WILDFIRE HOTSPOT CLUSTERING AND FALSE ALARM DATA DETECTION DERIVED FROM SUOMI NPP SATELLITE OVER THAILAND REGION

Thanarporn Khanpanya[1], Panupat Horma[1], Supavit Nounkhaow[1], Prasit Maksin[1], and Wasanchai Vongsantivanich[1]
[1]Geo-Informatics & Space Technology Development Agency (Public Organization),
88 Moo 9 Tambon Thung Sukala, Amphoe Siracha, CHONBURI 20230
Email addresses: thanarporn.kha@gistda.or.th, panupat.h@gistda.or.th, supavit@gistda.or.th,
prasit@gistda.or.th, wasanchaiv@gistda.or.th

KEY WORDS: False alarm filtering, Forest fire, VIIRS

ABSTRACT: Wildfire has recently become one of the most significant disaster in Thailand, especially in the northern part during the dry season; as it can be deadly, destroying homes, wildlife habitat and timber, and polluting the air with harmful emission on human health. The increasing number of burned area also worsen by climate change. This compounding effects leads to serious impact on ecosystem. Hence, space-based instruments have become the most essential means to monitor the hotspots due to their daily temporal revisit for nation-wide coverage data provision. Visible Infrared Imaging Radiometer Suite (VIIRS) instrument which onboards the Suomi NPP satellite is one of the most well-known instruments that have been used in wildfire monitoring. It is capable of providing full coverage of Thailand region at 375 m nominal resolutions of fire hotspot data. Although VIIRS instrument is suitable for detecting fire activity, some hotspot data are misinterpreted such as the ones in the factory area. Therefore, it is necessary to detect such false-alarm hotspot data. Additionally, the seasonal distribution of wildfire helps establishing the wildfire clustering map to observe continuation of wildfires/hotspot data for detecting wildfire propagation, and also to detect the damaged areas for assessing the vulnerability of the areas. False-alarm points and wildfire clustering observations can be a few formations of data management to help in reducing the risk of disaster. As in this study, false-alarm hotspot detection applied an idea of hotspot occurrence per pixel should have only one hotspot; therefore, a distance between pairs of a hotspot within one pixel range (375 m) will be suspected as false alarm points. Furthermore, wildfire clustering applied the idea of Density-based spatial clustering of applications with noise (DBSCAN) which is a grouping of continuous hotspots under a condition of specific distances derived from scan and track of each pixel.

1. INTRODUCTION

Fire played an important role in Thailand's forests clearance for agriculture in the last 20 years since it had been a major agricultural tool for many centuries in Southeast Asia. According to [12], wildfires commonly occur in the northern part of Thailand every dry season, and can be severe by effects of climate change every year. When wildfires occur, they can cause; (1) loss of habitat and wildlife species – like Doi Suthep-Pui National Park that has approximately 2000 species of ferns and flowering plants and about 360 species of birds in the park. (2) air pollution and health – wildfires have led to increased regional air pollution and reached unhealthy levels in several provinces in northern Thailand, which can result in health impacts such as respiratory problems and cardiovascular conditions [3]. There are several projects related to wildfire management, according to [2], the Royal Forest Department (RFD) fire management programme had maintained 85-km firebreaks with 10% of potential that can protect watershed area. Moreover, Community Coordinators (CCORD) directly linked the project and the villagers to conduct extension work related to community organization and awareness raising on fire management, which is the most important factor in reducing the number of and damage caused by fires. As two previous management projects of fire can be implemented in only specific areas, then space-based sensors have become the most essential means to monitor hotspots due to the necessity to monitor wide area for wildfires/hotspot occurrences for disaster risk reduction and their daily temporal revisit for nation-wide coverage data provision.

The Visible Infrared Imaging Radiometer Suite (VIIRS) is the new generation U.S. operational imager with finer spatial resolution and radiometric quality, which is the payload on Suomi-NPP satellite and it is the main sensor for hotspot monitoring in Thailand. The 375-m resolution of VIIRS active fire products are currently used due to their full finer spatial information; VIIRS is able to detect smaller fires. As VIIRS active fire products are presented in points of hotspot, clustering may need to be implemented to discover groups of hotspots where the distances between the members of each cluster are closer than to members in other clusters, to help delineate perimeters and direction of ongoing large fires. Incidentally, there is a major of limitations of using VIIRS active fire products to be considered, which is false alarm detections. The occurrence of false alarms can affect to data interpretation which relates to work operation of fire management – misinterpretation of data can waste time of onsite work operation.

