Earth and Space Science

RESEARCH ARTICLE
10.1029/2021EA001936

Twenty Years Monitoring of Soil CO_2 Flux and Seismicity at Cava dei Selci Gas Discharge (Colli Albani Volcano, Italy)

Luca Tarchini[1], Maria Luisa Carapezza[1], Domenico Granieri[2], Alberto Frepoli[1], Nicola Mauro Pagliuca[1], and Massimo Ranaldi[1]

[1]INGV—Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy, [2]INGV—Istituto Nazionale di Geofisica e Vulcanologia, Pisa, Italy

Key Points:

- Cava dei Selci is the main degassing site of Colli Albani volcano
- Monitoring of soil gas discharge was conducted since 2000
- Increase of the CO_2 output seems related to extensional deep earthquakes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. L. Carapezza,
marialuisa.carapezza@ingv.it

Citation:

Tarchini, L., Carapezza, M. L., Granieri, D., Frepoli, A., Nicola Mauro, P., & Ranaldi, M. (2022). Twenty years monitoring of soil CO_2 flux and seismicity at Cava dei Selci gas discharge (Colli Albani volcano, Italy). *Earth and Space Science*, *9*, e2021EA001936. https://doi.org/10.1029/2021EA001936

Received 5 DEC 2021
Accepted 18 APR 2022

Author Contributions:

Conceptualization: Luca Tarchini, Maria Luisa Carapezza, Domenico Granieri
Data curation: Luca Tarchini, Maria Luisa Carapezza, Domenico Granieri, Alberto Frepoli, Nicola Mauro Pagliuca
Formal analysis: Luca Tarchini, Domenico Granieri, Alberto Frepoli, Nicola Mauro Pagliuca, Massimo Ranaldi
Funding acquisition: Maria Luisa Carapezza
Investigation: Luca Tarchini, Maria Luisa Carapezza, Domenico Granieri, Alberto Frepoli, Nicola Mauro Pagliuca, Massimo Ranaldi

Abstract Cava dei Selci (CdS) is the main degassing site of the Colli Albani quiescent volcano and since 20 years it is the site of geochemical volcano monitoring. Emitted gas consists mostly of CO_2 (≥98 vol.%) with minor H_2S, and helium isotopes suggest it has a significant magmatic component. The diffuse soil CO_2 flux was monitored in the period 2000–2020, with 55 surveys on a target area. The total CO_2 output fluctuates from 5.6 to 24.8 t d^{-1}. The soil CO_2 flux per unit surface (average 2.323 kg m^{-2} d^{-1}) is the highest of 15 Italian actively degassing volcanic and geothermal areas. Soil CO_2 flux and environmental parameter data collected over 4-year of continuous monitoring (2004–2008) were analyzed by stochastic Gradient Boosting Trees regression (sGBT), Multiple Linear Regression, and Principal Component Regression. Only sGBT predicts the entire data set and effectively identifies the relationship between soil CO_2 flux and environmental parameters. Residuals indicate two anomalous degassing periods (March-2005, summer-2007). Colli Albani area is affected by moderate seismicity ($Md \leq 4$). 575 earthquakes occurring from 2009 to 2021 were analyzed determining their location, hypocenter depth, and focal mechanism (of 43 selected events). Evaluation of seismic events occurred across geochemical surveys within 30 km from CdS shows that there is a relationship between CO_2 flux, earthquake focal mechanism and depth: shallow strike-slip hypocenters are associated to low fluxes, deep normal-faulting hypocenters to high CO_2 output.

Plain Language Summary Cava dei Selci is a locality of the Colli Albani volcanic province characterized by continuous soil gas discharges, with an estimated total CO_2 output of 5.6–24.8 tonnes per day. Anomalously high levels of soil CO_2 were recorded in March 2005 and summer 2007. Monitoring of soil CO_2 flux has been conducted since 2000 for a period of 20 years, showing that degassing is principally controlled by NW-SE extensional faults, and is thus linked to the seismic activity in the region. Here we show a relation between soil CO_2 discharge and the depth of the foci of moderate magnitude extensional earthquakes recorded in the Colli Albani province in the period 1997–2021.

1. Introduction

Carbon dioxide (CO_2) is, after steam, the main gas released in active volcanic and geothermal areas. During magma ascent through the crust, the solubility of CO_2 in magmas decreases. This, combined with the low reactivity of CO_2 during high temperature water rock interaction processes, causes ready escape of this gas toward the surface (Gerlach & Graeber, 1985; Papale et al., 2006; Sigvaldason, 1989). Besides the convective emissions associated with fumaroles and mofettes, CO_2 is released also diffusively from the soil (Burton et al., 2013 and references therein). Therefore, CO_2 output is the geochemical parameter most widely used for volcanic surveillance. In Italy, starting from 1990 at Vulcano, and from 2000 at Stromboli, Etna, Campi Flegrei, Vesuvio, and Colli Albani, diffuse soil CO_2 flux surveys from target areas have been systematically carried out. At the same time, some automatic stations for the continuous monitoring of soil CO_2 flux have been installed (Badalamenti et al., 1994; Carapezza & Diliberto, 1994; Carapezza et al., 2009, 2012; Carapezza & Federico, 2000; De Gregorio & Camarda, 2016; Diliberto et al., 2002; Federico et al., 2008; Granieri et al., 2003, 2010, 2013; Liuzzo et al., 2013). Anomalies in soil CO_2 flux have been recorded before volcanic eruptions at Stromboli (Carapezza et al., 2004) and during unrest crises at Vulcano (Di Martino et al., 2020; Granieri et al., 2006), Campi Flegrei (Chiodini et al., 2010), and Santorini (Tarchini et al., 2019) confirming that CO_2 output is a promising parameter for volcanic surveillance. However, soil CO_2 flux alone is not sufficient to evaluate whether an unrest crisis will evolve or not into a volcanic eruption. The contemporaneous occurrence of both geophysical (seismic and geodetic) and geochemical (chemical and isotopic composition, temperature, and flow of fumarolic gas)

Methodology: Luca Tarchini, Domenico Granieri, Alberto Frepoli
Supervision: Maria Luisa Carapezza
Validation: Maria Luisa Carapezza
Writing – original draft: Luca Tarchini, Maria Luisa Carapezza, Domenico Granieri, Alberto Frepoli, Nicola Mauro Pagliuca
Writing – review & editing: Luca Tarchini, Maria Luisa Carapezza, Domenico Granieri, Alberto Frepoli

anomalies is needed to herald an impending eruption. It is thus useful to process and interpret together the time series of both soil CO_2 flux and seismic activity.

Colli Albani is a quiescent volcano located in central Italy (Lazio region), about 20 km to the south-east of Rome. It is characterized by the presence of several CO_2-dominated cold gas permanent discharges (Carapezza et al., 2019), the most important of which is Cava dei Selci (hereafter CdS) in Marino Municipality (Rome). The gas is dominated by CO_2, with ∼1 vol.% of hydrogen sulfide (H_2S), and its $^3He/^4He$ isotopic composition suggests the presence of a deep magmatic source (Carapezza & Tarchini, 2007).

At CdS, gas is released from both the soil and discrete vents. In the winter rainy season, a water pool frequently forms in the lowest part of CdS depression and dozens of cows and sheep have been killed by gas inhalation while drinking at the pool. In addition, a man lost his life near the gas discharge (Carapezza et al., 2003).

In the frame of Colli Albani volcano monitoring activity, the diffuse soil CO_2 flux of CdS gas discharge has been monitored in the last 20 years (Carapezza et al., 2003). In addition, from 2004 to 2008, soil CO_2 flux was also continuously recorded by an automatic station together with relevant ancillary environmental parameters (atmospheric pressure, air and soil temperature and humidity, wind direction and speed).

In this paper, we describe the results of 20-year monitoring of diffuse degassing from CdS, and analyze and discuss the variation of this parameter through time and in relation to local seismicity. Moreover, we compare and discuss common soil CO_2 flux data analysis techniques, to select the most appropriate for volcanic surveillance.

2. Geological Setting

The CdS gas discharge is located at the NW periphery of the Colli Albani composite volcano (Figure 1a). This is a potassic-alkaline volcanic complex belonging to the Roman comagmatic province (Washington, 1906), whose eruptive activity began ∼600 ka ago and continued up to at least 36 ka (Marra et al., 2016). The volcano is presently considered to be in a quiescent state (Marra et al., 2016), experiencing anomalous ground uplift and periodic seismic crises accompanied by deep gas discharges. It is thus in the list of Italian volcanic centers to keep under continuous surveillance because of a nonnegligible probability that it could erupt again (Carapezza, Barberi, et al., 2010; Trasatti et al., 2018).

The most recent eruptive vents of the complex are aligned NW-SE (Anzidei et al., 2008) along the main structural feature in the area (Acocella & Funiciello, 2006), cutting across the floor of the caldera which presently contains the Lake Albano. The main CdS gas discharge is located on this lineament (Figure 1a), which likely acts as a permeable fault system allowing deep gases to migrate and reach the surface (Carapezza et al., 2020).

Water floodings from the lake have repeatedly occurred up to Roman times (IV Century B.C.; Funiciello et al., 2010). The floods generated mud flows (lahars), which covered the north-western downslope side of the volcano, including the CdS area, with an impermeable surficial cover. In places where this impermeable cover has been removed by excavation, such as at CdS (the Italian word "cava" means quarry), the gas flows through to the surface. Where the impermeable mud cover is still in place, rising gas pressurizes shallow confined aquifers, which can produce gas blowouts when reached by drilling. This is the case of the Rome gas blowout zone (RBZ in Figure 1a) where several accidents occurred in 2003–2020 (Carapezza et al., 2020, 2021). As a consequence of shallow geognostic drillings (10–15 m depth), a gas blowout also occurred in 2008 (Carapezza, Ricci, et al., 2010) at a distance of less than 100 m from the CdS gas discharge (Figure 1b).

Gas of magmatic origin rising from depth, mixes with CO_2 produced by metamorphic decarbonation of the Mesozoic calcareous basement, which hosts the main regional aquifer of Colli Albani (Mazza & Capelli, 2010). It then escapes to the surface along leaky faults. Interpretation of gravity anomalies showed that the basement has a horst-graben structure (Cesi et al., 2008). Natural gas discharges of Lazio region are all located above the faults bordering structural highs of the basement (Carapezza et al., 2019), which include the NW-SE Ciampino high where CdS rests. Here, the top boundary of the carbonate basement rises at a depth of about 800 m from the surface (Giordano et al., 2014).

