IL NUOVO CIMENTO **45 C** (2022) 81
DOI 10.1393/ncc/i2022-22081-0

COMMUNICATIONS: SIF Congress 2021

Automatic detection of volcanic ash clouds using MSG-SEVIRI satellite data and machine learning techniques

F. TORRISI([1])([2])([*])

([1]) *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Osservatorio Etneo Catania, Italy*
([2]) *Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Università degli Studi di Catania - Catania, Italy*

received 31 January 2022

Summary. — Volcanic ash emissions can pose serious hazard to population living at the edge of an active volcano and can cause widespread disruption to aviation operations. Here an innovative machine learning (ML) approach, developed in Google Earth Engine (GEE), is proposed to detect volcanic ash clouds. It exploits the MSG-SEVIRI (Meteosat Second Generation - Spinning Enhanced Visible and Infrared Imager) images in the Thermal Infrared (TIR) range. This ML procedure was applied to the sequence of paroxysmal explosive events occurred at Mt. Etna between February and March 2021. It was demonstrated that machine learning algorithms combined with high temporal resolution satellite data offer a good solution to automatically detect, track and map a volcanic ash cloud.

1. – Introduction

During explosive volcanic eruptions, large volumes of ash and gases can be ejected into the atmosphere, forming volcanic ash and gas clouds that are transported over great distances by the wind. The dispersion of volcanic ash in the atmosphere represents a serious threat for aviation safety [1], whereas the tephra fallout, together with gas emission, may strongly affect population health [2] and damage the environment as well as infrastructures [3].

Until now, different techniques and instrumentation have been employed to measure volcanic emissions. *In situ* measurements are typically confined to particles measurements and sulphur dioxide concentrations in the atmosphere [4], whereas Global Positioning System signal strength (signal-to-noise ratio, GPS-SNR) data plays a critical

([*]) E-mail: federica.torrisi@ingv.it, federica.torrisi@phd.unict.it

role for detecting clouds, since they are very sensitive to large ash particles, providing a clearer picture of a volcanic plume [5]. Other presented methodology concerns the development of video surveillance systems exploiting infrared cameras for monitoring eruptive phenomena in volcanic areas [6].

To date, satellite remote sensing provides one of the most temporally and spatially practical tools for identifying and tracking volcanic clouds. Space observations traditionally use the Thermal Infrared (TIR) bands (located around $8.5\,\mu$m and $12.0\,\mu$m), as these have proved to be the most suitable to distinguish volcanic ash from meteorological clouds. Thus, different satellite techniques have been developed to monitor volcanic clouds drifting in the atmosphere. For example, the 2-Band technique [7, 8] is based on a threshold test: the positive or negative values of the brightness temperature difference (BTD) between the channel at $11.0\,\mu$m and the channel at $12.0\,\mu$m indicate respectively the absence or the presence of ash particles. The 2-Band method is very powerful under the supervision of a user, but it is ineffective for automated ash cloud detection as artefacts could lead to a large amount of false alarms [9]. Therefore, other satellite remote sensing techniques have been developed to identify volcanic clouds, such as the 3-Band method [10, 11], based on the use of three spectral bands centered at $8.5\,\mu$m, $11.0\,\mu$m and $12.0\,\mu$m, on the effective absorption optical depth ratios, known as *β-ratios*, which give quantitative information on whether individual pixels are affected by volcanic ash ($\beta(8.7, 11)$ and $\beta(12, 11)$), and on brightness temperature differences ($\mathrm{BTD}_{11\text{–}12}$ and $\mathrm{BTD}_{11\text{–}8.7}$). Therefore, threshold-based techniques are commonly employed to assess if a pixel is contained in a volcanic cloud. It is more appropriate to replace static threshold methods by techniques in which thresholds dynamically adapt to the image scene characteristics, which result less prone to false alarms [12]. Hybrid algorithms were also developed for the volcanic cloud detection, by combining the use of BT threshold tests and look up table (LUT) based inversion calculations [13]. Setting threshold for volcanic cloud detection is problematic, but computer methods including statistical measures or machine learning algorithms can be well-suited to this task [14].

