IL NUOVO CIMENTO **45 C** (2022) 80
DOI 10.1393/ncc/i2022-22080-1

COMMUNICATIONS: SIF Congress 2021

Machine learning and best fit approach to map lava flows from space

E. AMATO([1])([2])([*])

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

received 31 January 2022

Summary. — Estimating the areal coverage of newly erupted lava is both a crucial component of volcano monitoring and a powerful tool for characterizing lava flow emplacement behavior. Here, it is presented a methodology based on machine learning, developed in the Google Earth Engine platform, and best fitting approach, to map the extent of lava flows solely using freely available and open-source data from space-borne instruments. Radar and optical satellite data are used as input to different machine learning techniques and best fitting models, so that the methodology is able to operate in all weather conditions. The satellite-driven approach has been successfully used for mapping lava flows automatically during the long sequence of summit eruptions occurred at Mt. Etna between December 2020 and October 2021.

1. – Introduction

The recent eruptive activity of the Mount Etna (Sicily, Italy) was dominated by a number of lava fountain events, mostly occurring from the South-East Crater (SEC), located in the summit area. The last explosive phase started on 13 December 2020 and terminated on 23 October 2021, giving rise to a long sequence of paroxysmal lava fountain episodes of short duration (a few hours) and high intensity. These lava fountains fed several-km high eruptive columns and ash plumes, as well as short-lived lava overflows from the South-East Crater rim. Typically, these lava flows represent the greatest hazards to the tourist facilities on the upper flanks of Mt. Etna [1, 2]. Therefore, the timely mapping of the areas inundated by active lava flows is of obvious interest to hazard management during a volcanic eruption. However, estimating key lava flow parameters (*i.e.*, locations, intensities, extents, and advancing of lava flows) is not easy with traditional field methods because of the large size and poor accessibility of the inundated areas. Nowadays, the growing availability of satellite remote sensing (RS) data offers a means of gathering this information thanks to a number of sensors operating in the visible, infrared, thermal, and microwave regions of the electromagnetic spectrum that allow a detailed and timely view of active lava flow fields [3-11]. In addition, the free access to

([*]) E-mail: eleonora.amato@ingv.it, eleonora.amato01@unipa.it

a great amount of optical and radar satellite images with high spatial resolution, means it is now feasible to use machine learning approaches by taking into account different satellite data to obtain reliable lava flow maps [3, 5, 7, 12].

In the last few years, the use of machine learning (ML) techniques is increasing because of the ability to automatically process a great amount of data in a short time, avoiding the setting of a manual threshold and using the specific spectral response of the pixels of the images for the classification of new, unknown data [4,13-15]. Machine learning algorithms, usable also for volcanic applications [16], can involve either supervised or unsupervised learning. When using an unsupervised algorithm, only the number of clusters or classes for the classification and segmentation is chosen. When using a supervised algorithm, a training set needs to be prepared, representative of the features under investigation, to allow the algorithm generalizing for new data. Using high spatial resolution satellite images as input, unsupervised techniques have been successfully applied to map automatically lava flows in cases in which the lava is visible, clear and totally hot; otherwise, when the lava is not completely visible or it is composed of hot and cooling portions, it is preferable to use supervised techniques to capture all the features of the lava flow field [7, 8]. In cases in which high spatial resolution images are not available during the eruption, due to missing satellite problems or adverse weather conditions, the ML techniques can be usefully integrated with analytical methods [7], as well as best fitting approaches, to circumvent these problems. In particular, best fitting approaches have proven to be effective in recovering the missing data exploiting the trend of high temporal resolution satellite data in determining the extent of lava fields [17].

Here, it is presented a satellite-driven methodology developed within the Laboratory of Technologies for Volcanology (TechnoLab) at INGV in Catania that uses machine learning and best fitting approaches to detect and accurately map the areal extent of active lava flows. The methodology is able to operate in all weather conditions and allows obtaining the areal coverage of lava flows, combining satellite data, open-access platforms, and machine learning and best fitting approaches. The data input of the ML algorithms is related to different high spatial resolution missions (ranging from 10 to 100 m), in particular the images acquired by the Synthetic Aperture Radar (SAR) sensor on board of Sentinel-1 (from ESA), the MultiSpectral Instrument (MSI) on Sentinel-2 (from ESA), the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) on Landsat-8 (from NASA & USGS), and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor on EOS AM-1 Terra (from NASA). Finally, the high temporal resolution images (with a refresh frequency of 15 min and low spatial resolution of 4.5 km at Mediterranean latitudes) collected from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) instrument on board of the geostationary Meteosat Second Generation (MSG) satellite (from ESA EUMETSAT) are used as reference data for the fitting model.

