better than in the North. The reason for the difference is the number of weather pattern clusters found by K–Means. Specifically, except for the Southern region (the number of clusters found is equal to the number of weather patterns found

according to the forecaster's analysis), the number of clusters found by K–Mean in the North and Central regions are less than 1 compared to the forecaster's analysis. This result may lead to under–estimate the number of heat waves in these regions. For the Southern region, it is possible that the input data will be in the interference area of the two data clusters.

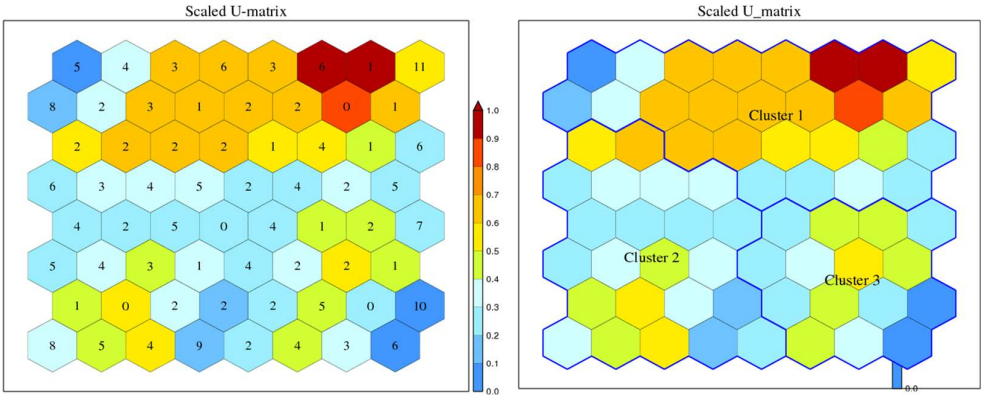

Figure 3. Similar to Figure 2 but for Central region.

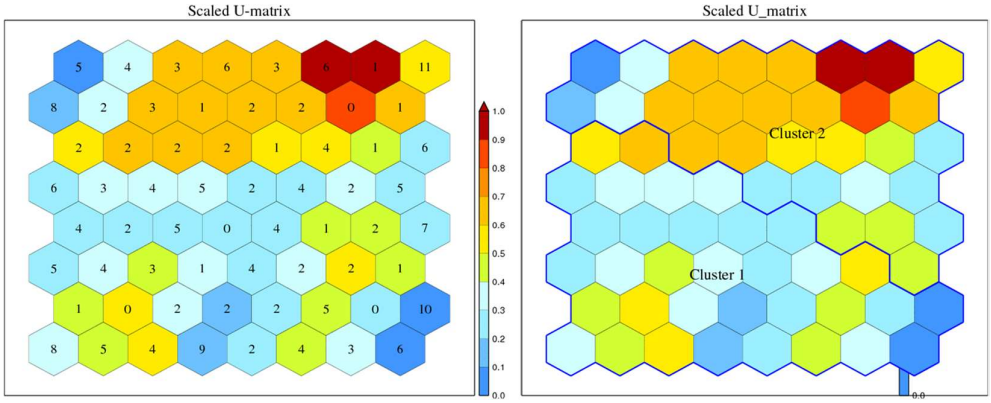

Figure 4. Similar to Figure 2 but for Southern region.