With the launch of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument on board of the Meteosat Second Generation (MSG) geostationary satellites (from ESA EUMETSAT), it is now possible to obtain high-temporal and high-spectral resolution of short and rapidly evolving eruptive phenomena from space. In addition, the plentiful data offered by geostationary MSG-SEVIRI sensor (with a standard refresh frequency of 15 min) is leading to the increased use of data-driven approaches, including machine learning (ML) techniques, to develop an automated satellite approach for the detection of volcanic clouds.

Nowadays, ML techniques offer new perspectives in terms of processing a large amount of data and of solving complex problems. The combination of machine learning and satellite remote sensing in volcano monitoring has the potential of analyzing global data in near real-time for mapping and monitoring purposes, but also for hazard analysis. Here, a first step towards the development of an algorithm which automatically detects and tracks volcanic clouds is presented. This tool is written in Google Earth Engine (GEE) [15] and ML techniques are employed, exploiting the TIR bands of the images acquired by the sensor SEVIRI. Two classifiers were designed, one unsupervised and the other supervised, both with the purpose of detecting the volcanic clouds. These two techniques were compared to determine which was the most suitable to perform this function. As the supervised classifier showed greater ability, then a supervised model was built and trained for the purpose of detection. These algorithms were applied to the sequence of paroxysmal events occurred at Mt. Etna between February and March 2021.

2. – Materials

2˙1. *Case studies.* – Between mid-February and late March 2021, the eruptive activity of Mt. Etna (Sicily, Italy) was characterized by a series of paroxysmal lava fountain events of short duration (a few hours) and high intensity. The series of events took place from the South-East Crater (SEC), located in the summit area, at almost regular frequency, around every 2–3 days within 2 months (from INGV weekly bulletins at `www.ct.ingv.it`). All the events are generally characterized by a starting phase with a gradual increase of the explosive activity and a subsequent paroxysmal phase, in which there is an intensification of explosions with the formation of an eruptive column feeding a sustained ash plume. At Etna, lava fountains are always associated with the formation of large ash plumes rising up to several kilometres above the crater and spreading by the wind in the atmosphere. The explosive events considered in this study are those occurred on 16th, 18th, 19th, 21st, 23rd, 24th and 28th February 2021 and on 12th March 2021. Eruptive columns and ash plumes, associated with these lava fountains, caused a widespread fallout of ash transported beyond the boundaries of Sicily and southern Italy and to Greece and Malta.

2˙2. *Data source.* – The geostationary instrument MSG-SEVIRI has the capability of measuring radiation in 12 spectral channels covering the visible (VIS) to the infrared (IR) part of the spectrum [16]. The radiances in the different thermal channels are converted to brightness temperatures (BT). For the retrieval, only the infrared channels were considered, such that they can be applied during day and night. The spectral bands used are the TIR bands centered at 8.7 μm, 10.8 μm and 12.0 μm. The spatial resolution of SEVIRI ranges from 3 km at the equator to 4.5 km at Mediterranean latitudes. The temporal resolution is 15 min for the full disc and 5 min in rapid scan mode, which covers mainly Europe. The high temporal resolution offered by SEVIRI enables almost continuous monitoring of drifting volcanic ash clouds.

3. – Methods

Since the late 1970 s, volcanic clouds are monitored using geostationary satellites [17]. Typically, RGB (Red, Green, Blue) composite images are used for highlighting specific features of volcanic ash clouds, as colors allow discriminating between the different components (like thin volcanic ash, SO_2 gas plume, mixed ash and SO_2 gas, water cloud) [18]. However, the detection of a volcanic cloud is not a simple problem, since it is difficult to distinguish ash clouds and thin cirrus clouds [19]. Moreover, volcanic clouds may have a different consistency depending on the type of eruption. The detection task has so far mainly been performed using threshold methods with the supervision of an operator, but it could be very useful to apply ML approaches to make this procedure automatic and applicable to a large amount of data. ML techniques have already been successfully used to analyze satellite images for volcanic monitoring applications in [20-23].