The satellite-driven methodology for mapping lava flows has been used for the summit eruptions occurred at Mt. Etna between December 2020 and October 2021. These eruptive events have been chosen as case studies because all eruptions were well documented by satellite measurements that have been extremely useful for monitoring the eruptive activity. All high spatial resolution satellite data have been retrieved and analyzed in the Google Earth Engine (GEE) cloud-based platform [18]. Despite the need for a great memory to store the data and for the power to process themselves, the GEE platform allows to view satellite data remotely and process them with various techniques, without overloading the devices in use. Moreover, the volcanic radiative power (VRP) values [3], derived from SEVIRI images for the best fitting approach, were downloaded

from LAV@HAZARD, a web-GIS platform for volcanic hazard monitoring developed by the TechnoLab [19].

2. – Materials

2˙1. *2020–2021 Etna paroxysmal events.* – Between 13 December 2020 and 23 October 2021, a long sequence of 54 paroxysmal lava fountain events occurred at Mount Etna volcano (Sicily, Italy) (fig. 1). All the lava fountains occurred from the South-East Crater (SEC) located in the summit area. The paroxysms were generally characterized by a starting phase with a gradual increase of the explosive activity and a subsequent paroxysmal phase, in which there was an intensification of explosions with the formation of an eruptive column feeding a sustained ash plume, rising up to several kilometers above the crater. All the lava fountains were also associated with the formation of short-lived lava overflows from the South-East Crater rim. In the table in fig. 3, the lava-forming eruptions from December 2020 to October 2021 are listed as identified by the INGV weekly bulletins. The volcanic activity of Etna is continuously monitored by the INGV, which publishes weekly bulletins at `www.ct.ingv.it`.

2˙2. *Data source.* – Sentinel-1 satellite mission is part of the Copernicus programme of the European Space Agency (ESA). The Sentinel-1 constellation consists of two identical satellites both equipped with C-band Synthetic Aperture Radar (SAR) instruments (frequency 5.405 GHz, wavelength ∼5.55 cm), providing data in dual or single polarizations. Sentinel-1 data are acquired in different spatial resolutions (from 10 m to 40 m) depending on the operational mode. Sentinel-1 images acquired in C-band Wide swath (IW) mode, with a temporal resolution of 12 days, are made available by Google Earth Engine, with Ground Range Detected (GRD) scenes, with thermal noise removal, radiometric calibration and with final terrain-corrected values converted to decibels via log scaling ($10 \log_{10}(x)$).

Fig. 1. – Location of the Mt. Etna volcano (white boxes), in Sicily, Italy, and the relative summit area (red boxes), with four summit craters, the Voragine (VOR), the Bocca Nuova (BN1, BN2), the North-East Crater (NEC), and the South-East Crater (SEC), and a valley in the east part of the volcano itself (Valle del Bove), where the lava frequently flows.

Sentinel-2 satellite mission is also part of the Copernicus programme of the ESA and consists of a constellation of two identical satellites. The revisit frequency of one satellite is 10 days, resulting in a revisit frequency of 5 days for the constellation and 2-3 days at mid-latitudes. Both satellites are equipped with a MultiSpectral Instrument (MSI) with 13 bands in the visible, near infrared, and shortwave infrared part of the spectrum, and with "Aerosols", "Water vapor", "Cirrus" bands, at spatial resolutions of 10, 20 and 60 m. Sentinel-2 images are made available as surface reflectance and top-of-atmosphere (TOA) reflectance scaled by 10000 by Google Earth Engine.

Landsat-8 is a satellite launched in the frame of a joint National Aeronautics and Space Administration (NASA) and United States Geological Survey (USGS) mission. It carries two sensors, namely the Operational Land Imager (OLI) characterized by 9 spectral bands, in the visible, near infrared and shortwave infrared, and a "Coastal aerosol" band (spatial resolution of 30 m), and with a Panchromatic band and a "Cirrus" band at a spatial resolution of 15 m; the second sensor is the Thermal Infrared Sensor (TIRS) with 2 bands of 100 m spatial resolution, resampled to 30 m on Google Earth Engine. Landsat-8 images, acquired at 16-day intervals, are made available as atmospherically corrected surface reflectance and calibrated top-of-atmosphere (TOA) reflectance from the OLI and TIRS sensors by Google Earth Engine.