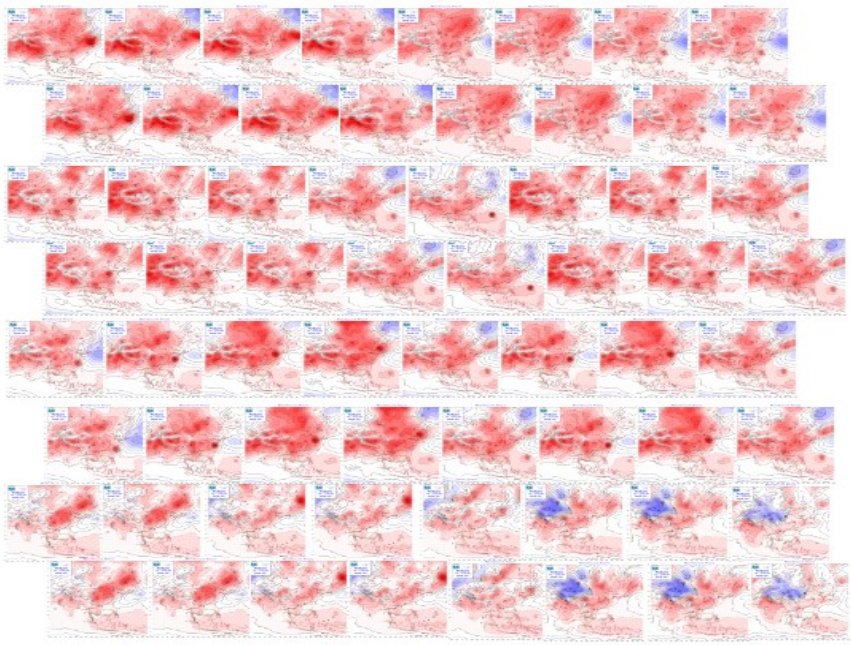

Figure 5a. The significant pressure of mean sea level maps of 8 × 8 U–Matrix for Northern region.

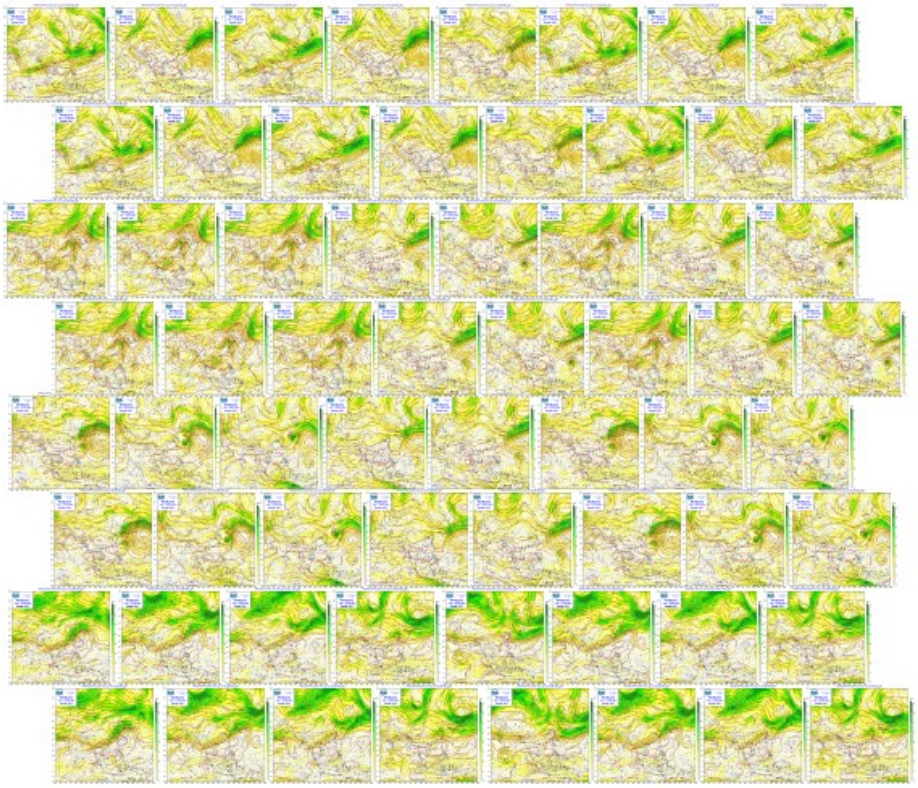

Figure 5b. The significant geopotential height at 500hpa level of 8 × 8 U–Matrix for Northern region.