In this study, we present a methodology to generate wildfire cluster and to detect false alarms (which is in section 3). Furthermore, an overview of VIIRS I-Band active fire product and related researches will be presented in Literature Reviews section (section 2), and results and discussions will be held in section 4 and 5, respectively.

2. LITERATURE REVIEW

This section presents an overview of VIIRS I-Band active fire product and existing methods in the literature considering wildfire cluster and false alarm detections.

2.1 VIIRS I-Band Fire Product

Suomi NPP (S-NPP) satellite was placed in a sun-synchronous orbit with an equator crossing local time of ~1:30 PM (ascending node) and ~1:30 AM (descending node). The sun-synchronous orbit is a unique path that takes the satellite over the equator at the same local (ground) time in every orbit. The VIIRS instrument on board S-NPP satellite with 375 m (I-bands) and 750 m (M-bands) resolution data.
The I-band data of VIIRS consist of five distinct single-gain channels (Table 1).

Channel	Spatial Resolution (m)	Spectral Resolution (µm)	Primary Use
I1	375	0.60 – 0.68	Cloud and water classification
I2	375	0.846 – 0.885	Cloud and water classification
I3	375	1.58 – 1.64	Water classification
I4	375	3.55 – 3.93	Fire detection
I5	375	10.5 – 12.4	Fire detection and cloud classification

Table 1 I-band (375 m) data of VIIRS spectral intervals for all channels

The pixel of VIIRS varies in size with an increasing of scan angle, typical to other sensors such as AVHRR and MODIS, and will be minimized due to a unique data aggregation scheme performed for all VIIRS channels [8][11]. This scheme describes the nominal resolution after native pixels are spatially aggregated across three distinct regions (Figure 1). The native I-band image resolution prior to onboard aggregation is approximately 125 × 375 m at nadir. In the first region (scan angles from 0° (nadir) to ± 31.72°), every three native pixels are aggregated in the along scan (cross-track) direction to form one effective data sample. In the second region (scan angles from ± 31.72° to ± 44.86°), every two native pixels are aggregated to form one effective data sample. Finally, no aggregation is performed in the third region (scan angles from ± 44.86° to ± 56.28°) such that each native pixel represents one effective sample. Consequently, the effective footprint ranges from the nominal 375 m resolution (383 × 360 m) at the sub-satellite point to 795 × 784 m at a maximum scan angle of 56.28° [7][8].

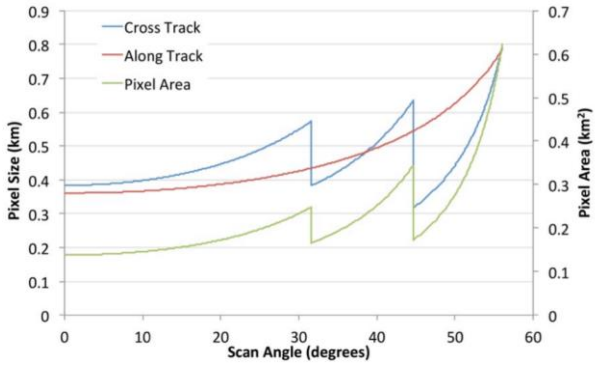

Figure 1 Spatial resolution of VIIRS imager data (I bands) as a function of scan angle. The three distinct regions describe data aggregation zones extending from nadir to the edge of the swath [7].

2.2 Cluster Analysis

• Machine Learning Clustering

Cluster analysis, or clustering, is an unsupervised machine learning task that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. There are several types of clustering algorithms such as K-Means, Mean Shift, DBSCAN, Agglomerative Hierarchical, etc., but we will only mention to K-Means and DBSCAN in this section.