The Colli Albani area is characterized by moderate seismicity (max magnitude $Md = 4.0$) with sporadic low-magnitude seismic swarms (e.g., Amato et al., 1994). The most recent seismic crisis occurred in 1989–1990, when a swarm produced more than 3,000 events with $1.5 \leq Md \leq 4.0$, and hypocenters distributed within a

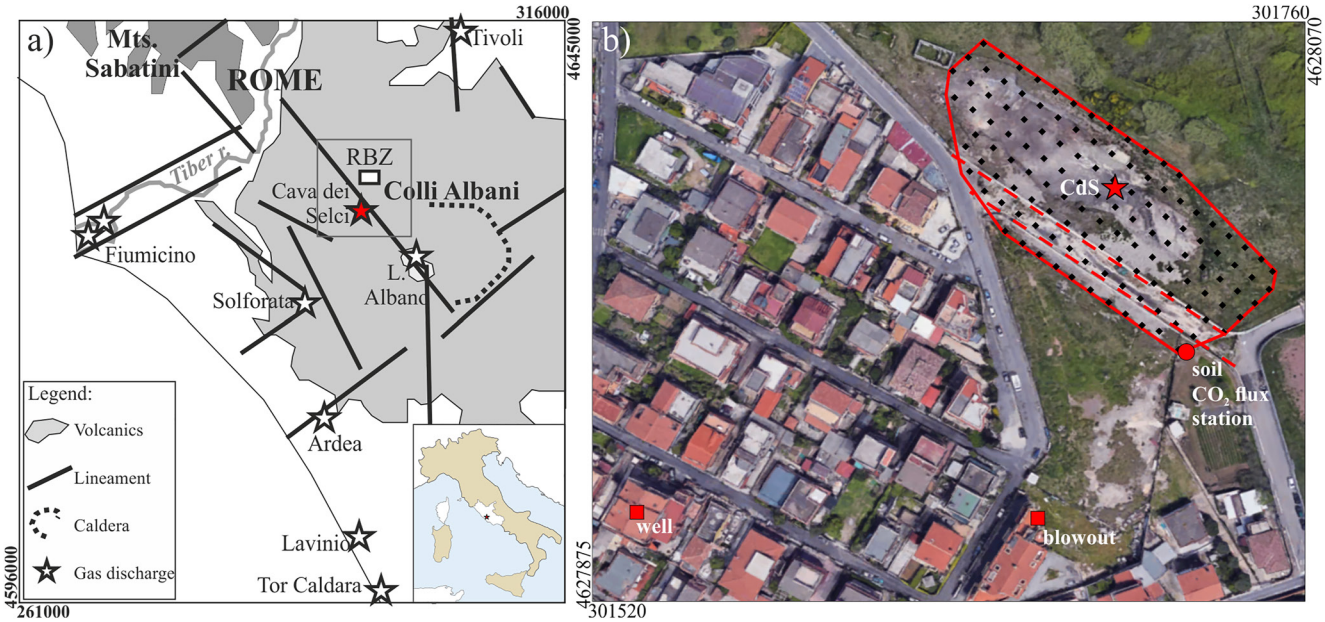

Figure 1. (a) Colli Albani volcanic complex (light gray area) with indication of Lake Albano, Cava dei Selci (CdS) gas discharge (red star), Rome gas blowout zone (RBZ) and location of the main gas discharges of the area (white stars); black segments are the main tectonic lineaments (after Acocella and Funiciello (2006)). (b) CdS gas discharge and nearby houses. The target area for diffuse soil CO_2 flux surveys is within the red perimeter; black dots are the measurement points. Red star indicates the CdS gas sampling site. Red dot is the site of the automatic monitoring station. Red squares indicate the gas sampling sites of 2008 blowout and of the well sampled in 2010. Red dashed lines indicate the unpaved road built in 2009. Satellite image from Esri. Coordinates in UTM datum WGS 1984 zone 33°N.

NW-SE striking ellipsoidal seismic belt of 6 × 12 km centered on the western part of the volcano (Lake Albano; Amato et al., 1994; Trasatti et al., 2018). According to Frepoli et al. (2010) the seismicity of the area shed light on present-day active tectonics confirming the change in principal stress orientation moving from the pure extensional stress regime of the Apennine chain to the transtensive regime of the Colli Albani and the eastern area of Rome.

The NW periphery of Colli Albani, centered on CdS, was investigated by an extensive survey of soil CO_2 flux measurements, carried out in February and March 2000, to identify areas displaying anomalously high CO_2 emissions. The results allowed identifying the area most prone to gas discharge in the central depression of the old quarry and showed that high CO_2 fluxes display an elliptical shape with major axis striking NW-SE (Carapezza et al., 2003, 2019). Since 2000, the diffuse soil CO_2 fluxes was periodically monitored in the old quarry, where it was established a "target area" of about ~6,000 m^2 (Carapezza et al., 2003).

3. Equipments and Methodology

3.1. Gas Sampling and Soil Gas Flux Procedures

Gas samples were collected inserting a 50-cm probe in the soil. To ensure the removal of air, gas was repeatedly pumped using a 100 cc syringe by a three-ways valve into two-ways Pyrex bottles (with vacuum valves at both sides). Chemical and isotopic ($\delta^{13}C_{CO_2}$ and $^3He/^4He$) analyses of gas were carried out at the INGV-Palermo labs, following the procedures described by Paonita et al. (2012).

We measured soil CO_2 flux by a West Systems accumulation chamber, following the method described by Chiodini et al. (1998). The device is equipped with an IR Licor-LI820 detector for CO_2 (single-beam dual wavelength NDIR; range 0–2 vol.%; accuracy: 3% of reading). Soil CO_2 flux are carried out on the CdS target area whose grid includes 63–142 measurement points and covers a 3,700–6,200 m^2 rectangular surface (generally 5,000–5,500 m^2; Figure 1b). Measurement sites, regularly 7 m spaced, are all flagged to ease their relocation during subsequent field campaigns. Between 10 and 15 points were located along the vegetated rim of the quarry in order to evaluate the contribution of organic CO_2 produced by local biologic activity (i.e., by soil respiration; Rey et al., 2002).

Fifty-five surveys were carried out in this target area between May 2000 and August 2020, with two time-breaks (2012–2016 and 2017–mid-2019). Most of the surveys include CO_2 measurements over the complete grid (typically 130 samples) but in the rainy periods a water pool forms in the center of the target area, making impossible the soil gas flux measurement in some sites (i.e., 10–20 points).

The total CO_2 output of each survey has been geostatistically estimated both by kriging (originally proposed by D.G. Krige in the 1950s; Wackernagel, 2003) and by the conditional sequential Gaussian simulation (sGs) algorithm of Deutsch and Journel (1992). The former method is a weighted summation of simple kriging residuals (observed minus mean) to make an estimation of values at unsampled points using information from surrounding sampled points, which eventually produces a single representation of the attribute (e.g., the soil CO_2 flux) and yields minimum error variance at each location. The latter is a conditioned and sequential simulation of a Gaussian cumulative distribution function that generates n equiprobable representations of the simple kriging, each reasonably honoring both the global statistics and the spatial features of the observed variable (Cardellini et al., 2003; Lewicki et al., 2005). Results of the surveys are summarized in Table S1 in Supporting Information S1; for the whole data set see Tarchini et al. (2021).

From February 2004 to June 2008, an automatic station to monitor soil CO_2 flux and environmental parameters was operative outside the SSE margin of CdS (location in Figure 1b; technical characteristics in Table S2 in Supporting Information S1). The time series of the soil CO_2 flux, and of environmental parameters (atmospheric pressure, air and soil temperature and humidity, wind direction and speed) that can potentially influence it, consists of dozens of thousands of hourly measurements for each parameter. The short-lasting data gaps within the observational period are due to device maintenance or failure. Part of the soil CO_2 flux measurements were removed from the pristine data set because of the high error due to vapor condensation on the mirror of the infrared gas analyzer occurring occasionally in the middle of the day, or to low battery voltage for poor solar radiation during winter time. Missing data of the environmental parameters (in particular the soil temperature) are due to sensors break or malfunctioning.

3.2. The Stochastic Gradient Boosting Trees Regression Method

In order to evaluate the influence of the environmental parameters (atm P, air and soil T and H, wind direction and speed) on the continuous data set of soil CO_2 flux, we applied the stochastic Gradient Boosting Trees (sGBT) regression method (module in Statistica 8.0). Results were then compared with two more classical statistical approaches: Multiple Linear Regression (MLR) and Principal Component Regression (PCR; Carapezza et al., 2009, 2012; Granieri et al., 2003, 2010; Laiolo et al., 2016; Oliveira et al., 2018; Viveiros et al., 2008, 2009, 2015).

The sGBT is an improvement of the Gradient Boosting Trees (GBT) method, which is a general and powerful machine-learning algorithm, mainly used for predictive data mining (Friedman, 1999; Friedman et al., 2000). The Gradient Boosting approach differs from traditional regressions, which produce a single best model. Conversely, the boosting technique is an additive regression model that sequentially adds a large number of relatively simple trees to optimize the predictive performance. The method computes a sequence of trees where each subsequent tree is built to minimize the residuals of the preceding. The first regression tree (the root) is the best partitioning, or splitting, of one (single decision stump) or more predictors that maximally reduces the deviances from the variable to be predicted, called pseudoresiduals, that is, the loss in predictive performance. At the following step, another best partitioning (that could contain quite different predictors and split points compared with the previous) is determined in order to reduce deviations of the pseudoresiduals, and then added to the model. In the following steps, new splittings are fitted to the pseudoresiduals of the previous n-terms model to furtherly reduce the residual error variance. This addition of learning trees advances until there is no significant reduction in the pseudoresiduals. This "additive weighted expansion" produces a high fit of the predicted values to the observed ones, even if the relation between dependent and predictor variables is complex (e.g., nonlinear or with interactions between predictors).

The sGBT improvement on the GBT method, results from the introduction of randomness by allowing at each step the splitting of the data set into a subset to train the model and a random subset to test the model. In fact, each consecutive tree is trained and tested only from randomly selected data subsets, thus each tree is built to

reduce residuals of the preceding trees on independent, and different at each step, randomly chosen samples. This stochastic procedure prevents the overfitting of the predictive response.

The main parameters that control the sGBT are:

1. The learning rate, a shrinkage weight with which the consecutive simple trees are added into the prediction equation (shrinkage values of 0.1 or less usually lead to better predictive validity preventing overfitting; Friedman, 1999; Friedman et al., 2000)
2. The random test-data proportion, a percentage of randomly chosen observations to serve as test sample in computation
3. The subsample proportion, a subsample percentage used for drawing the random training sample at each boosting step
4. The maximum number of nodes, the maximum number of splittings allowed from each parent node
5. The number of consecutive trees, which is the maximum number of additive trees before the test error increases (examples of model optimization can be found in Elith et al., 2008; Ghojogh & Crowley, 2019; Park & Ho, 2019)

The sGBT was here computed with the following parameters: Learning rate = 0.10; Random test-data proportion = 0.40; Random training-subset proportion = 0.60; Max No. of nodes = 5 (splitting to six nodes gave the same results); Number of trees = 297. Results show that the model is robust enough to hyperparameterization and that the training was stopped while the test error was still slowly decreasing (see Figure S1 in Supporting Information S1). Figure S2 in Supporting Information S1 shows the importance plot of the predicting variables for this sGBT model.