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is a high spatial, spectral and radiometric resolution multispectral sensor on board of NASA EOS AM-1 Terra satellite. The recurrent cycle is 16 days. Data are acquired in 14 spectral bands, including visible and near infrared radiation (VNIR), shortwave infrared radiation (SWIR) and thermal infrared radiation (TIR). All the ASTER images, with spatial resolution ranging from 15 to 90 m, with calibrated at-sensor radiance, orthorectified and terrain corrected, are made available by Google Earth Engine.

The Spinning Enhanced Visible and Infrared Imager (SEVIRI) is on board the geostationary Meteosat Second Generation (MSG) Satellites (from ESA EUMETSAT). SEVIRI has the capability to measure radiation in 12 spectral channels covering from the visible (VIS) to the infrared (IR) part of the spectrum. In general, the spatial resolution of SEVIRI depends on distance from the sub-satellite ground point (thus latitude, longitude) and on viewing zenith angle. Moreover, the spatial resolution is down to 1 km for the High-Resolution Visible channel and of 3 km at sub-satellite point for standard channels. The Volcanic Radiative Power (VRP) values calculated at Etna volcano using SEVIRI images are made available via the web-GIS LAV@HAZARD platform.

3. – Methods

The satellite-driven methodology for lava flow mapping in near real-time involves machine learning and best fitting techniques to process satellite data. Figure 2 shows the flow chart of the methodology used.

3˙1. *Machine learning techniques.* – Satellite data can become inputs to several different techniques of machine learning in order to segment lava flow images using a *pixel-based* and *scene-by-scene* approach. The pixel-based approach is founded on the classification of each pixel of the image under analysis one after the other, deciding which class each pixel belongs to, exploiting only the information due to the differences in the spectral response of an object (already contained in the single pixel) [3, 4]. The scene-by-scene classification approach is based on the use of a single image to apply the model, using a portion of this to train the model (drawing manually polygons that enclose the train pixels, emphasizing specific features of the classes object of the study), making the model

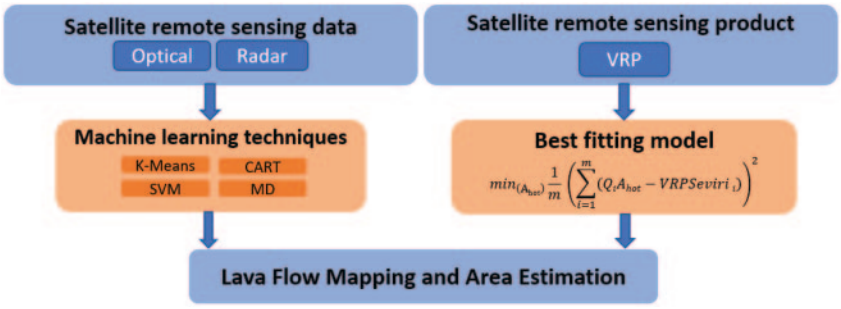

Fig. 2. – Flow chart of the methodology.

capable of classifying all the rest of the pixels of the image, or clustering directly the entire image [5, 7, 20].

In this study, different pixel-based and scene-by-scene ML techniques have been used, the unsupervised K-Means and the supervised Support Vector Machine (SVM), Classification and Regression Trees (CART) and Minimum Distance (MD), all implemented in GEE [7]. The unsupervised learning has been used in presence of totally hot and clear lava, while the supervised learning in presence of lava flow with hot and cooling parts, since a more advanced algorithm is necessary. Thus, the supervised techniques allow choosing the training data in such a way that the specific spectral characteristics of the hot and cooling points are taken into account and used for the analysis [7].

Firstly, a Normalized Index, defined by $NI(x, y) = \frac{x-y}{x+y}$, with x and y bands (*e.g.*, [3]), is applied to the image under analysis, in order to emphasize the peculiar characteristics of the bands used for the index. Otherwise, a normalization of the bands in the same range is done, before using them. This first step is necessary when data with different ranges of values are used and is named *feature scaling* [14, 21]. This step is requested to make data homogeneous, avoiding that the data with a range of values bigger than other ones can superimpose their information content with respect to the other data. In the cases in which the NI index is used, feature scaling has not been performed, because the NI is a normalized index. Otherwise, when no index has been calculated, a feature scaling is used. Subsequently, a specific ML technique is used for lava and background classification and segmentation, exploiting the different spectral response between the pixels under analysis, and obtaining a map of the lava flow. Finally, for each case under analysis, the total area of the lava flow obtained is calculated by summing up all the areas of the pixels classified as lava by the classifier (firstly computing in GEE the area of one pixel of the image and subsequently considering all the lava pixels) [7, 22]. Thus, all the pixels are classified, the lava flow is segmented respect to the background and the area of the lava flow is estimated.