In this work, ML techniques are implemented in GEE to identify volcanic clouds, thanks to the adoption of TIR bands that when opportunely combined allow discriminating more clearly a volcanic cloud and its components from the background. In fig. 1, the algorithm steps from the combination of the SEVIRI TIR bands to the detection of volcanic clouds through the application of ML algorithms are presented.

3˙1. *Ash RGB images composite.* – An Ash RGB image is an image composite designed to detect volcanic ash and SO_2 gas, which are the main components of a volcanic cloud.

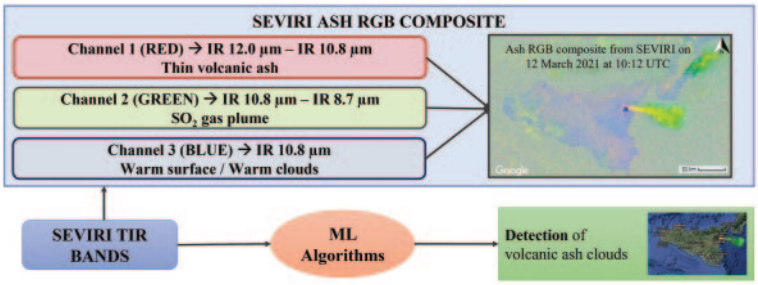

Fig. 1. – Steps of the procedure implemented in GEE. Combining the three SEVIRI TIR bands (IR 8.7 μm, IR 10.8 μm and IR 12.0 μm), Ash RGB images are realized, which allow recognizing some components of the volcanic clouds thanks to their different pixel colors (specifically, red indicates the presence of pure ash, green the presence of pure SO_2, while yellow indicates a mix of ash and SO_2). These images are processed by ML algorithms in order to detect volcanic clouds.

This type of composite can be generated by combining the SEVIRI TIR bands centered at 8.7 μm, 10.8 μm and 12.0 μm in this way:

- Red Channel: $\mathrm{BTD}_{12.0-10.8} = \mathrm{IR}\ 12.0\,\mu\mathrm{m} - \mathrm{IR}\ 10.8\,\mu\mathrm{m}$
- Green Channel: $\mathrm{BTD}_{10.8-8.7} = \mathrm{IR}\ 10.8\,\mu\mathrm{m} - \mathrm{IR}\ 8.7\,\mu\mathrm{m}$
- Blue Channel: IR 10.8 μm

The $\mathrm{BTD}_{12.0-10.8}$ highlights the presence of thin volcanic ash, which tends to have a strong reddish color, while $\mathrm{BTD}_{10.8-8.7}$ emphasizes the presence of SO_2, which is indicated by green pixels; instead, the regions of a volcanic cloud characterized by both ash and SO_2 are marked by yellow pixels. The shades of the color of an Ash RGB image depend on the concentrations of the elements that each color indicates. Thin volcanic ash has a good color contrast against water and ice clouds, whereas SO_2 has a good color contrast against ice clouds. The aim to create an Ash RGB image is to provide easily understandable visual information, using intuitive colors to recognize the different components of the clouds [24]. However, there are limitations, such as the dependence of the colors on the satellite viewing angle and the difficult identification of ash and SO_2 when they are mixed with cirrus clouds.