The Principal Component Regression (PCR) was performed by a multiple linear regression on the Varimax-normalized factors obtained from the principal component analysis of the environmental predictors (results in Table S3 in Supporting Information S1). Therefore, the forward-stepwise multiple linear regression has been applied on all factors, excluding wind direction.

The forward-stepwise (MLR) was performed on environmental variables selected by the model: air T (first predictor, 23.65% of CO_2 variance), soil RH (4.04%), air P (0.35%), wind speed (0.1%). This choice allowed to avoid overfitting due to the high correlations between some predicting variables (i.e., air T and soil T, air RH, and soil RH). Wind direction resulted in a high p-level and was not included in the linear-regression model. Results of MLR are reported in Table S4 in Supporting Information S1.

3.3. Seismicity

The seismic data presented in this study (data set in Frepoli and Pagliuca (2022)) were recorded by four seismic networks: INGV National Seismic Network (INGV Seismological Data Centre, 2006), RAN Italian Accelerometric network (only events with local or Richter magnitude $M_L \geq 2.5$), Regional Abruzzo-Lazio network (RSA, De Luca et al., 2009; Frepoli et al., 2017), IESN (Italian Earthquake Seismic Network; see Figure S3 in Supporting Information S1). Both national and regional networks have been significantly extended in the last two decades through installation of new three component stations, improving the detection and location of microseismicity. These data add to those of 2000–2008 previously published by Frepoli et al. (2010). Magnitudes are expressed either in *Md* (*duration magnitude* routinely used in the INGV Bulletin of the 1980s and 1990s, computed by measuring the length of the waveform on the seismogram), or in M_L (*local or Richter magnitude*, computed through the maximum *S* wave amplitudes measured on the seismogram) or in M_w (*moment magnitude*, authoritatively used to compute the magnitude of moderate to strong earthquakes, based on the seismic moment).

To accurately relocate the studied seismicity, we used the 1D SLAM velocity model (Frepoli et al., 2017). This 1D V_p velocity model was computed by the application of a genetic algorithm (GA), a method widely used to solve highly nonlinear inverse problems (Holland & Holland, 1975; Sambridge & Gallagher, 1993), on an extensive seismic data set of the Central Apennines recorded in the period 2009–2013 which also included the areas of Colli Albani and Mts. Prenestini-Tiburtini. We used the constant value of 1.85 for V_p/V_s determined with the Wadati method (Chatelain, 1978; Figure S4 in Supporting Information S1).

Table 1
Chemical and Isotopic Composition of CdS Gas

Site	Source	Date (d/m/yy)	He (ppm)	Ar (vol.%)	N_2 (vol.%)	CO_2 (vol.%)	CO (ppm)	H_2 (ppm)	H_2S (vol.%)	CH_4 (ppm)	$\delta^{13}C_{CO_2}$ (‰ versus PBD)	He/Ne	R/Rac
CdS	1	13 March 2000	2.6	n.a.	0.18	98.98	b.d.l.	b.d.l.	0.79	415	1.39	9.80	1.46
CdS	1	10 March 2004	2.2	n.a.	0.47	98.30	b.d.l.	b.d.l.	1.19	388	n.a.	n.a.	1.38
CdS	1	21 September 2004	2.5	n.a.	0.67	98.75	7.2	b.d.l.	0.53	420	n.a.	n.a.	1.40
CdS	2	29 July 2005	1.9	n.a.	0.13	99.03	b.d.l.	b.d.l.	0.80	422	0.75	42.51	1.34
CdS	1	6 February 2007	2.4	0.0630	0.38	98.59	0.5	b.d.l.	0.93	452	0.78	69.30	1.44
CdS	1	30 June 2008	2.4	0.1040	0.64	98.59	b.d.l.	b.d.l.	0.62	451	1.18	3.20	1.45
CdS_blowout	1	30 June 2008	2.8	0.1621	1.11	98.29	0.1	b.d.l.	0.37	642	1.03	2.10	1.51
CdS	1	29 March 2010	2.3	n.a.	0.03	99.02	b.d.l.	b.d.l.	0.91	453	0.87	36.80	1.42
CdS_well	1	30 March 2010	2.6	n.a.	0.37	99.23	b.d.l.	b.d.l.	0.35	497	0.85	30.20	1.37
CdS	3	6 April 2016	1.1	0.0032	0.38	98.61	3.3	3.60	0.92	821	1.08	n.a.	1.41
CdS	This paper	11 June 2019	2.4	0.0105	0.26	98.88	b.d.l.	b.d.l.	0.80	478	0.86	17.44	1.44
CdS	This paper	24 June 2019	2.4	0.0145	0.29	98.86	b.d.l.	b.d.l.	0.79	480	0.93	11.52	1.45
CdS_station	This paper	30 Octoer2019	2.9	0.0750	1.17	98.75	b.d.l.	0.02	0	1	−1.10	1.14	1.31
CdS	This paper	30 August2020	1.9	0.0181	0.54	98.77	0.2	3.00	0.63	483	0.7	13.09	1.40

Note. Gas values are corrected for air contamination (max O_2 vol.% = 0.6 for CdS and 1.36 for CdS_station). n.a. = not analyzed; b.d.l. = below detection limit. Source: (1) Carapezza et al. (2012); (2) Barberi et al. (2007); (3) Venturi et al. (2019).

The waveforms of local earthquakes were extracted from the continuous data stream of the INGV National Seismic Network, and merged with those of the other available networks. Seismogram collection of the RSA and IESN networks was performed using the trigger times based on standard STA/LTA algorithm ratio methods. We manually picked the *P*-arrival and *S*-arrival times mostly on the unfiltered seismograms and we assigned a weight to each time pick on the basis of the onsets quality (Table S5 in Supporting Information S1).

A total of 1,062 earthquakes (487 from Frepoli et al., 2010) were relocated with the HypoDD algorithm (Waldhauser & Ellsworth, 2000). HypoDD is a specific software for relocating earthquakes where the residual between observed and calculated traveltimes difference (or double-difference) between two events at a common station are related to adjustments in the relative position of the hypocenters and origin times. Residuals for pairs of earthquakes at each station are minimized by weighted least squares. Solutions are found by iteratively adjusting the vector difference between nearby hypocentral pairs. The double-difference relocation brings structural details, such as the location of active fault plane into sharp focus.

4. Results

4.1. Gas Geochemistry

The chemical and isotopic analyses of gas sampled along the years from one of the most emissive CdS vents are reported in Table 1 together with those of the gas emitted in the 2008 blowout and from a nearby abandoned water well. Table 1 reports also the analysis of the soil gas sampled in the site of the automatic station, external to the gas discharge (sampling sites are shown in Figure 1b). Gas is dominated by CO_2, with concentrations ranging from 98.29 to 99.23 vol.% (avg. 98.76 vol.%), followed by H_2S (from 0.35 to 1.19 vol.%; average 0.74 vol.%), N_2 (from 0.03 to 1.17 vol.%; average 0.47 vol.%) and CH_4 (388–821 ppm; average 492 ppm). Concentrations of Ar range from 0.003 to 0.16 vol.%. Those of H_2 and CO are frequently below detection limit or of a few ppm; also helium has very low concentrations ranging between 1.1 and 2.9 ppm.

The deep origin of the gas, likely from the Colli Albani magmatic reservoir, is indicated by its helium isotopic composition (R/Rac = 1.31–1.51; Table 1), which is similar to that of the fluid inclusions in phenocrysts of Colli Albani volcanic rocks (R/Ra = 1.17–1.73; Martelli et al., 2004) and higher than in the gas discharges of the volcanic districts north of Rome (Carapezza et al., 2015; Cinti et al., 2011; Ranaldi et al., 2016). These R/Ra

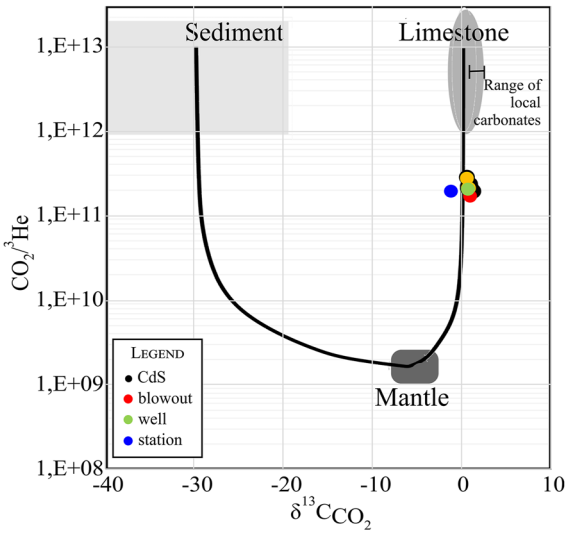

Figure 2. Carbon and helium isotopic composition of Cava dei Selci gas in the period 2000–2020 ($CO_2/^3He$ versus $\delta^{13}C_{CO_2}$ diagram modified after Sano and Marty (1995)). The gray areas are the broad ranges of terrestrial values for sediments and limestones; the $\delta^{13}C_{CO_2}$ range of local limestones is after Fornaseri and Turi (1969).

values are among the lowest measured in subduction zones worldwide, and this can be explained by the melting of a HIMU-like mantle chemically modified by the subduction of continental crust (Martelli et al., 2004). The $\delta^{13}C_{CO_2}$ values of CdS gas vary from −1.10‰ to 1.39‰ versus PBD (Table 1). This suggests a mixing between a deep CO_2 deriving from a crustally contaminated mantle or magma ($\delta^{13}C_{CO_2} = -3‰$ versus PBD; Chiodini et al., 2010) and a CO_2 deriving from metamorphic decarbonation at magma-limestone contact in the magma chamber ($\delta^{13}C_{CO_2} = \geq 2‰$ versus PBD; Chiodini et al., 2010; see Figure 2).

4.2. Soil CO_2 Flux Surveys at CdS Gas Discharge in 2000–2020

The mean soil CO_2 flux values calculated from 55 surveys from the target area collected in 2000–2020, range from 877 to 4,735 g m^{-2} d^{-1} (Table S1 in Supporting Information S1), with a mean arithmetic value for all the 55 surveys of ~2,100 g m^{-2} d^{-1} and a standard deviation (1σ) of ~900 g m^{-2} d^{-1}. Only a few points, representing less than 3% of the overall number of measurements, gave soil CO_2 flux values below 30 g m^{-2} d^{-1}, typical of vegetated soils (Rey et al., 2002), a value that was considered the local threshold between gas flux of biological and endogenous origin (Carapezza et al., 2003). Values of the CO_2 flux can differ by three orders of magnitude at distances of only a few dozens of meters, which suggests that operations during and postquarrying may have produced local compaction and fracturing in the surface lithology, creating preferential gas pathways to the surface.