• K-Means Clustering: K-Means algorithm is the most well-known iterative algorithm belongs to partitioning-based techniques grouping with *k* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. Furthermore, K-Means algorithm is also a popular data clustering algorithm which requires the number of clusters in the pre-determined data. Finding the appropriate number of clusters is the same as a trial-and-error process which made more difficult by the subjective nature of deciding what constitutes correct clustering. There is one major limitation of K-Means algorithm that is not suitable for being applied to this study that is to define number of clusters *k.*

According to [5] and [6], scan statistic is originally introduced in the field of health sciences by J. I. Naus as a family of local methods for cluster detection allowing to exactly locate clusters in space and/or in time and to test for their statistical significance. Then this scan statistic was implemented by Kulldorff which developed the SaTScan™ software and later integrated the spatial-temporal extensions. As Kulldorff's approach was to solve the problems about size and geographic distribution of cluster, means that if we look for spatial clusters of a population disease, the expected number of cases in an area is not necessarily proportional to the size of that area, but the observed distribution of the population at risk has to be considered. From [6], a brief introduction of scan statistic methods is given as follow:

• **Spatial scan statistics:** The basic module integrated in the SaTScan™ software, consists of a purely spatial analysis aiming at detecting spatial clusters. To identify areas inside region Z with higher aggregation of events c_i, the region Z will be scanned by circular moving windows with increasing of the radius r up to a fixed limit (r_{max}) on a different event-location. While a window scans the whole region Z, the number of events falling inside and outside the circular window is calculated (Figure 2).

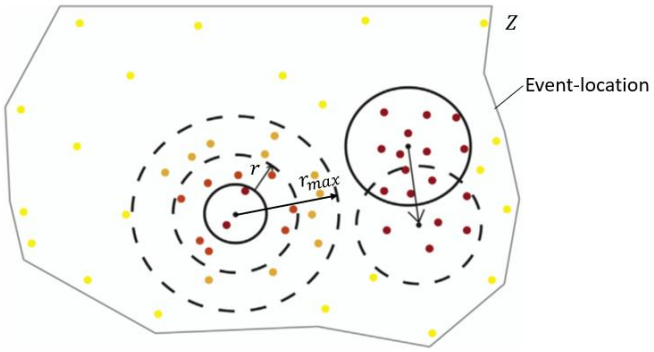

Figure 2 The scanning windows with increasing their radius (circles on the left), and moving around the study region (circles on the right) [6].

This is similar to the idea of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering technique for which the number of clusters is not needed to be pre-determined. From Figure 2, the way of the scanning windows is with increasing radius and move around the study area can be as finding cluster relying on only one of parameters of DBSCAN namely *eps* (the maximum distance between two samples for one to be considered as in the neighborhood of the other), which is applied in this study.

• **Spatio-temporal scan statistic:** The spatio-temporal scan statistic uses the scanning cylinders instead the circular windows to detect the cluster (Figure 3). The height of the cylinder is a measure of the time dimension (t) while the area of the base remains as the measure of the geographical spatial dimension ($s = \pi r^2$). Both the height, t, and the base, s, of the cylinder grow up to maximum values. As there is the time dimension, the observed cases should be aggregated into time intervals such as a week, a month, or a year, to reduce the computing time.

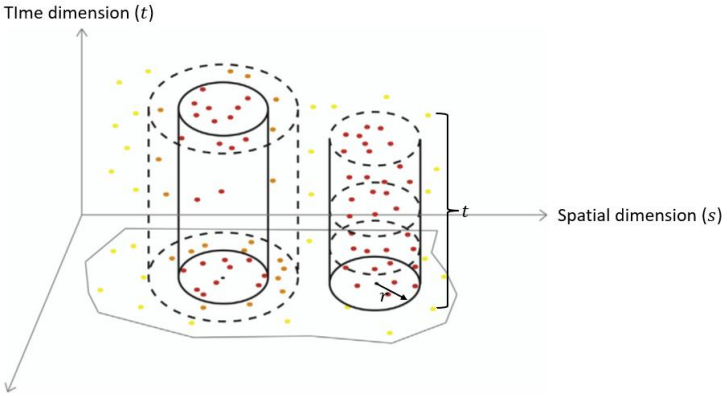

Figure 3 The scanning windows increasing their radius (on the left), and their height based on a time-frame (on the right) [6].

In this study, we also applied the idea of spatio-temporal scan statistic to both false alarms detection and wildfire clustering as we considered the time dimension to observe the repeated occurrence of hotspots in the nearby location.