In fig. 2, three examples of SEVIRI Ash RGB images are reported and these events are considered as reference in this study, since the type of volcanic cloud is different for

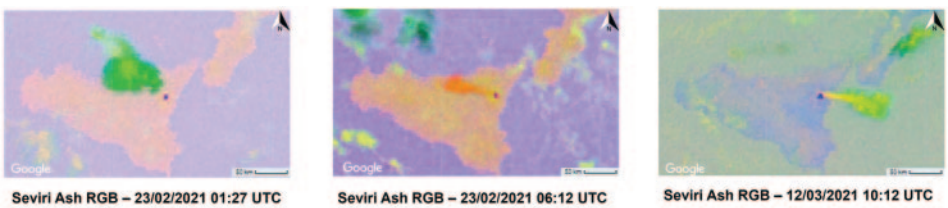

Fig. 2. – SEVIRI Ash RGB images.

each of them: a) the case of 23 February 2021 01:27 UTC depicts a pure-SO_2 cloud, b) the event of 23 February 2021 06:12 UTC a pure-ash cloud, and c) the episode of 12 March 2021 10:12 UTC a volcanic cloud characterized by both ash and SO_2.

3˙2. *Machine learning techniques.* – The satellite data were analyzed using two types of ML techniques, unsupervised and supervised. Unsupervised learning algorithms exploit unlabelled data in order to extract common features based on the similarity of the data, whereas supervised machine learning techniques use labelled data sets to train the algorithm, which after the training phase becomes capable to be applied to a new set of data. The idea is to apply both unsupervised and supervised algorithms for the detection of volcanic clouds and to evaluate which of the two approaches is more suitable. The algorithms employed are the unsupervised K-means algorithm (K-means) and the supervised Support Vector Machine (SVM).

3˙2.1. Unsupervised algorithm. The K-means algorithm is an unsupervised ML approach, used to analyze and cluster unlabelled data sets, without the need of human intervention; the unlabelled data are grouped according to their similarity [25]. The number n of clusters to partition our dataset is established a priori by the user. The goal is to discriminate the volcanic clouds from the background, then a number of cluster $n = 2$ is chosen. K-means is an iterative algorithm, which tries to assign data points to a cluster such that the sum of the squared distance between the points and the cluster centroids is at the minimum. First of all, cluster centroids are randomly initialized and each observation is allocated to the nearest cluster, according to the Euclidean distance. Subsequently, the cluster centroids are computed another time by taking the average of all the data points that belong to each cluster. These steps are repeated until an optimal convergence is reached.

The K-means lends itself well to the detection of volcanic ash clouds, since in the TIR bands they have generally spectral characteristics easily distinguishable from those of the background. As it is shown in fig. 3, the three TIR bands are fed to a K-means unsupervised classifier, which groups all the pixels of the image in two cluster: volcanic ash cloud (in blue) and background (in white).

3˙2.2. Supervised algorithm. The Support Vector Machine (SVM) is a supervised ML approach, used for classification, regression and outliers detection. The objective of this algorithm is to find a hyperplane in an N-dimensional space (N is the number of features), which is able to classify the data points. The design parameters for SVM are the kernel type, which function is to take data as input and transform them into the required form, the parameter γ, which is the "spread" of the kernel, and the cost c, which is the penalty for misclassifying a data point [26].

The SVM algorithm implemented in this work for the detection of volcanic clouds has a Radial Basis Function (RBF) as kernel, with $\gamma = 0.5$ and $c = 10$. It exploits the SEVIRI Ash RGB images and uses information provided by the combination of TIR bands to catch the main features related to volcanic clouds. In order to train the classifier to discriminate volcanic clouds from the background, a training dataset was gathered. It consisted of small areas belonging to the volcanic clouds and to the background, which are manually drawn in GEE from the three SEVIRI Ash RGB images of reference (23 February 2021 01:27 UTC, 23 February 2021 06:12 UTC, 12 March 2021 10:12 UTC). Thanks to the intuitive colors of the Ash RGB images, it was easy to recognize and to select correctly the regions inside the cloud as training samples, taking in mind that red, green and yellow pixels correspond respectively to pure-ash, pure-SO_2 and the mixture