Several studies, focused on the comparison of interpolation techniques (e.g., Cardellini et al., 2003; De Bortoli Teixeira et al., 2011; Lewicki et al., 2005), concluded that the conditional sequential Gaussian simulations procedure produces more realistic images of the spatial distribution of the CO_2 fluxes as well as a better quantification of the total CO_2 output over the degassing area since it honors the univariate (histogram) and the bivariate (variogram) statistics of the original measurements (Cardellini et al., 2003). Therefore, the maps of the soil CO_2 flux and the total soil gas output, of each CdS survey, were elaborated by the sGs procedure. To allow a comparison, total outputs were also compared with those estimated by the ordinary kriging technique (OK; 2000–2008 data after Carapezza, Barberi, et al. (2010), 2009–2020 data this work; Figure S5 in Supporting Information S1).

The total soil CO_2 output from the target area over 20 years (from 2000 to 2020, with interruptions in 2013–2016 and 2018) is highly fluctuating (Figure 3a), with a minimum emission of 5.6 t d^{-1} in December 2003 and a maximum emission of 24.8 t d^{-1} in January 2006 (maps in Figures 3b and 3c, respectively). The estimated mean CO_2 output over the whole period is 12.1 t d^{-1} and the standard deviation (1σ) is 4.5 t d^{-1} (Figure 3a). The "average" output condition is well represented by the May 2020 survey (map in Figure 3d) whose estimated total CO_2 flux is 12.4 t d^{-1}. The probability plots of the soil CO_2 flux measurements for the December 2003 January 2006 and May 2020 surveys (Figures 3b-3d) show that their statistical distribution is the result of three combined log-normal populations representing high, average and low flux families (HF, AF, LF, respectively). The HF population of January 2006 is characterized by a very high mean value (8,039 g m^{-2} d^{-1}) in comparison with those of May 2020 (2,242 g m^{-2} d^{-1}) and December 2003 (356 g m^{-2} d^{-1}). This suggests a different contribution of the deep source over time. The LF populations (mean value of 48.5 g m^{-2} d^{-1} in December 2003, 49.7 g m^{-2} d^{-1} in January 2006 and 57.6 g m^{-2} d^{-1} in May 2020) are likely the best indication of the local biogenic contribution in different seasons. All the statistical parameters of the partitioned populations are summarized in Table S6 in Supporting Information S1.

The soil CO_2 flux maps show typically a highly emissive area in the internal sector of the investigated grid, with elongation NW-SE (Figure 3b).

Part of the target area was affected in October 2009 by the building of a road (dashed lines in Figure 1b), which, though unpaved, might have modified the soil gas permeability; however, no significant changes have been recorded in the gas release from that zone. Since then, the main gas emission zone has been fenced off, but small animals (cats, birds) are still frequently found dead within the CdS depression.

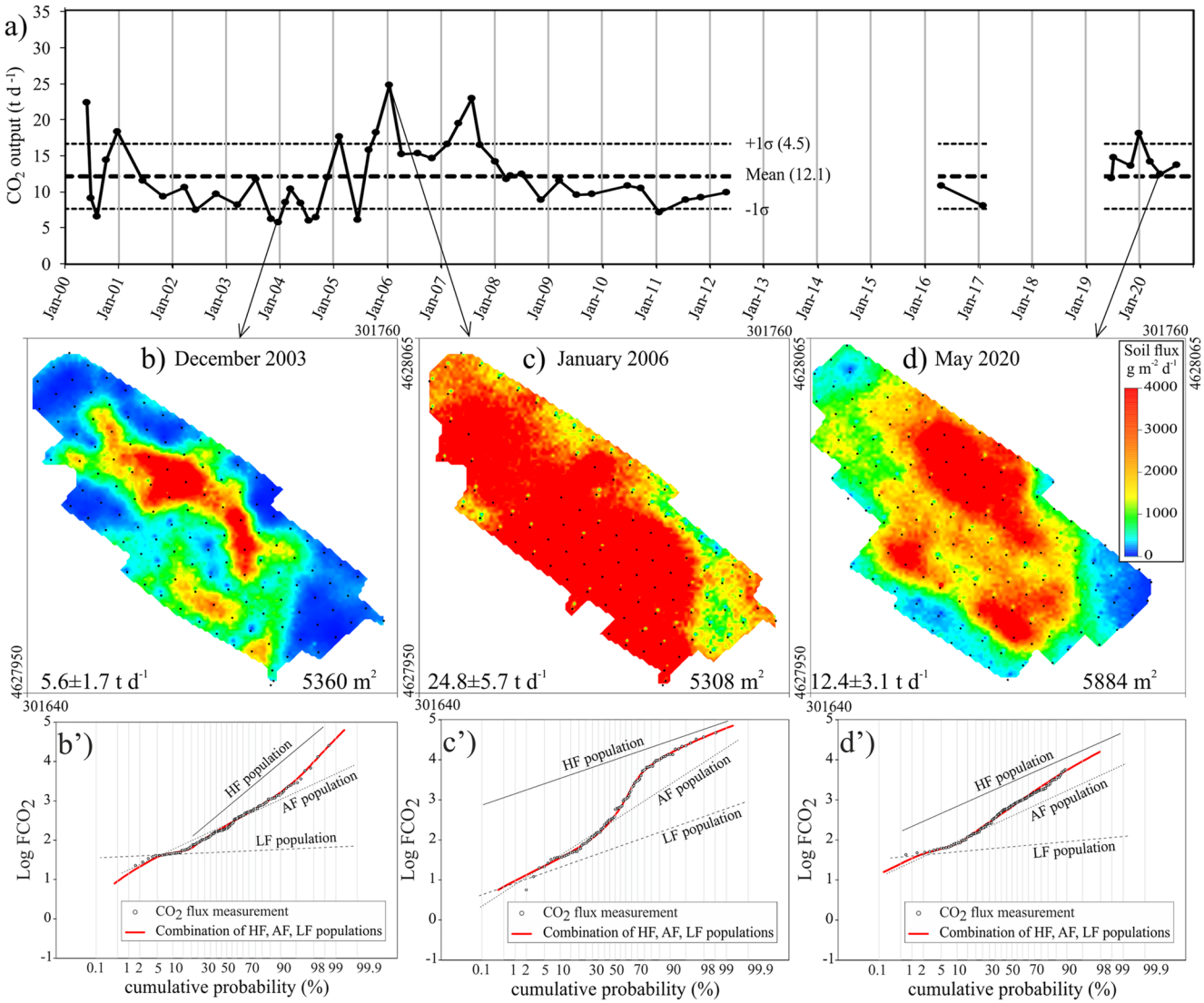

Figure 3. (a) Time variation of total diffuse soil CO_2 emission from the fixed grid of Cava dei Selci (CdS), with indication of the mean and the standard deviation (dashed horizontal lines). (b–d) Maps of lowest, highest and medium soil CO_2 emission from the target area (all maps have the color scale of Figure 3d). In each map the estimated total soil CO_2 output is indicated together with the extension of the surveyed area. Black dots are measurement points. The probability plots of the soil CO_2 flux from these surveys are shown in (b'–d').

4.3. Soil CO_2 Flux and Environmental Parameters: 2004–2008 Time Series

4.3.1. Analysis of the Acquired Time Series

Data of soil CO_2 flux and environmental parameters acquired from the continuous CdS station have been statistically treated and main results are presented in Table 2.

The data correlation (last column in Table 2) highlights that soil CO_2 flux is positively correlated with air temperature ($r = 0.47$) and soil temperature ($r = 0.43$) and inversely correlated with soil moisture and air relative humidity ($r = -0.45$ and $r = -0.32$, respectively). Wind direction, wind speed and air pressure are not correlated with CO_2 flux ($r = 0.11$, 0.09, and 0.05, respectively).

In Figure 4a, the main correlated parameters are first grouped by the hour of measurement (local time, LT) and then averaged. It can be observed that in CdS station site, higher soil CO_2 emissions occur in the afternoon (maximum at 18:00, LT) as a consequence of soil heating due to the solar radiation in the warmest hours of the day (air T maximum at 15:00–16:00, LT and soil T maximum at 20:00, LT) and with relatively low soil moisture

Table 2
Descriptive Statistics and Correlation Matrix of the Data Sets Collected by the CdS Automatic Station, From February 2004 to June 2008

Parameter	Valid data (#)	Mean	Min	Max	Std. Dev.	Correlation (r) with CO_2 flux
Air T (°C)	32,486	15.0	−5.5	43.4	8.2	0.47
Air RH (%)	32,522	74.4	2.0	98.7	21.2	−0.32
Air P (hPa)	28,036	997.4	948.4	1020.8	7.0	0.05
Soil T (°C)	24,085	16.5	2.3	28.6	6.2	0.43
Soil RH (%)	30,436	49.7	17.5	100.0	16.5	−0.45
Wind speed (m s^{-1})	32,632	0.9	0.05	9.7	1.1	0.09
Wind dir (°N)[a]	20,884	187.8	0.0	358.6	37.8	0.11
Soil CO_2 flux (g m^{-2} d^{-1})	28,171	335.3	10.1	4286.4	244.7	–

[a]As the lower detection limit of the anemometer is 0.05 m/s, we considered as "calm" all wind speeds <0.05 m/s and the corresponding wind directions were considered as not measurable.

(minimum soil moisture at 14:00, LT). Similar findings are obtained considering the monthly means of diffuse CO_2 flux measured during 4 years. The time series shows higher values during the warmer months (i.e., July and August), when the conditions of the local soil are typically dry (Figure 4b). Indeed, as shown by Camarda et al. (2019), variations in soil moisture and temperature can affect the transport of soil CO_2 through the soil.