• **Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Clustering:** DBSCAN is a clustering algorithm with no supposition needed about number of clusters and not relying on polygons and requires two input parameters namely *eps* and *MinPts*; *eps* represents the maximum distance between two samples for one to be considered as in the neighborhood of the other, and *MinPts* represents the minimum number of points to be considered as a member of a cluster within the radius of *eps*. DBSCAN detects groups of clusters by giving each data point a circle with the radius of determined *eps*, followed by the membership evaluation for each point enclosed in that circle. There are three categories that a point would be assigned to; a core point, border points, and noise points. A point will be assigned as core points when it has at least a number of member (enclosed) point equal to *MinPts* within radius *eps*. A point will be assigned as border points when a point is within *eps* but has a number of member points of less than *MinPts*. Finally, if the point doesn't belong to the core or border points, it is assigned as noise points. Noise points are not a member of any clusters. A cluster is then defined as a set membership containing a combination of core points surrounded by border points [1]. The pseudocode for the DBSCAN algorithm is presented in Figure 4.

```
DBSCAN(D, eps, MinPts)
  C = 0
  for each unvisited point P in dataset D
    mark P as visited
    N = getNeighbors (P, eps)
    if sizeof(N) < MinPts
      mark P as NOISE
    else
      C = next cluster
      expandCluster(P, N, C, eps, MinPts)

expandCluster(P, N, C, eps, MinPts)
  add P to cluster C
  for each point P' in N
    if P' is not visited
      mark P' as visited
      N' = getNeighbors(P', eps)
      if sizeof(N') >= MinPts
        N = N joined with N'
    if P' is not yet member of any cluster
      add P' to cluster C
```

Figure 4 DBSCAN Algorithm [10].

2.3 False Alarm Detection

According to Schroeder et al. (2014), false alarms can occur at the areas of high solar reflection, resulting in peak of brightness temperature on channel I4 over large and bright surfaces such as concrete pavement, and at water bodies over which specular reflection (Sun glint) may occur. Furthermore, [8] also reported that a high rate of low confidence daytime active fire pixels was detected by active fire detection algorithm in eastern China. And the analysis has shown that many of the low confidence active fire pixel detections are in fact false alarms associated with large industrial buildings, having highly reflective and/or warm rooftops and surrounded by more rural landscapes.

To optimize false-alarm detection algorithm, [13] had developed landcover and hotspot persistence-based spatial filters from GlobeLand30 global landcover mapping data with layers that include 'places', 'buildings', 'landuse' and 'roads' used to identify urban areas. In addition, false alarms were masked as the thermal anomaly pixels persistently (and unrealistically) detected at multiple times in the same locations, and again found that these were mostly related to the type of industrial buildings. As the definition of false alarms that [13] provided, we can use this definition to help analyzing false alarms in this study.

3. MATERIALS AND METHODOLOGY OF WILDFIRE CLUSTERING AND FALSE ALARM DETECTION

This section presents an overview of study area and data used in this study, and methodology of wildfire cluster and false alarm detections.

Materials

3.1 Study Area

Thailand is located in Southeast Asia between latitudes 5˚ 37′ N and 20˚ 15′ N and longitudes 97˚ 22′ E and 105˚ 37′ E. Physiographical, the country can be divided into five regions; the northern, northeastern, central, eastern and southern regions, with the total 77 provinces [4]. In Thailand, forest cover has declined from 53.5% in 1961 to 31.6% in 2014 as a result of population growth, infrastructure development, agricultural expansion, illegal logging and uncontrolled forest fires.

Figure 5 Study area location map
(https://east-usa.com/world/thailand-map.html)

3.2 VIIRS I-Band Fire Product (used in this study)

Visible Infrared Imaging Radiometer Suite (VIIRS) instrument was firstly launched in October 2011 aboard the Suomi-National Polar-orbiting Partnership (S-NPP) satellite. Two sets of full global coverage were provided at 375 m and 750 m nominal resolutions every 12 h or less depending on the latitude. The VIIRS instrument incorporates fire-sensitive channels, including a dual-gain, high-saturation temperature 4 μm channel, enabling active fire detection and characterization. In this study, 375 m (I-band) moderate resolution data of VIIRS (375 m active fire data; VNP14IMG) were dominantly used due to its higher spatial resolution, thus the VNP14IMG active fire product can capture more fire pixels than any active fire products [9].

Methodology

3.3 Wildfire Hotspot Cluster Detection

To detect wildfire clusters, HashMap data structure was applied to reduce the processing time. For each data point corresponds to grid index (i, j), we chose 8-neighbor-grids around (i, j) as illustrate in Figure 6.

Actual Pixel Size

The actual pixel size of hotspots needs to be informed to define the distance used in grouping wildfire clusters. Since the actual pixel size varies with the scan and track, then the distance can be calculated from the formula of distance between two points with an adjustment of $\Delta x, \Delta y$ to be scan and track.