3˙2. *Best fitting model.* – The Volcanic Radiative Power (VRP) values derived from the radiometric measurements of the SEVIRI sensor are used as input to the best fitting model. The original SEVIRI data are downloaded from the EUMETSAT system, which provides them with a latency of around 3 h. The original spectral radiance data recorded by SEVIRI, in the Middle Infrared (MIR) at 3.9 μm and Thermal Infrared (TIR) at 12.0 μm, are processed by LAV@HAZARD in order to identify the pixels containing thermal anomalies associated with the eruptive activity at Mt. Etna [19, 23]. Once the hot pixels have been identified (hotspot detection), the volcanic radiative power (VRP) is calculated using the MIR method [24].

Firstly, in the MIR atmospheric window (3.4–4.2 μm) and at the temperatures of the active fires (∼600–1500 K), the Plank's radiation law $B(\lambda, T)$, that models the relationship between emitted spectral radiance and emitted temperature (T), is approximable with a fourth-order power law. Thus, $L_{MIR,hot}$, that is the MIR emitted spectral radiance of a fire, can be written as (see [24-27])

$$L_{MIR,hot} = \epsilon_{MIR} B(\lambda_{MIR}, T) \sim \epsilon_{MIR} a \sum_i p_i T_i^4, \tag{1}$$

where p_i is the portion of the pixel at temperature T_i (in K), ϵ_{MIR} is the surface spectral emissivity in the MIR band, a is the empirical constant (in W m^{-2}sr^{-1} μm^{-1} K^{-4}) depending on the specific sensor employed. This equation is defined like the Stefan-Boltzmann's law.

Using this last eq. (1), it is possible to formulate the eq. (2),

$$VRP_{true} = \epsilon \sigma A_{pixel} \sum_i p_i T_i^4 \tag{2}$$

in watt, where A_{pixel} is the pixel size (4.5 10^6 m^2 for the resampled SEVIRI pixels), ϵ is the emissivity and σ is the Stefan-Boltzmann constant (5.67 10^{-8} W m^{-2} K^{-4}). The VRP values calculated by the LAV@HAZARD platform is obtained combining eq. (1) and eq. (2), resulting in

$$VRP = \frac{\epsilon \sigma A_{pixel} L_{MIR,hot}}{a \epsilon_{MIR}}. \tag{3}$$

Thus, it is possible to consider the physical formulation (from eq. (2)) $VRP = \epsilon \sigma A_{hot} T_{hot}^4$, with A_{hot} the effect of the hotspot area and T_{hot} its integrated temperature, and it is possible to analyse also the background contribution, that should be subtracted for real remote sensing data. In this way, finally, $VRP = QA$, with $Q = \epsilon \sigma (T_{surf}^4 - T_{background}^4)$, where T_{surf} is computed with the Stefan Cooling Problem methodology [3, 17], can be written and, using the VRP_{Seviri} from LAV@HAZARD, the best fitting algorithm is defined like in eq. (4),

$$\min_{A_{hot}} \frac{1}{m} \left(\sum_{i=1}^{m} (Q_i A_{hot} - VRP_{Seviri_i}) \right)^2, \tag{4}$$

minimizing the function to obtain the best area A_{hot} for the data used like target or reference data (VRP_{Seviri}), with m the number of data in the final vector to minimize. The VRP_{Seviri} data used for the best fitting approach are retrieved by the cooling curves associated with the paroxysmal events, *i.e.*, the data between the pick value of VRP during the eruption and the values of VRP near to the zero at the final phase of the event [17].