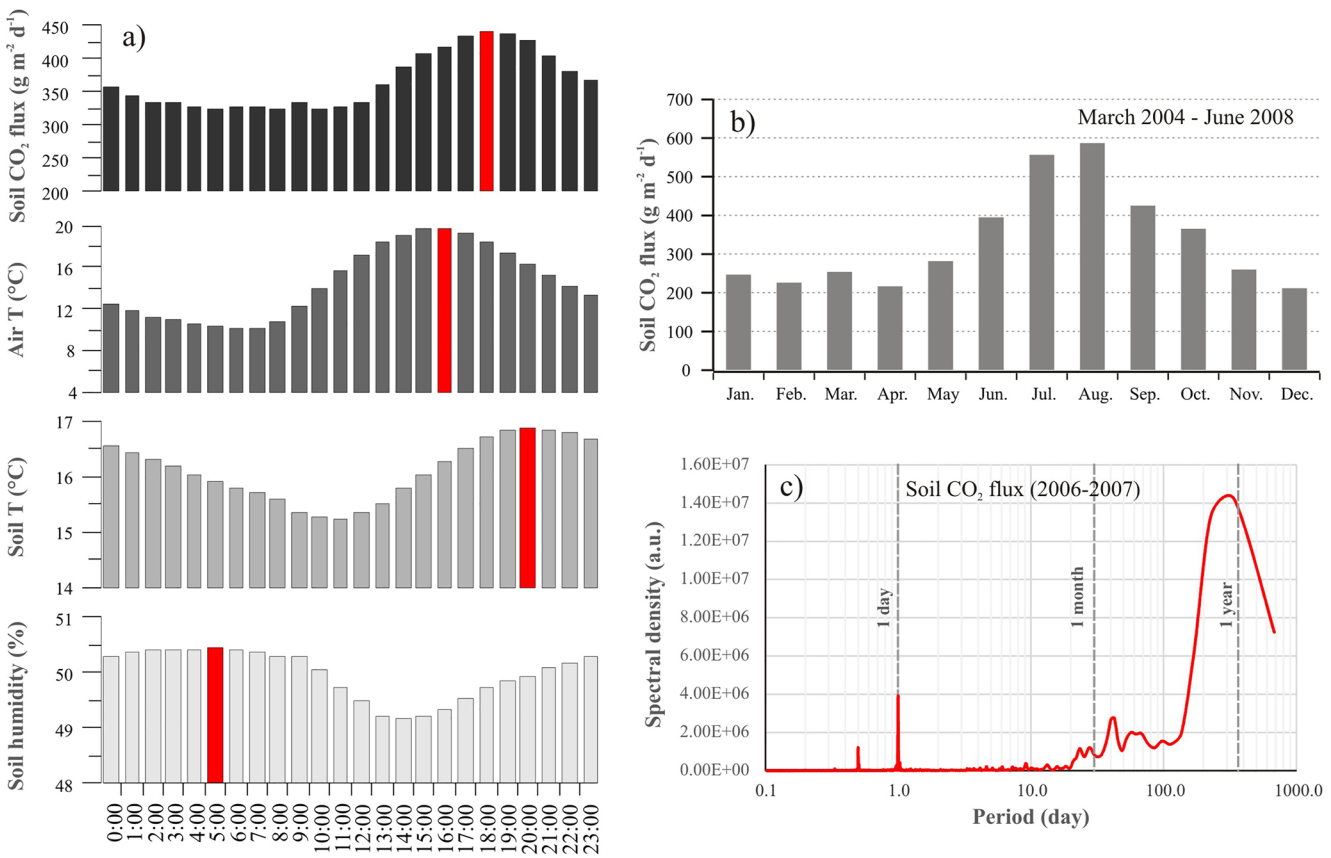

Figure 4. (a) Soil CO_2 flux, air T, soil T, and humidity clustered by local hour; each data set includes more than 30,000 measurements (~24,000 for the soil T series, see Table 2); red bars denote the maximum values of the series; (b) mean of the diffuse CO_2 flux clustered over monthly periods; (c) spectral analysis of the soil CO_2 flux over the period 2006–2007.

The soil CO_2 flux series was examined for the presence of periodicities through the application of the Fast Fourier Transform (FFT). Since this spectral analysis requires that the series is consecutive, that is, that no missing data be present, we first filled the voids with extrapolated values based on the extreme measurements of the empty intervals. The analysis was applied only for the 2006–2007 period, in which missing data are about 16% (2,885 missing CO_2 flux measurements out of a total of 17,520). The soil CO_2 flux spectrum (Figure 4c) highlights both the high frequencies due to the diurnal (and weaker semidiurnal) cycle and the low frequencies due to a 341-day cycle, representing the annual periodicity. As in other studies (e.g., Oliveira et al., 2018; Viveiros et al., 2014), the divergence to the theoretical 365 days can be imputed to the limited length (just 2 years) of the considered time series.

4.3.2. Regression Results

The time variation of the soil CO_2 flux recorded for over 4 years by the CdS automatic station is shown in Figure 5a, together with the data predicted by the stochastic Gradient Boosting Trees-sGBT regression and its relative residuals (i.e., observed minus predicted data in Figure 5a″).

In Figures 5b and 5c, we present, respectively, the results obtained by the MLR (Multiple Linear Regression) and by the PCR (Principal Component Regression). Their relative residuals are reported in Figures 5b″ and 5c″, respectively.

Given the features of the sGBT algorithm, it was possible to predict the soil CO_2 flux response of the whole data set (32,632 responses, one for each hour of automated operation).

The sGBT predicted up to 56.5% of soil CO_2 flux variance (R^2 in Figure 5a′). The PCR predicted 27.9% of soil CO_2 flux variance (R^2 in Figure 5b′). Due to the method limitation, PCR could only predict 17,395 valid cases, that is, 53.31% of the total observations. Figure 5b″ shows the PCR residuals.

MLR predicted up to 28.1% of the soil CO_2 variance (R^2 in Figure 5c′). Given some missing environmental data (see subchapter 3.2 for explanation), MLR could only predict 22,047 soil CO_2 flux responses out of the total 32,632 soil CO_2 flux measures (67.56%). Figure 5c compares the observed values with the MLR predicted ones, and Figure 5c″ shows the residuals.

As shown in Figures 5b, 5c and 5b′, 5c′, PCR and MLR do not predict high soil CO_2 fluxes (maximum predicted with PCR = 893.4 g m^{-2} d^{-1} and with MLR = 781.5 g m^{-2} d^{-1}), as opposed to sGBT, which predicted values up to 2750.9 g m^{-2} d^{-1} (Figures 5a and 5a′).

In addition, sGBT provides a complete predicted data set (i.e., 32,632 responses) while MLR and PCR provided only 22,047 and 17,395 responses, respectively, due to missing data in the recorded series. The reason being that sGBT can estimate responses for each hourly recording, regardless of missing observations in predictors (i.e., environmental parameters) or response variable (i.e., CO_2 flux). This depends on the different computational approach of these methods. MLR and PCR "predict" in the general form

$$Ri = a_0 + \sum_{i=1}^{n} (a_i \times P_i)$$

where Ri are the predictions, a_0 and a_i are constants and P_i are the predictors. So, if only one of the predictors is missing in ith case, the ith response is not estimated. On the contrary, sGBT works by reducing the residuals of a loss-function at each following tree, and when a predictor value is missing it is replaced with the overall mean of the predictor for the residual analysis. In this way, every case is eventually included in the computations.

Looking at the shape of the curves of observed versus predicted values (Figures 5a′–5c′), two families with different slopes emerge: a first one including the majority of data and a smaller one including the highest observed values (up to 4,286 g m^{-2} d^{-1}) which are largely underestimated by PCR and MLR models. The high values of the second family correspond to the positive peaks recorded in March 2005 and August 2007 (Figure 5, observed series), which likely indicate an anomalous gas flux of endogenous origin. This is particularly clear because of the high values of the residuals (exceeding 2σ) shown in Figures 5a″–5c″. In August 2007, the maximum prediction of sGBT was 2750.9 g m^{-2} d^{-1}; although the observed and sGBT predicted trends are similar (Figure 5a), a huge residual anomaly is recorded (Figure 5a″) suggesting a dominant endogenous gas origin. It has to be noted that sGBT results in more negative residuals during the 2007 period of high soil CO_2 flux. Conversely, the sGBT

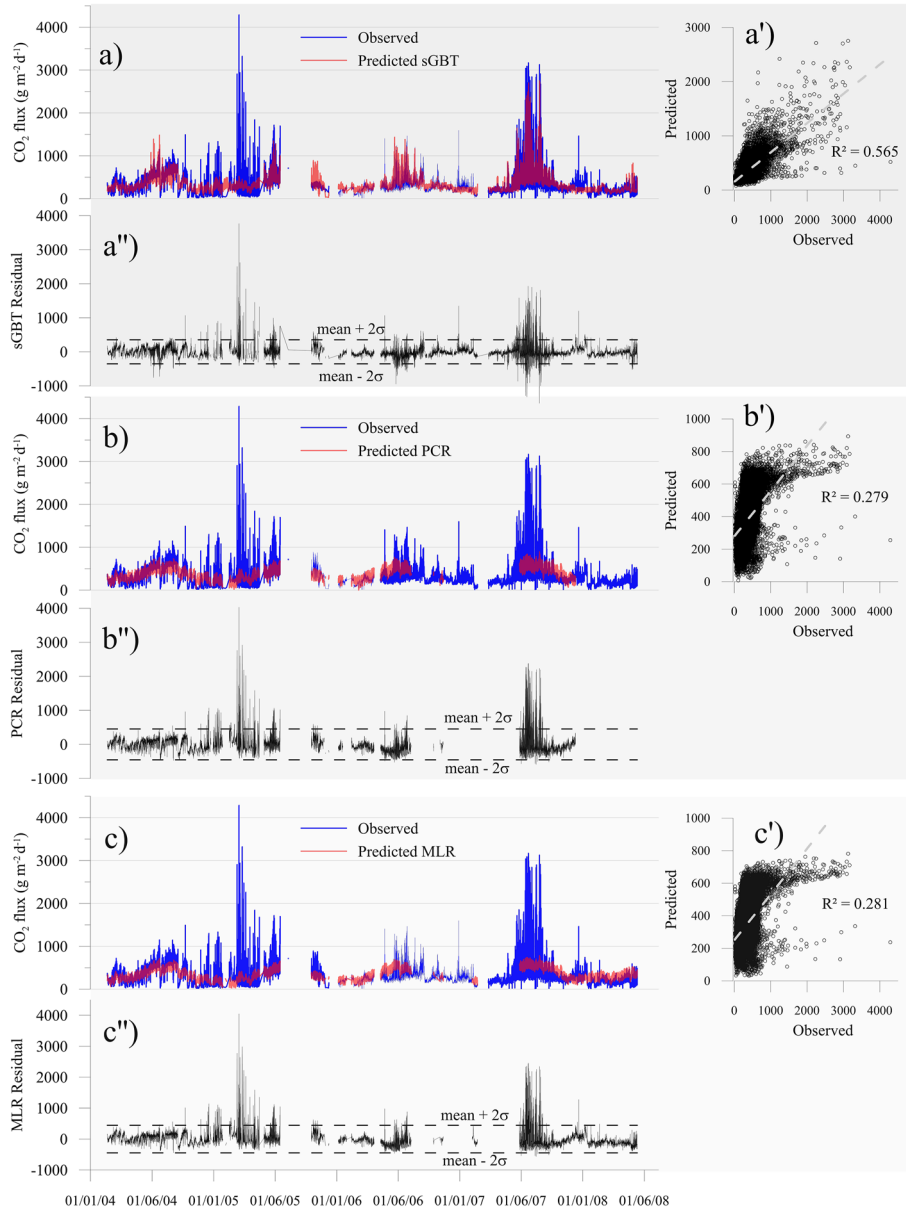

Figure 5. (a–c) Time series of observed data (blue lines) compared with values predicted (red lines) by sGBT, PCR, MLR methods, respectively; (a′–c′) Observed versus predicted values and (a″–c″) residuals (i.e., observed minus predicted) with dashed lines indicating $\pm 2\sigma$. All units in g m^{-2} d^{-1}.

residuals standard deviation is lower (161.5 g m^{-2} d^{-1}) than the PCR and MLR standard deviations (respectively 236.0 and 222.9 g m^{-2} d^{-1}). This likely depends on the nature of the observed CO_2 flux: in fact, periods of "low" emissions span from absolute minima to few hundreds of g m^{-2} d^{-1}, while periods of high emissions are characterized by large variations from relative minima to maxima of many hundreds of g m^{-2} d^{-1}. Therefore, as sGBT succeeds in predicting the high emission rates, these predictions are also more distant to the minima recorded values resulting in lower residuals. On the contrary, PCR and MLR do not succeed in predicting very high values, hence the residuals are more distant to high observations and closer to low observations.