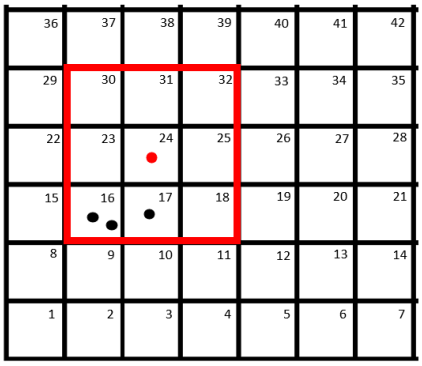

Figure 6 Boundary of an interested hotspot determination using HashMap

3.4 False Alarms Detection

To observe false alarms, we defined false alarms in this study as a group of hotspots which occurred in the same region consistently for a given period of time (set to be one month in this study). Therefore, false alarms can be detected by being a group of hotspots within the radius of 375 m repeatedly appears in the nearby region for more than 30 days per year (empirically set) with spatial and temporal detections.

Spatial False-Alarms Detection

We used the method of cluster detection (applied HashMap data structure) to help detecting false alarms spatially (Figure 6). After having the window of hotspots, calculate the distance between the pair of interested hotspot and neighbor hotspots using the Great Circle Distance with Haversine formula, given by:

$$D = 2R\ arcsin\sqrt{sin^2\left(\frac{x2-x1}{2}\right) + \cos(x1)\cos(x2)\ sin^2\left(\frac{y2-y2}{2}\right)} \quad (1)$$

where D is the distance between two points calculated from latitude and longitude, R is the Earth radius (≈ 6,372.8 km), $x1, y1$ is latitude and longitude of the first position, $x2, y2$ is latitude and longitude of the second position. Distances between the interested hotspot and nearby hotspots will be paired to each other using pairwise function to have Distance Matrix with the size of n^2, where n is the total number of hotspots.

Temporal False-Alarms Detection

Hotspots getting from spatial detection need to be considered as the hotspots that repeatedly appear for more than 30 days per year applied the idea of defining key from HashMap, which can be calculated as:

$$key(time) = [(j \times No.of\ column) + i] + [(No.of\ column\ \times No.of\ row) \times rev] \quad (2)$$

where $key(time)$ is key in temporal, i, j is the grid position along x and y axis, $No.of\ column$ is the total number of grid columns, $No.of\ row$ is the total number of grid rows, and rev is the satellite revolution across Thailand.

To observe false alarms in temporal domain, we need to consider spatial keys that consist of hotspots appear repeatedly over the varying time. In Figure 7, there are the iteration of hotspots in the same spatial keys which will be considered on the condition that there are more than or equal to two iteration hotspots. Consequently, there spatial keys number 16, 17, and 23 are considered with the period of occurrence for more than 30 days per year. In case if the both spatial and temporal conditions of false alarms detection are adapted, those hotspots will be considered as false alarms.

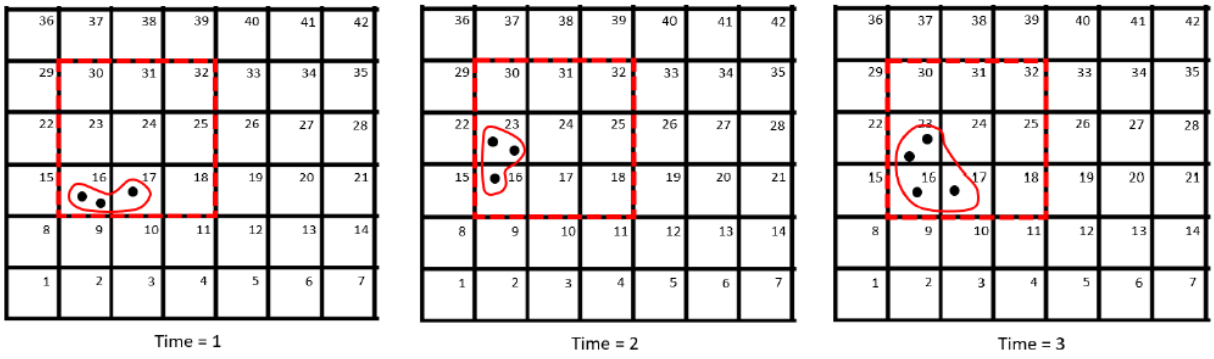

Figure 7 Repeated hotspots over the same spatial keys at the varying time (Time = 1, 2, and 3)

4. RESULTS

4.1 Wildfire Hotspot Cluster Detection

DBSCAN leads to identification of several clusters. The size of clusters will vary by the distance according to scan and track values of each data point and dates to consider continuation of hotspots that becomes clusters. The number of clusters will vary by the occurrence duration of wildfire, wildfires in Thailand dominantly occur during January to May of every year. As shown in Figure 8, the highest number of hotspots are in March, thus the number of clusters in March will be higher than other months.