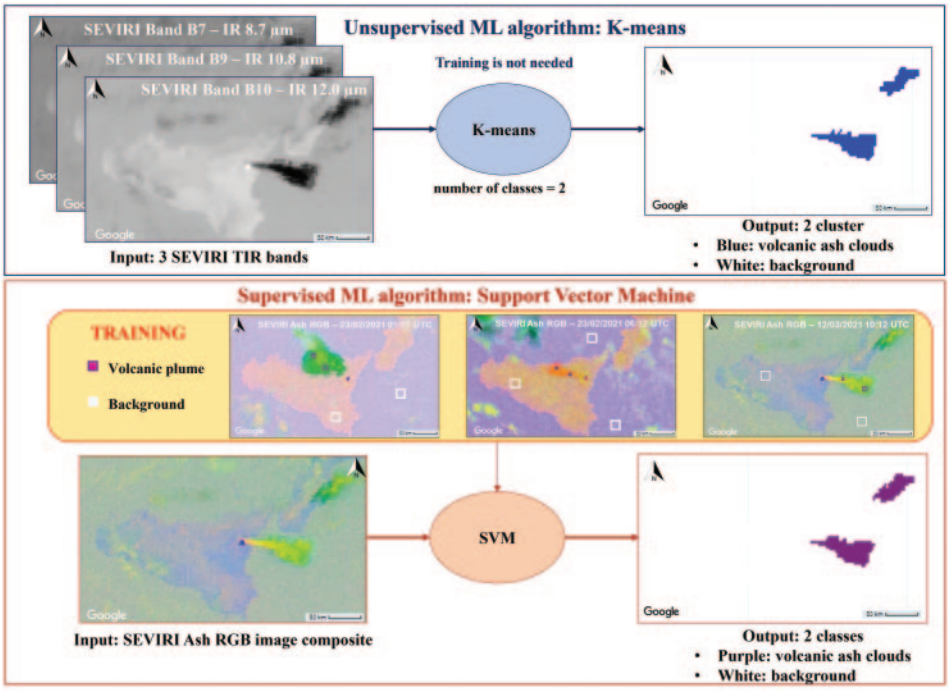

Fig. 3. – Scheme of the two ML approaches developed. At the top, the unsupervised K-means algorithm, which has as input the 3 SEVIRI TIR bands centered at 8.7 μm, 10.8 μm and 12.0 μm and gives as output two clusters, the volcanic ash cloud and the background. At the bottom, the supervised SVM algorithm, which uses as input a SEVIRI Ash RGB image, exploits a training dataset and produces as result two classes, the volcanic ash cloud and the background.

of ash and gas. In the selection of training samples belonging to the volcanic clouds, particular care was taken not to include any pixels belonging to meteorological clouds or to cirrus clouds, so that the recognition of a volcanic cloud is as accurate as possible. In fig. 3, the steps of the implemented SVM algorithm are schematically reported, showing that it uses as input SEVIRI Ash RGB images and as training dataset samples from the three SEVIRI Ash RGB images of reference. It produces as output the classification of volcanic clouds (in purple) and of the background (in white). After the training phase, the classifier is ready to be applied to new images.

4. – Results and discussions

The performances of the two ML algorithms proposed for the detection of volcanic clouds, one unsupervised (K-means) and one supervised (SVM), were compared to establish which is the main one suitable for this purpose. These algorithms were applied to the three images served as reference images (23 February 2021 01:27 UTC, 23 February 2021 06:12 UTC, 12 March 2021 10:12 UTC). The K-means does not need any training dataset and uses as input the three SEVIRI TIR bands (8.7 μm, 10.8 μm and 12.0 μm). On the contrary, the SVM exploits some sample regions as the training set from the three SEVIRI Ash RGB images of reference and uses as input SEVIRI Ash RGB images, as it is shown in fig. 3.