Anyway, it is worth noting that the three residual curves (Figures 5a″–5c″) indicate the same two main anomalous periods (March 2005 and summer 2007). Both gas anomalies are shortly preceded by high total CO_2 output from the target area (17.6 t d^{-1} in February 2005 and 22.9 t d^{-1} in July 2007). Unfortunately, the station was out of service when the maximum gas output from the target area was recorded (24.8 t d^{-1} in January 2006). This time

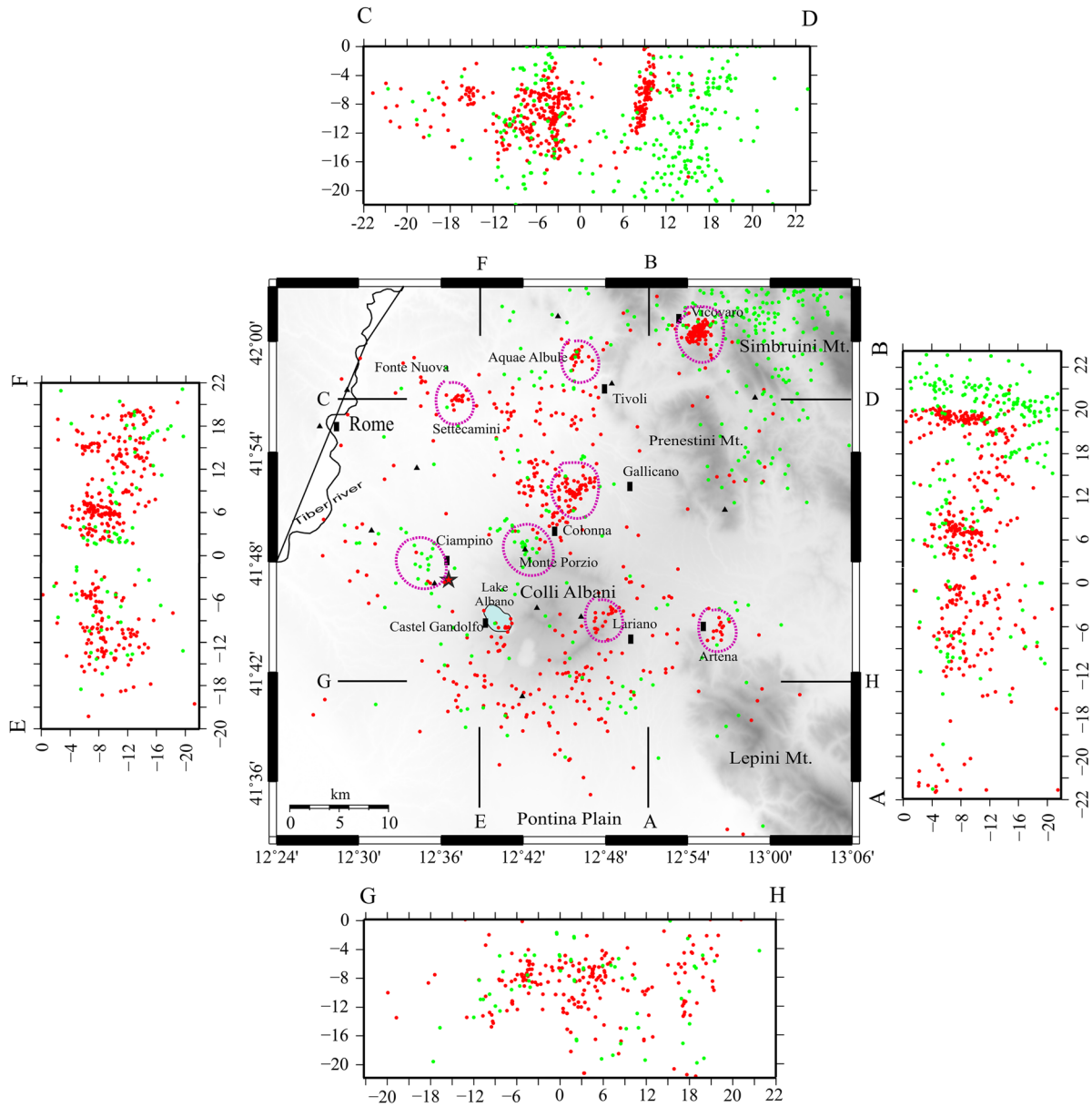

Figure 6. Seismicity distribution of the study area and vertical cross sections; green dots represent the 1997–2008 seismicity (after Frepoli et al. (2010)), red dots represent the 2009–2021 seismicity. Red star: Cava dei Selci gas discharge; magenta dashed ellipse: earthquakes' cluster.

difference in recording high flux anomalies at the station likely depends on its relative distance from the main gas-discharging tectonic fractures.

4.4. Colli Albani Seismic Data Analysis

The seismicity of Colli Albani and surrounding areas in the period 1997–2008, reported as green dots in Figures 6 and 7a, has been studied by Frepoli et al. (2010), who identified two major clusters of seismicity: Monte Porzio (January-March 2006, maximum $M_L = 2.7$) and Ciampino area (April 2008, maximum $M_w = 3.9$).

To integrate this information, we analyzed the seismicity from January 2009 to March 2021 and determined 575 earthquake hypocenter depths with acceptable uncertainty. For the whole data set see Frepoli and Pagliuca (2022). Most of the hypocenters were located within clusters to the north (Vicovaro, Tivoli, Settecamini

and Colonna-Gallicano clusters), and to the east (Lariano and Artena clusters) of Colli Albani (red dots in Figure 6). Seismicity is more scattered toward south and west of Colli Albani.

Hypocenter depth extends down to 20 km with a pronounced maximum between 6 and 14 km. The average root mean square residuals of the 575 selected events is 0.11s while the average location errors are 0.32 km (horizontally) and 0.55 km (vertically) with a confidence level of 90%. The final 2009–2021 data set consists of 11,163 *P* and 9,380 *S* wave arrivals. The magnitude, with values range from 0.3 to 3.6, was extracted from the INGV earthquake database (ISIDe Working Group, 2007).

4.4.1. Focal Mechanism Computation and HypoDD Relocation Results

We have computed 43 new fault-plane solutions using the FPFIT code of Reasenberg and Oppenheimer (1985). Table S7 in Table Supporting Information S1 shows the parameters of the 54 selected fault-plane solutions, 11 of which after Frepoli et al. (2017); the stress regime is indicated following Zoback (1992).

Focal mechanism selection was performed on the basis of the values of Q_f and Q_p, the two FPFIT output quality factors (Table S8 in Supporting Information S1). Q_f gives information about the solution misfit of the polarity data F_j, whereas Q_p reflects the solution uniqueness in terms of the 90% confidence region for strike (Δs), dip (Δd), and rake (Δr). The quality factors range from A to C for decreasing quality. All focal mechanisms with one or both quality factors C, and with less than 10 polarities, were rejected. The majority of the selected focal mechanisms represent pure normal faults and normal faults with strike-slip component (44.4%), whereas the 42.6% are pure strike-slip mechanisms. Only 7.4% are reverse with strike-slip component and 5.6% are odd mechanisms. The *T* axis of the selected solutions suggests a widespread NE-SW extension regime in the whole study area with an average trend of N30°E (Figures 7d and 8). The only two sectors characterized by NW-SE and E-W oriented *T* axes are in the southern side of Colli Albani and in Aquae Albule-Tivoli and Vicovaro areas (Figure 7d).

Focal mechanisms of the small January 2006 Monte Porzio seismic sequence (Figure 7a) accounts for a ~N150° striking normal fault in accordance with the SW dipping alignment of hypocenters. The April 2008 Ciampino sequence is characterized by left-lateral N-S oriented strike-slip solutions (Figure 7a) arranged in an elongated subvertical cluster with hypocentral depth between 7.5 and 11 km. This transtensive regime in the Ciampino sector is linked to the structural character of the Rome area, where preexisting subvertical N-S faults are reactivated under the extensional field with left-lateral movement (Frepoli et al., 2010; Marra, 2001). This fault activity indicates that Colli Albani volcano is affected by active NE-SW extensional deformation, although the moderate magnitude associated with these structures accounts for very slow strain rates.

The 2009–2021 seismicity distribution of Colli Albani analyzed in this study is quite scattered (red dots in Figure 6). There are only two main seismic clusters with a consistent number of events: the first one in the Colonna-Gallicano area and the second one in the Vicovaro area. We performed the HypoDD algorithm for these two clusters to identify clear fault dip through hypocenter alignments. The small number of events did not allow applying the HypoDD algorithm for the smaller clusters (Settecamini, Aquae Albule-Tivoli, Lariano and Artena; Figure 6).

The Colonna-Gallicano cluster shows seismic activity in the period 2009–2011 (purple dots in Figure 9a; 21 events) and a reactivation begun in 2019 (green dots in Figure 9a; 43 events). The 2009–2011 seismicity is slightly shifted to the northeast. This seismicity shows a structure dipping to the SW at a high angle (around 65–70°) within 6–10 km of depth. Magnitude of these events is ranging between 0.5 and 3.6.

The largest shock is the Colonna earthquake of 23 June 2019 with seismic moment magnitude $M_w = 3.6$. Focal mechanism of the Colonna main shock is a strike-slip solution (Figure 7c). We computed eight more fault-plane solutions of this cluster. The majority of these solutions are strike-slip with left-lateral movement following the N-S nodal plane. Only two are normal fault solutions with *T* axis NE-SW oriented. We observe the same kinematics in the Ciampino cluster (2008) and in the recent seismicity of Fonte Nuova ($M_w = 3.3$; May 2020; Figures 8b and 8c).