4. – Results

All the 54 lava fountain events occurred at Mt. Etna between December 2020 and October 2021 were analyzed. The high spatial resolution satellite images were obtained and processed in GEE with automatic ML models. When these satellite images are

not available, best fitting models were used with VRP data. Different combinations of images are used for the ML models, due to the relative availability, *i.e.*, Sentinel-2 MSI, Landsat-8 OLI/TIRS, and Terra ASTER images, to develop a near real-time analysis. In addition, Sentinel-1 SAR images are fused together with the optical images, in case of adverse visibility conditions, and used for the ML models. In particular, two SAR images are used, the first one prior to the event ("Pre" image), the second one after the event ("Post" image), making a difference of them in order to emphasize the changes in the scene before and after the event [28]. It was also used the ratio instead of the difference between SAR images, for example when more events were analyzed in a cumulative way and the ratio emphasized the changes in the scene and reduced the speckle noise, typical of the SAR images [6, 29, 30]. In particular, the lava flow overflows occurred between 19 and 27 June 2021 were analyzed cumulatively because the adverse weather conditions made distinguishing the different lava flow fields with a satellite image for each event impossible. The total areal coverage of the lava flows of this June period was estimated using the ratio between two SAR images, fused with a Sentinel-2 image from 23 June 2021 (with the NI index for the bands used), an ASTER image from 25 June, and a Landsat-8 image of the same date (with the NI index for the bands used). The bands of the image obtained have been normalized in the same range and GEE filters were used to improve the performances of the segmentation.

The table in fig. 3 summarizes all the satellite images used as input to the ML techniques (*i.e.*, K-Means, SVM, CART, MD) and the results obtained for each event both using ML analysis and best fitting method. Figures 4 and 5 show the total areal coverage of the lava flows erupted at Mt. Etna volcano from December 2020 to October 2021. Figure 4 shows the maps obtained for each lava flow with the ML techniques, between December 2020 and October 2021, except for the events occurred on 19 February, the period between 04 March and 01 April, 25 May, 28 May, 04 June and 28 June 2021 events. Figure 5 shows all the lava flows maps obtained with the ML techniques overlapped. The different colors indicate the number of lava flows overlapped, highlighting the regions more inundated by the lava flows, *i.e.*, a grayscale from black (equal to value 0, showing no lava in the pixel) to white (equal to value 26, showing 26 lava flows overlapped in that pixel).

It is worth mentioning that for the cases in which the estimate of the areal coverage involves multiple lava fountains very close in time (*e.g.*, for 22 December 2020, 21 and 22–23 May 2021, 25 May 2021 and 28 May 2021 events), only the cooling curve and the VRP SEVIRI data related to the last fountain are used as reference for the best fitting model.

5. – Discussion

Following the manifestation of an effusive eruption once it has started has immediate applications to the real-time monitoring of lava flow hazards. A satellite-driven strategy could represent an operational monitoring system that should give during an eruption the current state of the effusive activity and the spatial-temporal evolution of the lava flow field. The results of the mapping of lava flows during the 2020–2021 Mt. Etna eruptions highlight different situations. In cases in which the lava is completely hot, the four different ML techniques, *i.e.*, k-means, SVM, MD, CART, are interchangeable and the unsupervised learning is enough for the segmentation. In contrast, in cases of lava flows with hot and cooling parts, only the supervised techniques are capable of providing a robust mapping of lava flows. When the bad weather