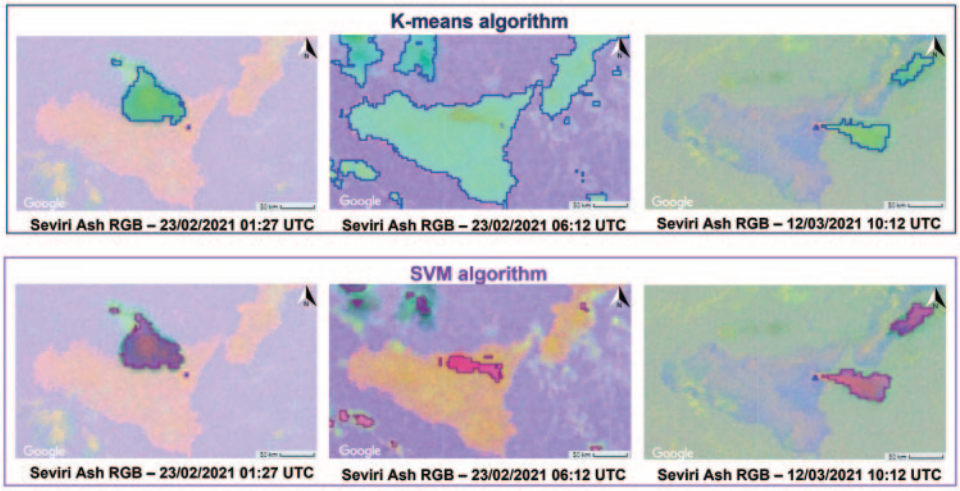

Fig. 4. – K-means and SVM outcomes of 23 February 2021 01:27 UTC, 23 February 2021 06:12 UTC and 12 March 2021 10:12 UTC events from GEE for the detection of volcanic clouds.

In fig. 4, the results of the application of the two algorithms are reported. In order to evaluate the performance and the robustness of the two ML classifiers, accuracy is used as the performance index and is defined as

$$Accuracy = \sqrt{\frac{A(test \cap real)}{A(test \cup real)}}, \tag{1}$$

where $A(test \cap real)$ and $A(test \cup real)$ are the areas of the intersection and union between the calculated and actual plumes, respectively. The actual area for each case study is manually retrieved through the visual analysis of the image. This method to calculate the accuracy is the same used in [27-29] to evaluate the ability of ML classifiers to map lava flows. The lower and upper bounds for this performance index are zero and one. A fitting score equal to one corresponds to a complete overlap; *i.e.*, the calculated area is perfectly coincident with the actual plume top area. On the other hand, a fitting score equal to zero indicates the maximum error, *i.e.*, lack of common areas between the calculated and actual plume top area. Table I shows the plume top area and the accuracy calculated applying the K-means and the SVM algorithm, for the events of 23 February 2021 01:27 UTC, 23 February 2021 06:12 UTC and 12 March 2021 10:12 UTC.

The application of K-means to the events of 23 February 2021 01:27 UTC and of 12 March 2021 10:12 UTC give results with a good level of reliability, the volcanic clouds and the background are completely distinguishable. In these cases, the K-means results slightly overestimate the real areal size of the volcanic plumes, in fact the area calculated is of 4416.55 km^2 for 23 February 2021 01:27 UTC and 2072.88 km^2 for 12 March 2021 10:12 UTC event, unlike the actual area values equal respectively to 3358.74 km^2 and 2482.19 km^2. Nevertheless, the accuracy is very high, respectively 0.90 and 0.91. An unreliable outcome is given by the application of the K-means to the case of 23 February 2021 06:12 UTC, where the volcanic cloud, characterized by pure-ash, is not detected: it

TABLE I. – *Areal extent of volcanic clouds of 23 February 2021 01:27 UTC, 23 February 2021 06:12 UTC and 12 March 2021 10:12 UTC estimated using the K-means and the SVM classifiers. The goodness of these approaches is evaluated by the accuracy index.*

Events	Real Area [km^2]	Area K-means [km^2]	Area SVM [km^2]	Accuracy K-means	Accuracy SVM
23 February 2021 01:27 UTC	3358.74	4416.55	3534.59	0.91	0.94
23 February 2021 06:12 UTC	1134.49	–	963.87	–	0.85
12 March 2021 10:12 UTC	2482.19	2072.88	2171.77	0.90	0.92

means that in the presence of ash-only clouds there are no spectral characteristics that allow us to discriminate volcanic clouds from the background without human effort.