The Vicovaro cluster shows the largest spatial extent and the highest number of earthquakes (81 events) of this study (Figure 9b, red dots). Seismicity of this cluster is concentrated in the period 15 December 2020–2027 March 2021. This cluster shows a swarm activity: it started with a magnitude $M_L = 2.1$ and reached the largest magnitude ($M_L = 2.6$) on 19 December 2020; 16 earthquakes have magnitude ranging from 2.0 to 2.6. The HypoDD relocation shows a cluster elongated from 2 to 12 km of depth dipping to the WSW. This cluster is

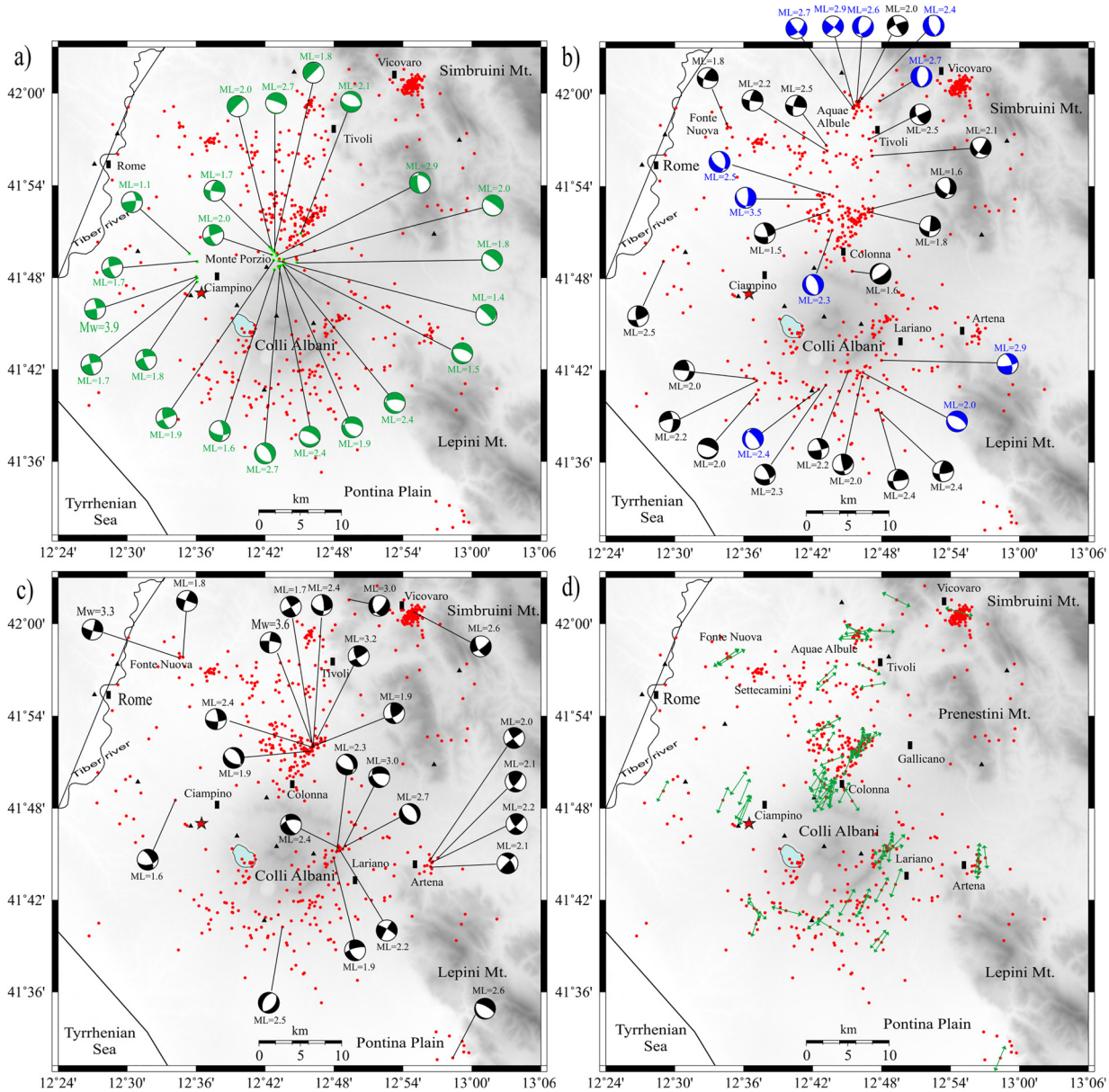

Figure 7. Location of the 54 selected fault-plane solutions. Event magnitude is shown close to each focal mechanisms; (a) period 1999–2008 (green beach balls after Frepoli et al. (2010); Monte Porzio and Ciampino clusters); (b) period 2009–2017; (c) period 2018–2020; (d) *T* axes distribution. Red star: Cava dei Selci gas discharge; light blue ellipse: Lake Albano.

located ~5 km SW of the Arsoli cluster studied by Frepoli et al. (2010) with seismicity belonging to the period 2000–2006 (Figure 9b, green dots). Both clusters show the same dip to the WSW at an angle of ~70°. The Arsoli cluster is more extended in depth (2–21 km) and is characterized by a gap of seismicity in the 4–6 km depth range. Unfortunately, due to the small magnitude of the events and the few polarities available, we computed only one focal mechanism for the Vicovaro cluster (Table S7 in Supporting Information S1, event no. 54). This fault-plane solution belongs to the largest magnitude earthquake within this cluster (M_L = 2.6; hypocentral depth = 6 km). One nodal plane of this solution is dipping to the WSW with an angle of ~70°, very similar to the orientation and the dip angle of the cluster. *T* axis is ESE oriented. Both Arsoli and Vicovaro clusters fall within the area characterized by pure extensional stress regime of the Apennine chain.

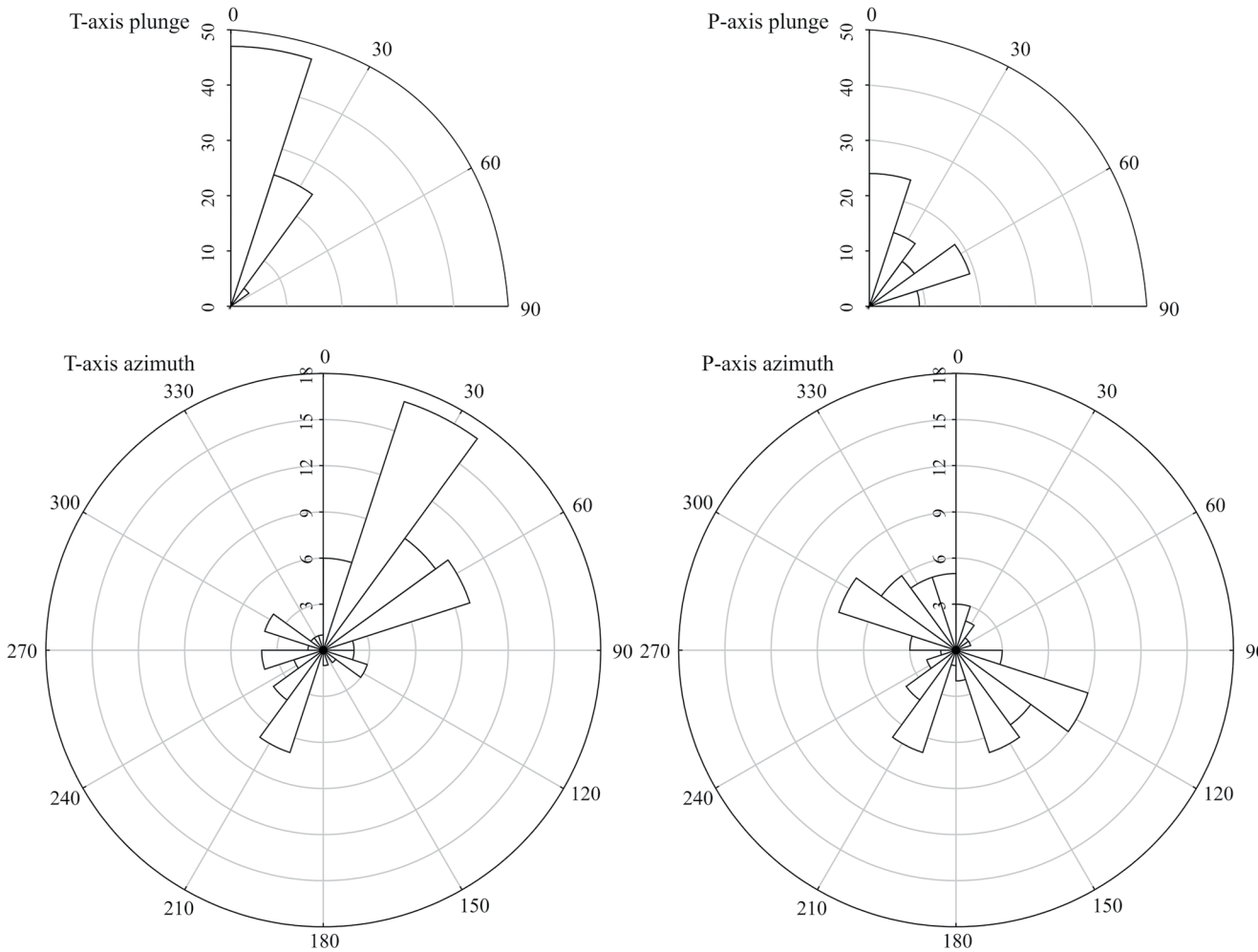

Figure 8. Rose diagrams showing *P* and *T* axes plunge and azimuth distribution of 76 fault-plane solutions: 22 fault-plane solutions after Frepoli et al. (2010) belonging to the Monte Porzio and Ciampino clusters (2006–2008), and 54 focal mechanisms from this study.

5. Discussion

5.1. Soil CO_2 Flux Surveys

From May 2000 to August 2020, 55 soil CO_2 flux surveys have been carried out at Cava dei Selci (Table S1 in Supporting Information S1). All data have been re-elaborated with the sGs method in order to (a) derive the total soil CO_2 output, (b) quantify the uncertainty of the total CO_2 output estimation, and (c) draw maps of the CO_2 fluxes over the degassing area. Data reprocessing by sGs method provides some important differences with respect to the results obtained by OK method, although the two series show the same temporal variation trend (Figure S5 in Supporting Information S1). In particular, Carapezza, Barberi, et al. (2010) who presented the 2000–2008 surveys processed by OK, attributed the maximum total CO_2 output (25 t d^{-1}) to the May 2000 survey. Data reprocessing lowers the total flux output of that survey to 22.4 t d^{-1}, and assesses the maximum CO_2 output (24.8 t d^{-1}) in January 2006, when the previous estimation was 13.8 t d^{-1} (Figure S5 in Supporting Information S1). Apart from the first period (2000–mid-2001) and the May 2020 survey, the series of the total CO_2 output derived with the sGs method shows values higher than those of the OK series. The sGs model is actually more rigorous in representing data with high variation over small distances, whereas the OK interpolator tends to promote the smoothing of data typically underestimating higher values. This may explain also some high differences of the total CO_2 output estimated by the two methods (e.g., July 2003 and January 2006; Figure S5 in Supporting Information S1). Moreover, as sGs method generates a set of equiprobable representations (100 in this application), each survey is now completed with its relative uncertainty (see Figure 3a).