N.	Event	Satellites	Algorithm	Area (x10^6 m^2)	N.	Event	Satellites	Algorithm	Area (x10^6 m^2)
1	14 December 2020	Sentinel-1 SAR (pre, post), Sentinel-2 MSI, Landsat-8 OLI/TIRS	SVM	0.47	28	30 May 2021	Landsat-8 OLI/TIRS	MD	0.11
2	22 December 2020	Sentinel-1 SAR (pre, post), Sentinel-2 MSI	K-MEANS	0.87	29	02 June 2021	Terra ASTER	K-MEANS	0.60
3	17 January 2021	Sentinel-2 MSI	MD	0.05	30	04 June 2021	MSG-SEVIRI	Best fit	0.25
4	18 January 2021	Sentinel-1 SAR (pre, post), Sentinel-2 MSI	K-MEANS	0.47	31	12 June 2021	Sentinel-2 MSI	MD	0.35
5	16 February 2021	Terra ASTER	MD	1.45	32	14 June 2021	Terra ASTER	MD	0.30
6	18 February 2021	Sentinel-2 MSI	K-MEANS	1.80	33	16 June 2021	Terra ASTER	CART	0.36
7	19 February 2021	MSG-SEVIRI	Best fit	1.03	34	17 June 2021	Sentinel-2 MSI	K-MEANS	0.27
8	21 February 2021	Sentinel-2 MSI	K-MEANS	1.61	35	19 June 2021	Sentinel-1 SAR (pre, post), Sentinel-2 MSI, Landsat-8 OLI/TIRS, Terra ASTER	SVM (cumulative approach)	1.65 (cumulative)
9	23 February 2021	Sentinel-2 MSI	K-MEANS	1.63	36	20 June 2021			
10	24 February 2021	Sentinel-2 MSI	MD	1.01	37	22 June 2021			
11	28 February 2021	Sentinel-2 MSI	MD	0.83	38	23 June 2021			
12	02 March 2021	Sentinel-2 MSI	K-MEANS	1.23	39	24 June 2021			
13	04 March 2021	MSG-SEVIRI	Best fit	1.13	40	25 June 2021			
14	07 March 2021	MSG-SEVIRI	Best fit	1.07	41	26 June 2021			
15	09 March 2021	MSG-SEVIRI	Best fit	0.58	42	27 June 2021			
16	12 March 2021	MSG-SEVIRI	Best fit	1.25	43	28 June 2021	MSG-SEVIRI	Best fit	0.44
17	15 March 2021	MSG-SEVIRI	Best fit	0.31	44	01 July 2021	Landsat-8 OLI/TIRS	MD	0.48
18	17 March 2021	MSG-SEVIRI	Best fit	0.48	45	04 July 2021	Sentinel-2 MSI	MD	0.49
19	19 March 2021	MSG-SEVIRI	Best fit	0.61	46	06 July 2021	Sentinel-2 MSI	K-MEANS	0.29
20	24 March 2021	MSG-SEVIRI	Best fit	0.86	47	08 July 2021	Terra ASTER	SVM	0.49
21	31 March - 01 April 2021	MSG-SEVIRI	Best fit	1.11	48	14 July 2021	Sentinel-2 MSI	MD	0.52
22	19 May 2021	Sentinel-2 MSI	K-MEANS	0.39	49	20 July 2021	Terra ASTER	K-MEANS	0.69
23	21 May 2021	Terra ASTER	SVM	0.14	50	31 July 2021	Terra ASTER	CART	0.63
24	22-23 May 2021				51	08 August 2021	Sentinel-2 MSI	K-MEANS	0.38
25	24 May 2021	Landsat-8 OLI/TIRS	K-MEANS	0.13	52	29 August 2021	Sentinel-2 MSI	K-MEANS	0.90
26	25 May 2021	MSG-SEVIRI	Best fit	0.23	53	21 September 2021	Sentinel-2 MSI	CART	0.48
27	27-28 May 2021	MSG-SEVIRI	Best fit	0.20	54	23 October 2021	Landsat-8 OLI/TIRS	CART	0.46

Fig. 3. – List of the 2020–2021 paroxysmal summit events of Mt. Etna (Italy), of the satellite used for obtaining the images, of the algorithm chosen to obtain the areal extension and the relative results (rounded to the second decimal digit).

conditions do not allow detecting the whole lava flow field with a single image, different images from different sensors are fused in a single image and processed using the ML techniques. Typically, in these cases the difference or the ratio between SAR images combined with optical images are used. Finally, when more events are not captured by any high spatial resolution images, the areal coverage of lava flows of the entire period is obtained considering a cumulative area, for example for the events that occurred in June 2021, or, for each event, applying the best fitting approach.

Fig. 4. – Top: events of 14 December 2020 (a), 22 December 2020 (b), 17 January 2021 (c), 18 January 2021 (d), 16 February 2021 (e), 18 February 2021 (f), 21 February 2021 (g), 23 February 2021 (h), 24 February 2021 (i), 28 February 2021 (l), 02 March 2021 (m). Middle: events of 19 May 2021 (a), 21 and 22–23 May 2021 (b), 24 May 2021 (c), 30 May 2021 (d), 02 June 2021 (e), 12 June 2021 (f), 14 June 2021 (g), 16 June 2021 (h), 17 June 2021 (i), 19–27 June 2021 (l). Bottom: events of 01 July 2021 (a), 04 July 2021 (b), 06 July 2021 (c), 08 July 2021 (d), 14 July 2021 (e), 20 July 2021 (f), 31 July 2021 (g), 08 August 2021 (h), 29 August 2021 (i), 21 September 2021 (l), 23 October 2021 (m).