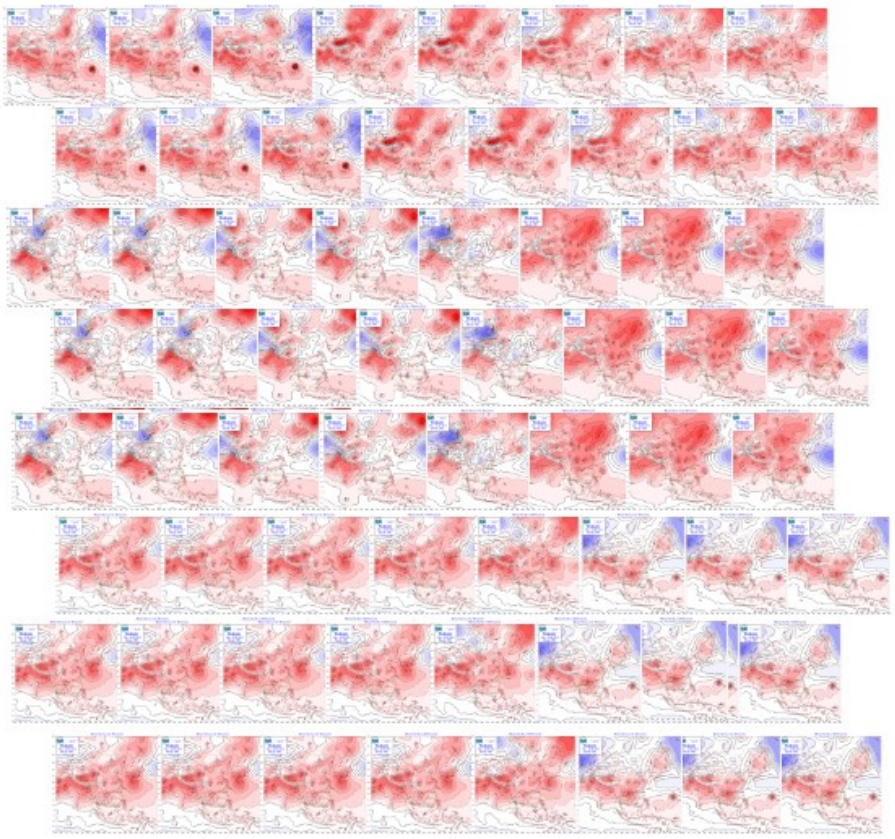

Figure 6a. The significant pressure of mean sea level maps of 8 × 8 U–Matrix for Central region.

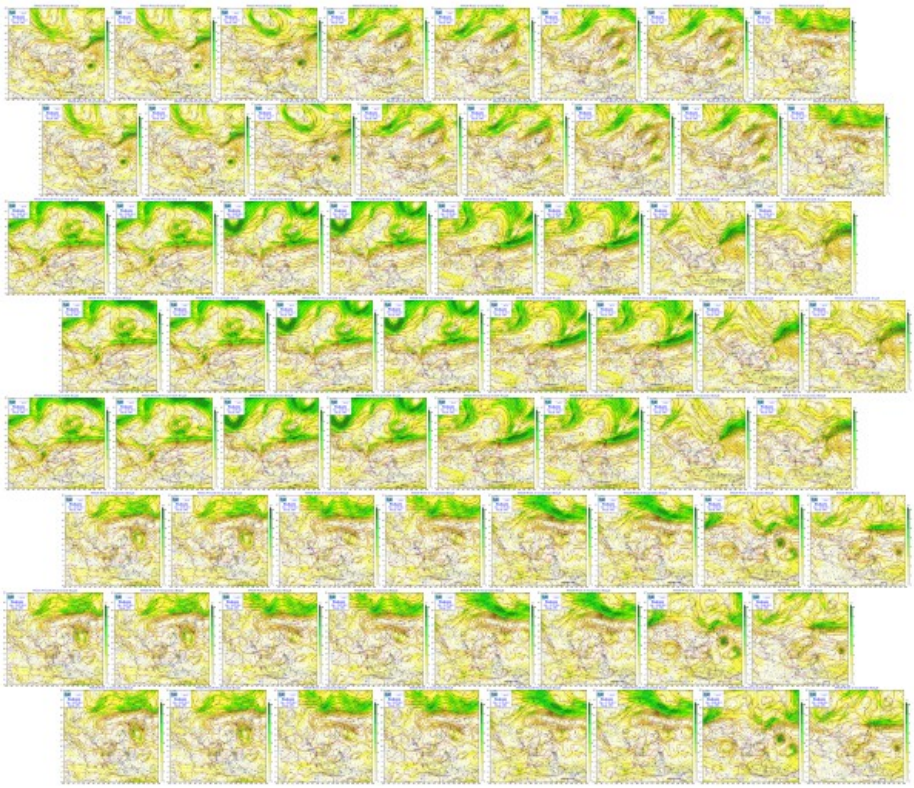

Figure 6b. The significant geopotential height at 500hpa level of 8 × 8 U–Matrix for Central region.