On the other hand, the results obtained using the SVM algorithm are acceptable for all the three cases, in fact the accuracy is always higher than 0.80. Also the pure-ash cloud of 23 February 2021 06:12 UTC was detected and this demonstrates a better reliability of the SVM algorithm compared to the K-means algorithm. However, sometimes different kinds of clouds with similar optical properties can appear the same, leading the detection to errors, as it happens in fig. 4 for the event of 23 February 2021 06:12 UTC, where the clouds in the south-western part of the image are not related to the volcanic clouds. The discrimination of volcanic clouds from cirrus clouds is a crucial point, but surely by increasing the training samples and taking them with more precision, it is possible to overcome this limit and to improve the performance of the proposed supervised classifier. In any case, the results obtained by the SVM algorithm are satisfactory, given the high accuracy values (0.94, 0.85 and 0.92 respectively for the events of 23 February 2021 01:27

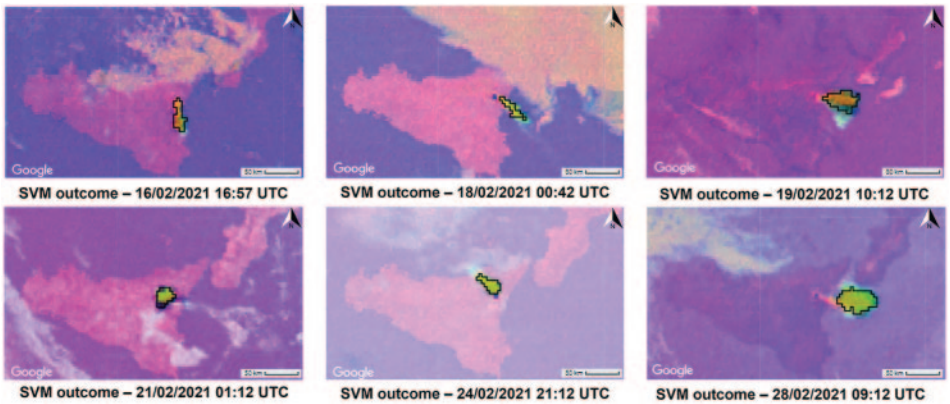

Fig. 5. – SVM outcome on 16, 18, 19, 21, 24 and 28 February 2021 from GEE for the detection of volcanic clouds. These study cases are used as testing phase, thus no training is performed.

UTC, 23 February 2021 06:12 UTC and 12 March 2021 10:12 UTC).

Finally, the SVM algorithm exhibits better performance than the K-means and for this reason it was chosen for the detection task, since it gives a positive and trustworthy contribution to the volcanic ash cloud identification. This supervised SVM model was applied to other paroxysmal events at Mt. Etna producing volcanic clouds, and specifically to those occurred during February 2021 (16, 18, 19, 21, 24 and 28 February 2021). These other study cases are used as a testing phase, thus no training is performed, and the results obtained, shown in fig. 5, demonstrate that the model is able to detect pretty well the volcanic clouds present in new images.

5. – Conclusions

An innovative ML approach, developed in Google Earth Engine (GEE), for identifying and tracking volcanic ash clouds using MSG-SEVIRI data has been presented. The high temporal resolution of the measurements from SEVIRI, together with good spatial resolution and spectral imaging capability, has permitted the accurate detection of volcanic clouds. Thermal Infrared remote sensing measurements have been used as inputs to two different Machine Learning algorithms. The comparison of the performances between unsupervised and supervised algorithms highlighted that the unsupervised K-means classifier is not always suitable to discriminate the plume from the background, *i.e.*, when the volcanic cloud is rich in ash. Thus, the supervised SVM approach was preferred and employed for the detection of volcanic clouds. The SVM classifier was trained on portions of volcanic clouds produced during the events at Mt. Etna of 23 February and 12 March 2021 and was tested on events not used during the training phase (paroxysms occurred on February 2021), showing a good ability to identify the overall volcanic cloud.