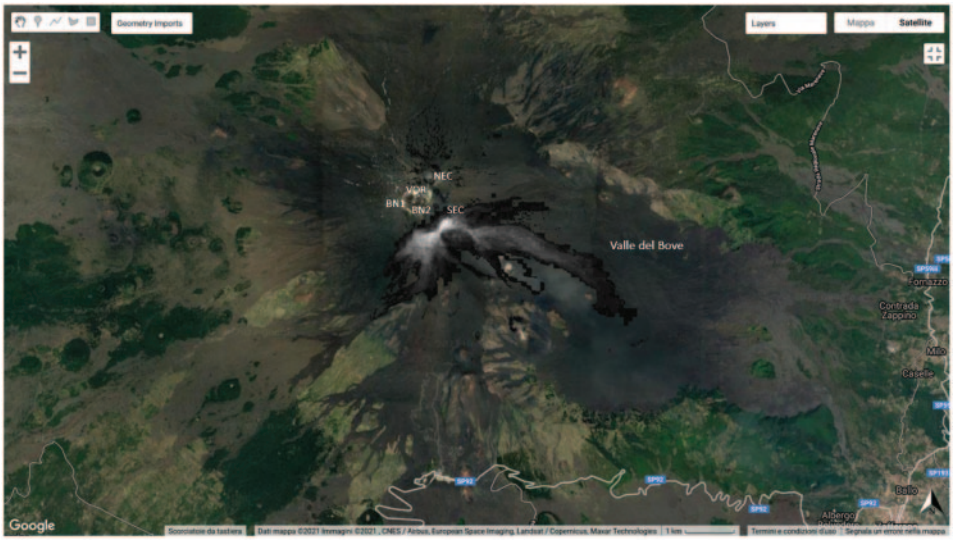

Fig. 5. – All the events overlapped (between December 2020 and October 2021, except for the 19 February, the period between 04 March and 01 April, 25 May, 28 May, 04 June and 28 June 2021 events), with a gradation of colors (black = 0, white = 26) to indicate the number of overlapped lava flows.

The best fitting approach does not provide a map of the lava flow under examination, it only gives an estimation of the lava flow extension to obtain the measured VRP, used as reference for that given event. The best fitting approach has been used for the eruptions occurred on 19 February, in the period between 04 March and 01 April, on 25 May, 28 May, 04 June and on 28 June 2021. In contrast, the ML techniques return a map of the lava flows and the relative areal extensions of these maps. In general, the flow maps produced from machine learning approach correspond well with the actual lava flow fields of eruptive events discriminating between pixels belonging to the lava flow and pixels belonging to the background, exploiting all the available satellite data. Possible uncertainties on estimations of lava area values were expected during the events occurred between 15 and 19 March 2021, in the second half of June 2021 and on 23 October 2021, when the weather was particularly cloudy. The overestimation was also expected in cases in which the best fitting model was used. Nevertheless, all the estimated values shown in the table in fig. 3 agree well with the values published for the relative events in the INGV weekly bulletins at `www.ct.ingv.it`.

The period under analysis was characterized by an intense volcanic activity, with overlaps of the flows, and this is confirmable seeing the overlapped lava flows in fig. 5. This figure shows all the maps obtained with ML techniques overlapped. It is possible to notice that some lava flows inundate the same location of the other ones, with a number of lava flows overlapped, considering a specific region (pixel), from a minimum of 0 to a maximum of 26 lava flows. If you compute the area of all the events overlapped, without considering the thickness of each event and the region with more lava flows, the result will be $5.40\,\mathrm{km}^2$. In contrast, if you sum all the areas for the same events, considering the results obtained with the ML methods, the area will be $21.53\,\mathrm{km}^2$, confirming the idea that more events are overlapped.

6. – Conclusion

The ability to accurately quantify the lava areal extent and map lava flow fields at high spatial and temporal resolution has direct relevance to volcanic hazards once an eruption has begun. It was found that the maps of lava flows generated by ML techniques and the estimations of lava areal coverage calculated by best fitting models are valuable tools for investigating lava flow behavior. The joint analysis of radar and optical satellite data has provided a superb means of documenting lava flow activity at Etna volcano during the 2020–2021 Mt. Etna summit eruptions. The availability of a variety of multispectral satellite data for the same eruptive event has given the possibility to use the machine learning techniques and statistical and analytical methods for the classification and segmentation of the hot and cooling lava flows in all weather conditions. Different ML approaches, unsupervised and supervised, have been applied to analyze a great amount of high spatial resolution multispectral data without any human intervention. The flow maps produced from machine learning algorithms correspond well with the actual lava flow fields. In cases of lack of data for missing satellite problems or adverse weather conditions, the areal coverage of lava flows was estimated using the VRP values derived from the geostationary SEVIRI sensor as input to a best fitting technique. Finally, the satellite-based techniques for mapping lava flows like those presented here for Etna can be used to generate, improve, or update existing lava flow maps for many active volcanoes worldwide. In addition, a satellite-derived map can tremendously reduce the time and effort needed for traditional fieldwork by detecting the location and boundaries of newly erupted lavas.