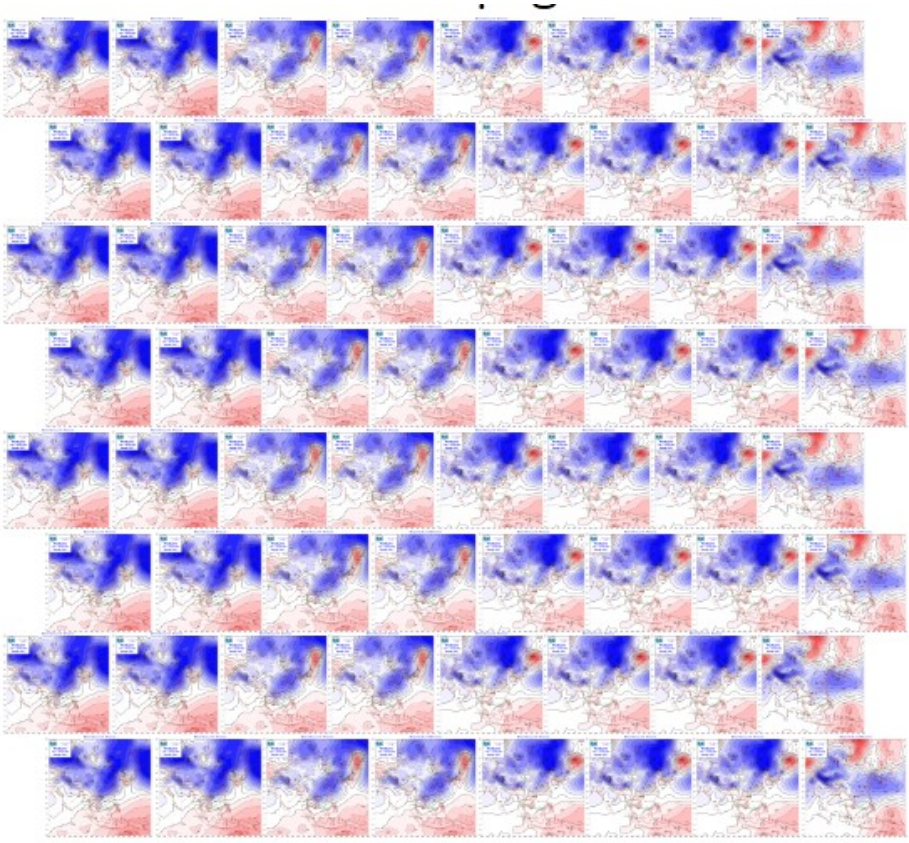

Figure 7a. The significant pressure of mean sea level maps of 8 × 8 U–Matrix for Southern region.