* * *

This work was developed within the framework of the Laboratory of Technologies for Volcanology (TechnoLab) at the INGV in Catania (Italy). We are grateful to EUMETSAT for SEVIRI data (`https://data.eumetsat.int`).

REFERENCES

[1] Casadevall T. J., *J. Volcanol. Geotherm. Res.*, **62** (1994) 301.
[2] Thordarsson H. *et al.*, *J. Geophys. Res.*, **108** (2003) 4011.
[3] Robock A., *Rev. Geophys.*, **38** (2000) 191.
[4] Flentje H. *et al.*, *Atmos. Chem. Phys.*, **10** (2013) 10085.
[5] Larson K. M., *Geophys. Res. Lett.*, **40** (2013) 2657.
[6] Andò B. *et al.*, *Comput. Geosci.*, **32** (2006) 80.
[7] Prata A. J., *Int. J. Remote Sens.*, **10** (1989) 751.
[8] Prata A. J., *Geophys. Res. Lett.*, **16** (1989) 1293.
[9] Guéhenneux Y. *et al.*, *J. Volcanol. Geotherm. Res.*, **293** (2015) 25.
[10] Pavolonis M. J., *J. Appl. Meteorol. Climatol.*, **49** (2010) 1992.
[11] Francis P. N. *et al.*, *J. Geophys. Res.*, **117** (2012) D00U09.
[12] Prata F. *et al.*, *Atmosphere*, **10** (2019) 199.
[13] Lee K. H. *et al.*, *Atmos. Res.*, **143** (2014) 31.
[14] Kox S. *et al.*, in *Joint 2013 EUMETSAT Meteorological Satellite Conference and 19th Satellite Meteorology, Oceanography and Climatology Conference of the American Meteorological Society*, `https://www-cdn.eumetsat.int/files/2020-04/pdf_conf_p_s11_06_kox_v.pdf` (2013).

[15] Gorelick N. *et al.*, *Remote Sens. Environ.*, **202** (2017) 18.
[16] Schmetz J. *et al.*, *Bull. Am. Meteorol. Soc.*, **83** (2002) 977.
[17] de Laat A. *et al.*, *Nat. Hazards Earth Syst. Sci.*, **20** (2020) 1203.
[18] EUMETRAIN, https://www-cdn.eumetsat.int/files/2020-04/pdf_rgb_quick_guide_ash.pdf.
[19] Ellrod G. P. *et al.*, *J. Geophys. Res.*, **108** (2003) 4356.
[20] Amato E. *et al.*, *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (IEEE) 2021, p. 1, https://doi.org/10.1109/ICECCME52200.2021.9591110.
[21] Amato E. *et al.*, *Earth Space Sci. Open Arch.* (2021). https://doi.org/10.1002/essoar.10509929.1.
[22] Corradino C.*et al.*, *Remote Sens.*, **13** (2021) 4080.
[23] Corradino C. *et al.*, *Earth Space Sci. Open Arch.* (2021). https://doi.org/10.1002/essoar.10510119.1.
[24] Torrisi F. *et al.*, *Earth Space Sci. Open Arch.* (2021). https://doi.org/10.1002/essoar.10509947.1.
[25] MacQueen J., Le Cam L. and Neyman J. (Editors), *Proceedings 5th Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press) 1967, p. 281.
[26] Suykens J. A. K. *et al.*, *Neural Process. Lett.*, **9** (1999) 293.
[27] Corradino C. *et al.*, *Remote Sens.*, **11** (2019) 1916.
[28] Corradino C. *et al.*, *Energies*, **14** (2021) 197.
[29] Corradino C. *et al.*, 17^{th} *International Workshop on Cellular Nanoscale Networks and their Applications (CNNA)* (IEEE) 2021, https://doi.org/10.1109/CNNA49188.2021.9610813.