* * *

This work was developed within the framework of the Laboratory of Technologies for Volcanology (TechnoLab) at the INGV in Catania (Italy). Sentinel-1, Sentinel-2, Landsat-8, EOS AM-1 Terra and EUMETSAT MSG data used in this work are provided by the European Space Agency (ESA), the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS), available through the Google Earth Engine platform and the LAV@HAZARD platform.

REFERENCES

[1] Del Negro C. *et al.*, *Sci. Rep.*, **3** (2013) 1.
[2] Del Negro C. *et al.*, *Geol. Soc. Am. Bull.*, **132** (2020) 1615.
[3] Harris A., *Thermal Remote Sensing of Active Volcanoes* (Cambridge University Press) 2013, https://doi.org/10.1017/CBO9781139029346.
[4] Corradino C. *et al.*, *Remote Sens.*, **11** (2019) 1916.
[5] Corradino C. *et al.*, *Energies*, **14** (2021) 197.
[6] Corradino C. *et al.*, *Remote Sens.*, **13** (2021) 4080.
[7] Amato E. *et al.*, in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (IEEE) 2021, p. 1, https://doi.org/10.1109/ICECCME52200.2021.9591110.
[8] Corradino C. *et al.*, in *2021 17th International Workshop on Cellular Nanoscale Networks and their Applications (CNNA)* (IEEE) 2021, p. 1, https://doi.org/10.1109/CNNA49188.2021.9610813.
[9] Amato E. *et al.*, *Earth and Space Science Open Archive* (2021), https://doi.org/10.1002/essoar.10509929.1.

[10] Torrisi F. *et al.*, *Earth and Space Science Open Archive* (2021), https://doi.org/10.1002/essoar.10509947.1.
[11] Corradino C. *et al.*, *Earth and Space Science Open Archive* (2021), https://doi.org/10.1002/essoar.10510119.1.
[12] Lu Z. *et al.*, *Remote Sens. Environ.*, **91** (2004) 345.
[13] Freeman E. A. *et al.*, *Ecol. Model.*, **217** (2008) 48.
[14] Bonaccorso G., *Machine Learning Algorithms* (Packt Publishing Ltd.) 2017.
[15] Rogic N. *et al.*, *Remote Sens.*, **11** (2019) 662.
[16] Corradino C. *et al.*, *Remote Sens.*, **12** (2020) 970.
[17] Ganci G. *et al.*, *Geophys. Res. Lett.*, **39** (2012) L06305.
[18] Gorelick N. *et al.*, *Remote Sens. Environ.*, **202** (2017) 18.
[19] Vicari A. *et al.*, *Ann. Geophys.*, (2011) https://doi.org/10.4401/ag-5347.
[20] Zerrouki N. *et al.*, in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE) 2014, p. 864, https://doi.org/10.1109/SMC.2014.6974020.
[21] Géron A., *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (O'Reilly Media, Inc.) 2019.
[22] Marchese F. *et al.*, *Remote Sens.*, **13** (2021) 3074.
[23] Corradino C. *et al.*, *Energies*, **12** (2019) 1216.
[24] Wooster M. J. *et al.*, *Remote Sens. Environ.*, **86** (2003) 83.
[25] Wooster M. J. *et al.*, *J. Geophys. Res.: Atmos.*, **110** (2005) D24311.
[26] Zhang T. *et al.*, *Remote Sens. Environ.*, **198** (2017) 407.
[27] Xu W. *et al.*, *Remote Sens. Environ.*, **261** (2021) 112460.
[28] Wadge G. *et al.*, *J. Volcanol. Geotherm. Res.*, **199** (2011) 142.
[29] Wadge G. *et al.*, *Geol. Soc. London Mem.*, **21** (2002) 583.
[30] Arnold D. *et al.*, *Remote Sens. Environ.*, **209** (2018) 480.