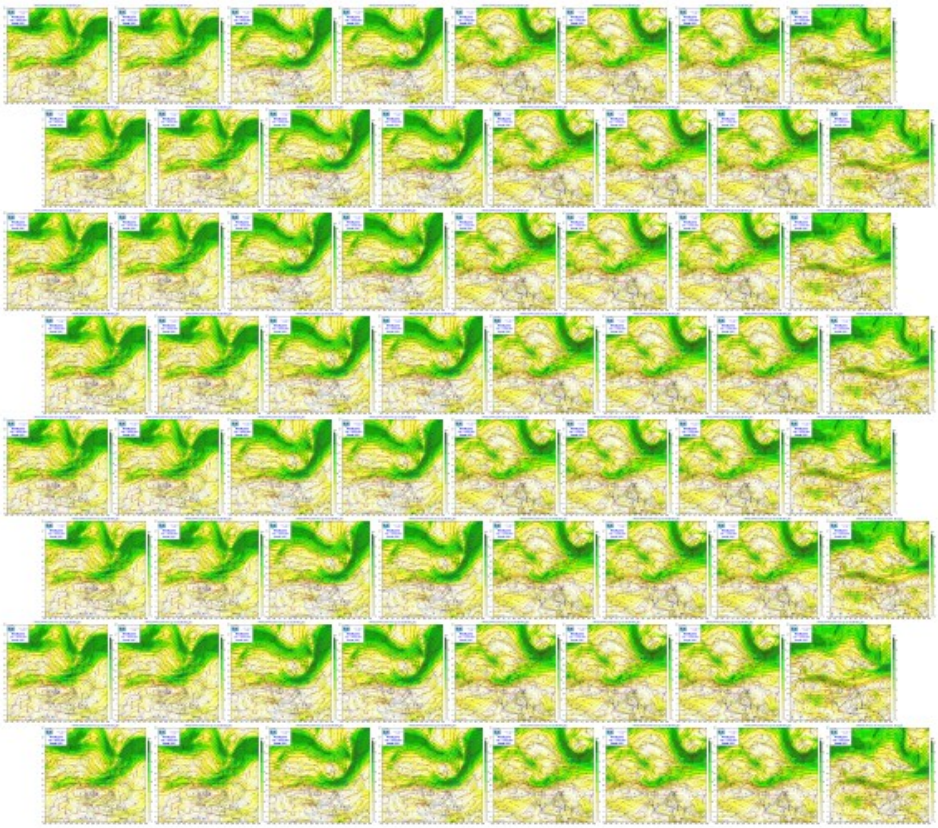

Figure 7b. The significant geopotential height at 500hpa level of 8 × 8 U–Matrix for Southern region.

By reanalyzing significant weather pattern maps of given clusters, we determine dominant weather pattern that caused heat waves in Viet Nam as following:

1. For Northern region: Western hot depression, Northwest Pacific Subtropical High Pressure, South China depression and Northern cold air mass.

2. For Central region: Western hot depression, Northwest Pacific Subtropical High Pressure and South–west monsoon.

3. For southern region: Western hot depression and Northwest Pacific Subtropical High Pressure.

Compared with the clusters of weather patterns determined by the forecaster, the weather patterns found by SOM and K–Means have similarities when only has a large–scale pattern caused heat wave. However, in case of association of 2 or more weather patterns with spatial scale differences, SOM could not be captured all cases, specially has activities of small–scale patterns. This is also the reason why the number of heat waves detected from the SOM is less than that determined by the forecaster's experience.

Table 2. The statistics of heat wave number in period of 1998–2018 for each research region by forecaster and SOM method.

Year	Northern Region		Central Region		Southern Region	
	Forecaster	SOM	Forecaster	SOM	Forecaster	SOM
1998	5	3	6	5	1	1
1999	4	3	8	7	1	1
2000	3	2	5	4	0	0
2001	2	2	7	5	0	0
2002	1	1	4	4	2	1

Year	Northern Region		Central Region		Southern Region	
	Forecaster	SOM	Forecaster	SOM	Forecaster	SOM
2003	8	5	11	9	4	4
2004	5	4	9	7	2	2
2005	4	2	9	8	1	1
2006	8	5	10	8	2	2
2007	6	3	7	5	5	4
2008	8	6	10	8	2	2
2009	8	6	9	7	2	2
2010	15	9	14	12	7	7
2011	6	4	8	6	2	2
2012	11	7	14	12	4	4
2013	7	6	12	12	7	6
2014	11	8	11	10	6	5
2015	16	12	11	10	5	4
2016	13	10	12	10	11	9
2017	9	7	15	12	2	2
2018	6	4	12	10	3	3
Sum	**156**	**109**	**204**	**171**	**69**	**62**

4. Conclusion

The paper has been studied and applied the SOM method combined with K–Means to objectively classify the weather patterns that cause heat wave in Vietnam based on dataset of pressure of mean sea level and 500hpa geopotential height of JRA25 reanalysis in the period 1998–2018. By using the criteria $Tx \geq 35^oC$ and taking a large scale, there respectively were 156 heat waves occurred in the Northern region, 204 in the Central region and 69 in the South in the period 1998–2018. Subjective analysis has shown that there are 5 main weather patterns causing heat wave in the North. For the Central and Southern regions, there are 4 and 2 main weather patterns, respectively. In general, the heat wave occurring in all 3 areas is related to the activity of the western hot depression and the northwest Pacific subtropical high pressure. The heat wave is caused by combination of 2 or more weather patterns mainly occur in the North and Central regions.