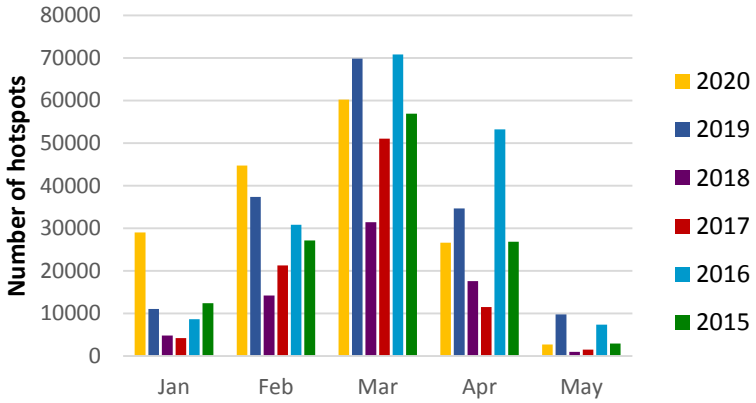

Figure 8 Annual number of hotspots in Thailand from 2015 to 2020

Figure 9 shows the sample of hotspot clusters which the distances of each hotspot calculated from the actual sizes of pixel which vary with the scan and track.

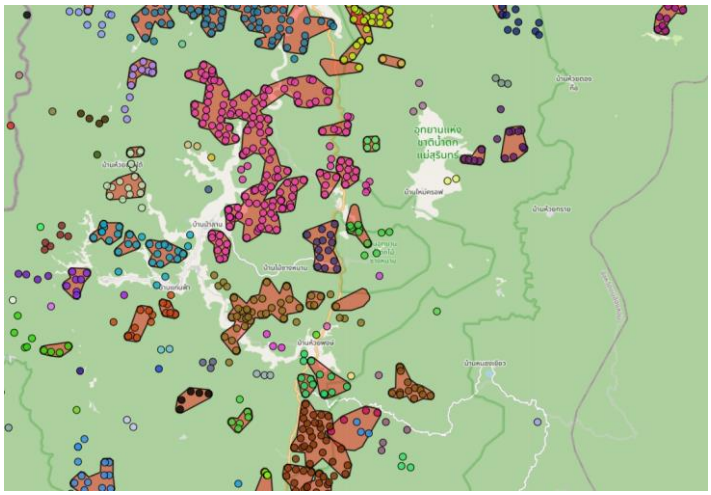

Figure 9 The wildfire hotspot clusters occurred in the regions of northern Thailand during March 1st to 11th, 2021

4.2 False Alarms Detection

After implementing the spatial and temporal detection to find false alarms, we found them and show some examples in Figure 10.

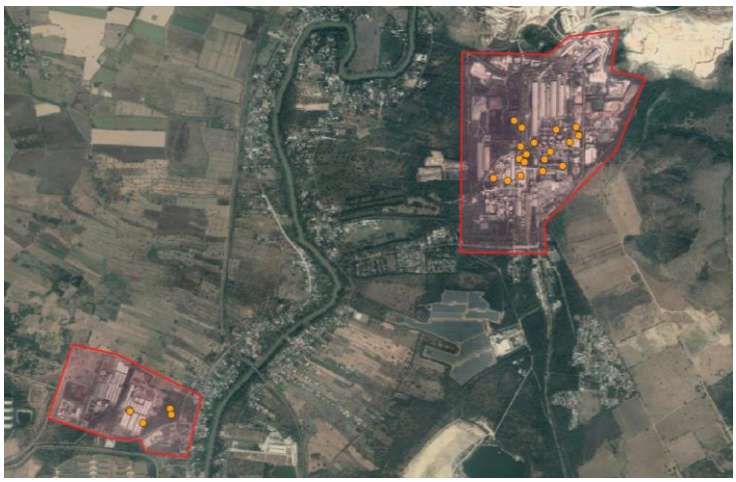

Figure 10 The sample of false alarms in the industrial building

In Figure 11, we have the groups of hotspots appear in the areas of industrial buildings (red-line boundary) with the distance of each hotspot is closer than 375 m and their occurrence is repeatedly for more than 30 days per year. To state that these hotspots are false alarms, we need to consider hotspot iteration in the same spatial key; as shown in Figure 8, there are hotspots in two spatial keys, which are key numbers 925018 (purple points) and 925019 (pink points). In key number 925018, hotspots occur since January 3rd, 2021 until May 29th, 2021, which the period of occurrence is adapted to state that these hotspots are false alarms or in the areas of high solar reflection.