The objective classification results based on SOM and K–Means methods are most appropriate in the Southern region, followed by the Central and Northern regions. The number of heat waves detected by the SOM method is less than that determined by subjective methods. The reason is that the number of weather pattern clusters classified from SOM and K–Means is less than in subjective analysis (except for the Southern region). The SOM method classifies well when there is only one dominant weather pattern in large–scale. When there are combinations of 2 or more weather patterns and the impact of these patterns are the same (especially with spatial differences), classification by SOM is difficult because the data will be in the intersection of the clusters. To improve the research results and overcome the above shortcomings, we suggest that it is necessary to optimize the U–matrix size to further enhance the ability to capture the small–scale weather patterns when there is a combination of many weather patterns at different spatial scales.

Author Contribution statement: Conceived and designed the experiments: M.T.T.T., T.L.D.; Analyzed and interpreted the data: H.V.V., T.L.D.; contributed reagents, materials, analysis tools or data: T.L.D.; manuscript editing: H.V.V., M.T.T.T.; Performed the experiments: M.T.T.T.; contributed reagents, materials, analyzed and interpreted the data, wrote the draft manuscript: H.V.V.

Acknowledgments: This work was supported by the Ministry of Natural Resources and Environment through the national Project "The impact of climate change on abnormal cold and warm spells in the winter at the Viet Nam northern mountain areas to serve for socio–economic development" (code: BDKH.25/16–20).

Competing interest statement: The authors declare no conflict of interest.

References

1. Ha, H.T.M.; Tan, P.V. The trend and variation of extreme temperatures in Viet Nam in the period of 1961–2007. *Sci. Technol. J. Ha Noi Univ. Sc.* **2009**, 412–422.
2. Lanh, N.V.; Thai, D.V. Prediction of maximum temperature and heat wave in Ha Noi in May–August months under the influence of western hot depression. *VN J. Hydrometeorol.* **2002**, *502*, 14–19.
3. Lanh, N.V. The impact of the depression patterns to summer weather in Viet Nam. *VN J. Hydrometeorol.* **2010**, *593*, 43–53.
4. Lanh, N.V. 2010. Heat wave and key weather patterns that cause heat wave in Viet Nam. *VN J. Hydrometeorol.* **2010**, *597*, 1–8.
5. Hang, P.M.; Dung, T.T.; Quang, N.D. The impact of western hot depression and Northwest Pacific Subtropical High Pressure to heat wave activities over Northern Central region during 2010–2015. *VN J. Hydrometeorol.* **2017**, *674*, 44–52.
6. Lan, H.P.; Dien, N.H. Prediction of maximum temperature by using ANN for northern delta region. *VN J. Hydrometeorol.* **2008**, *571*, 20–23.
7. Hoa, V.V.; Tien, D.D.; Duc, T.A.; Hung, M.K.; Quan, D.D.; Khiem, N.V.; An, N.V. Research on classifying typical synoptic patterns causing abnormal warm spells in early winter in northern area of Viet Nam by a Self–Organizing Map. *VN J. Hydrometeorol.* **2019**, *703*, 51–59.
8. Nishiyama, K.; Endo, S.; Jinno, K.; Uvo, C.B.; Olsson, J.; Bertndtsson, R. Idenfication of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a Self–Organizing Map. *Atmos. Res.* **2007**, *83*, 185–200.
9. Liu, Y.; Weisberg, RH. Sea surface temperature patterns on the West Florida shelf using growing hierarchical Self–Organizing Maps. *J. Atmos. Ocean Tech.* **2005**, *23*, 325–338.
10. Duc, T.A. Research on classifying heavy rainfall patterns in Viet Nam using Self–Organizing Map method. Climatology and Meteorology Thesis in master degree, Ha Noi University of Science, 2014, pp. 80.
11. Tuan, V.A. etc. Research on objectively classifying heavy rainfall patterns in Viet Nam. *Sci. Rep.* **2015**, pp. 179.
12. Seung–Yoon, B.; Kim, S.W.; Jung, M.I.; Roh, J.W.; Son, S.W. Classification of Heat Wave Events in Seoul using Self–Organizing Map. *J. Clim. Change Res.* **2018**, *9*, 209–221. Doi:10.15531/ksccr.2018.9.3.209.