Accordingly, false alarms over Thailand region are about 140 locations which are mostly found in the areas of industrial buildings, mine, solar farms, electric plants, roads/highways, and rooftops in urban areas.

Figure 11 False alarms in the industrial building with different key numbers: key number 925018 (purple points), key number 925019 (pink points)

5. DISCUSSION AND CONCLUSION

In this study, we analyze VIIRS active fire data over Thailand. We can detect wildfire hotspot clusters and false alarms using DBSCAN and found that the wildfire clusters which the number of clusters will vary by the occurrence duration of wildfire; the highest number of clusters will be in March due to the highest number of hotspots. Moreover, there are about 140 locations of false alarms over Thailand which are found in the areas of industrial buildings, mine, solar farms, electric plants, roads/highways, and rooftops in urban areas.

References

[1] Anwar M.T., Hadikurniawati, W., Winarno, E., Supriyanto, A., 2019. Wildfire Risk Map Based on DBSCAN Clustering and Cluster Density Evaluation. Advance Sustainable Science, Engineering and Technology.

[2] Hoare, P., 2004. A process for community and government cooperation to reduce the forest fire and smoke problem in Thailand. Agriculture, Ecosystems and Environment 104, pp. 35-46.

[3] Karen, N., 2020. Forest Fires Have Devastated Northern Thailand, Retrieved by September 24, 2020, from https://earth.org/forest-fires-have-devastated-northern-thailand/.

[4] Liengsiri, C., 2009. Forest in Thailand. The Royal Forest Department, Bangkok, pp. 1-46.

[5] Parente, J., Pereira, and M.G., Tonini, M., 2016. Space-time clustering analysis of wildfires: The influence of dataset characteristics, fire prevention policy decisions, weather and climate. Science of the Total Environment 559, pp. 151-165.

[6] Pereira, M.G., Caramelo, L., Orozco, C.V., Costa, R., Tonini, M., 2015. Space-time clustering analysis performance of an aggregated dataset: The case of wildfires in Portugal. Environmental Modelling & Software 72, pp. 239-249.

[7] Schroeder, W., and Giglio, L., 2016. Visible Infrared Imaging Radiometer Suite (VIIRS) 375 m Active Fire Detection and Characterization Algorithm Theoretical Basis Document 1.0. Department of Geographical Sciences, University of Maryland, pp. 1-21.

[8] Schroeder, W., Oliva, P., Giglio, L., Csiszar, I.A., 2014. The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. Remote Sensing of Environment 143, pp. 85-96.

[9] Vadrevu, K., and Lasko, K., 2018. Intercomparison of MODIS AQUA and VIIRS I-Band Fires and Emissions in an Agricultural Landscape—Implications for Air Pollution Research. Remote Sens. 978.

[10] Vijayalaksmi, S., Punithavalli, M., 2012. A Fast Approach to Clustering Datasets using DBSCAN and Pruning Algorithms. International Journal of Computer Applications 60(14).

[11] Wolfe, R.E., Lin, G., Nishihama, M., Tewari, K.P., Tilton, J.C., and Isaacman, A.R., 2013. Suomi NPP VIIRS prelaunch and on-orbit geometric calibration and characterization. Journal of Geophysical Research: Atmosphere 118, pp. 11,508–11,521.

[12] WWF-TH, Sustainable and Consumption Project, 2020. 2020 Northern Thailand forest fires snapshot, Retrieved July 4, 2020, from https://www.wwf.or.th/?362337/2020-Northern-Thailand-forest-fires-snapshot

[13] Zhang, T., Wooster, M.J., and Xu, W., 2017. Approaches for synergistically exploiting VIIRS I- and M-Band data in regional active fire detection and FRP assessment: A demonstration with respect to agricultural residue burning in Eastern China. Remote Sensing of Environment 198, pp. 407-424.