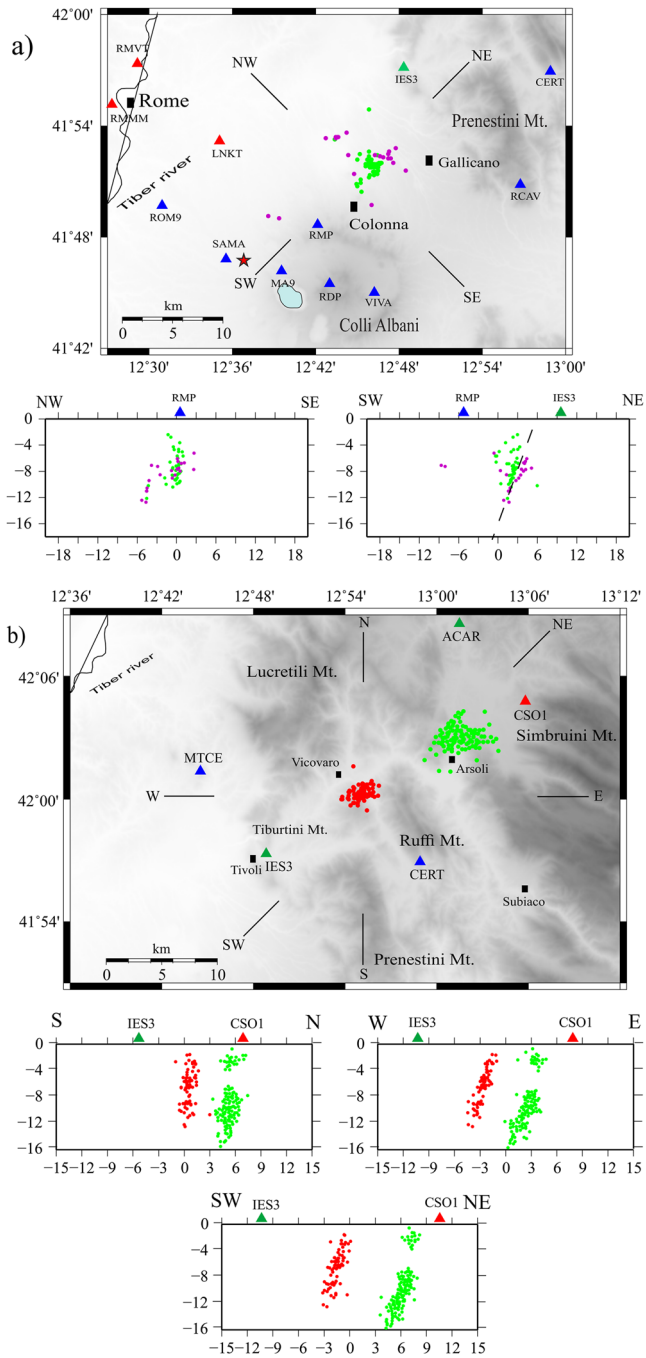

Figure 9. (a) Epicentral distribution of the Colonna-Gallicano clusters and SW-NE and NW-SE vertical cross sections. Purple dots indicate 2009–2011 seismicity; green dots 2019 seismicity. Red star: Cava dei Selci gas discharge; light blue ellipse: Lake Albano. (b) Epicentral distribution of the Vicovaro (red dots) and Arsoli (green dots) clusters with S-N, W-E, and SW-NE vertical cross sections.

The total CO_2 output shows relevant time changes (Figure 3a) with the values of the dry season (late spring-summer) on average lower than those of the wet season (late autumn-winter; Figure S6 in Supporting Information S1). Carapezza, Barberi, et al. (2010) attributed this variation to local overexploitation of the water for potable and irrigation uses in the summer with consequent inflow of new CO_2-undersaturated water. It has to be noted that soil CO_2 fluxes measured by the automatic station do not reproduce this feature. This discordance likely depends on the specific site-effect of the flux station: as shown in Table 1, the soil gas of the station site is chemically and isotopically different from the gas freely released at CdS. However, the mean value of the soil CO_2 flux (335 g m^{-2} d^{-1}, Table 2) recorded at the continuous station is over one order of magnitude higher than the local biological background (30 g m^{-2} d^{-1}, Carapezza et al., 2003) and together with the isotopic mark confirms the endogenous origin of the gas.

Anomalous values of the sGs series (higher than the mean $+2\sigma$, Figure 3a) were observed in May 2000 January 2006 and July 2007. May 2000 anomaly belongs to a period of high CO_2 release during which lethal accidents occurred to cows and sheep while watering at the CdS pool, and culminated with the death of a man by asphyxia in December 2000 (Carapezza et al., 2003).

The soil CO_2 flux maps show a typical NW-SE elongation (Figure 3b) and we argue that the NW-SE trending volcano-tectonic fissure, starting from the Lake Albano and crossing CdS area (Figure 1a), acts as a preferential pathway for the migration of the deep gas to the surface. This can also explain the correlation found between the soil CO_2 flux anomaly and the 2006 Castel Gandolfo earthquake (see Figure 6 for location).

During the last 20 years, similar diffuse soil CO_2 investigations have been conducted in other volcanic or geothermal areas of the Italian territory to estimate the total CO_2 output. The flux data from 15 different areas are reported in Table 3.

Although the number of measurements roughly depends on the extension of the investigated area, the sampling densities are very different from one area to another (Table 3). Despite this fact, the application of the sGs approach for all considered surveys encourages in performing the comparison, since the basic requirement for the application of the sGs methodology is an appropriate sampling design (Cardellini et al., 2003). When multiple surveys were performed in the same area (e.g., CdS, Solfatara, Vulcano), the mean values of total CO_2 output (in t d^{-1}) and soil CO_2 flux normalized by the surveyed area (in kg m^{-2} d^{-1}) are reported. The CdS level of the CO_2 output normalized by area is the highest amongst the 15 considered sites (Table 3). This primarily depends on the fact that CdS surveyed site covers the main gas discharge zone, as also shown by the very scarce proportion (0.07–0.25) of the low-flux populations in Table S6 in Table Supporting Information S1.

5.2. Relationship Between Soil CO_2 Flux and Seismicity

For each of the 55 soil CO_2 flux surveys carried out from February 2000 to August 2020, we extracted from the catalog the seismic events occurred during 30 days straddling the geochemical survey, and within 30 km distance from CdS gas discharge (data set in Frepoli and Pagliuca (2022)). We analyzed number, depth, magnitude of the events (M ranges from 0.3 to 3.9), and the associated focal mechanisms. In Figure 10, seismic data are compared with soil CO_2 output from CdS

Table 3
Comparison of Soil CO_2 Output From 15 Italian Volcanic or Geothermal Areas (in Decreasing Order of Soil CO_2 Flux Per Unit Surface)

Site	Samples (no.) avg. (min.-max.)	Surface (m^2)	Sampling density (No./1,000 m^2)	CO_2 output (t d^{-1})	CO_2 flux (kg m^{-2} d^{-1})
CdS, Colli Albani[1] (55 surveys)	114 (63–142)	5,252	21.7	12.1	**2.323**
Salcheto[2] (Tuscany)	153	20,250	7.6	36	1.778
Solfatara[3], Campi Flegrei (30 surveys)	439 (372–583)	938,000	0.5	1309.4	1.395
Vulcano[4], Levante Beach (2 surveys)	52 (39–66)	17,700	3.0	19	1.095
Bagni San Filippo[5] (Tuscany)	595	1,700,000	0.4	865	0.509
Solforata[6], Colli Albani	365	229,000	1.6	89.2	0.390
Vulcano[7], La Fossa crater (11 surveys)	258 (54–489)	600,000	0.4	228	0.380
Poggio dell'Olivo[8] (Latium)	196	820,000	0.2	253.1	0.309
Caldara di Manziana[2] (Latium)	538	723,968	0.7	162.4	0.224
Lipari[2] (Aeolian Island)	411	31,900	12.9	6.6	0.207
Tor Caldara[6] (Latium)	85	77,000	1.1	11.5	0.149
Latera[10] (Latium)	2,500	3,100,000	0.8	350	0.113
Favare[11], Pantelleria (Sicily)	222	93,000	2.4	7.8	0.084
Vesuvio[12]	502	1,800,000	0.3	138.7	0.077
Ischia-Donna Rachele area[13]	336	862,500	0.4	32.6	0.038

Note. (1) This work; (2) Cardellini et al. (2004); (3) Cardellini et al. (2017); (4) Chiodini et al. (2005); (5) Chiodini et al. (2020); (6) Carapezza et al. (2012); (7) Granieri et al. (2006); (8) Cardellini et al. (2003), (9) Costa et al. (2008); (10) Chiodini et al. (2007); (11) Granieri et al. (2014); (12) Granieri et al. (2013); (13) Chiodini et al. (2004).

target area measured in 2000–2020 and with sGBT residual values of soil CO_2 flux continuously recorded at CdS station in 2004–2008.

The maximum magnitude (M_w = 3.9, depth = 9.2 km) was recorded on 12 April 2008 during a small seismic sequence occurred in the Ciampino area near CdS, in a period characterized by medium CO_2 flux values (12.1 t d^{-1}). This sequence shows essentially strike-slip solutions with cluster hypocenters mainly between 7.5 and 11 km depth clearly N-S aligned, displaying a subvertical dip (Frepoli et al., 2010).

Conversely, high total soil CO_2 output from the CdS target area (exceeding the mean $+1\sigma$) was temporarily associated to the seismic sequence of Monte Porzio (January-March 2006), which included 16 events with mostly normal focal mechanism, few strike-slip solutions and clustered hypocenters. The earthquake depth was between 11 and 14 km; the main shock occurred on 4 January 2006 with M_L = 2.7 at 14.72 km depth.

A particular case is that of June 2019. On 11 June the total soil CO_2 output had been estimated to 11.8 t d^{-1}. On 23 June an earthquake occurred near Colonna, 15 km NE of CdS, with M_w = 3.6 and depth of 5 km, which was followed by a few earthquakes with M_L = 1.4. On the following day (24 June) a new survey found a significant increase of the total CO_2 output (14.7 t d^{-1}).

Comparing these cases, we could conclude that soil CO_2 flux is especially related to tectonic settings and earthquake depth: shallow hypocenters and strike-slip faults are associated to low soil CO_2 flux; normal faults and deep hypocenters are associated to high CO_2 flux values. Unfortunately, this is not always true, probably because fault systems are often characterized by a complex geometry of fractures.

Considering the earthquakes of the Ciampino area, we have low CO_2 flux (6.1 t d^{-1}) on 31 October 2003, shortly after the Ciampino seismic swarm occurred on 10–22 October with a depth of 5–13 km (main shock on 22 October, Md = 2.9, depth = 10 km). On 22 August 2005, a medium-high CO_2 flux (15.7 t d^{-1}) was preceded on 21 August, by the M_L = 1.6 Ciampino earthquake, with 13 km depth.

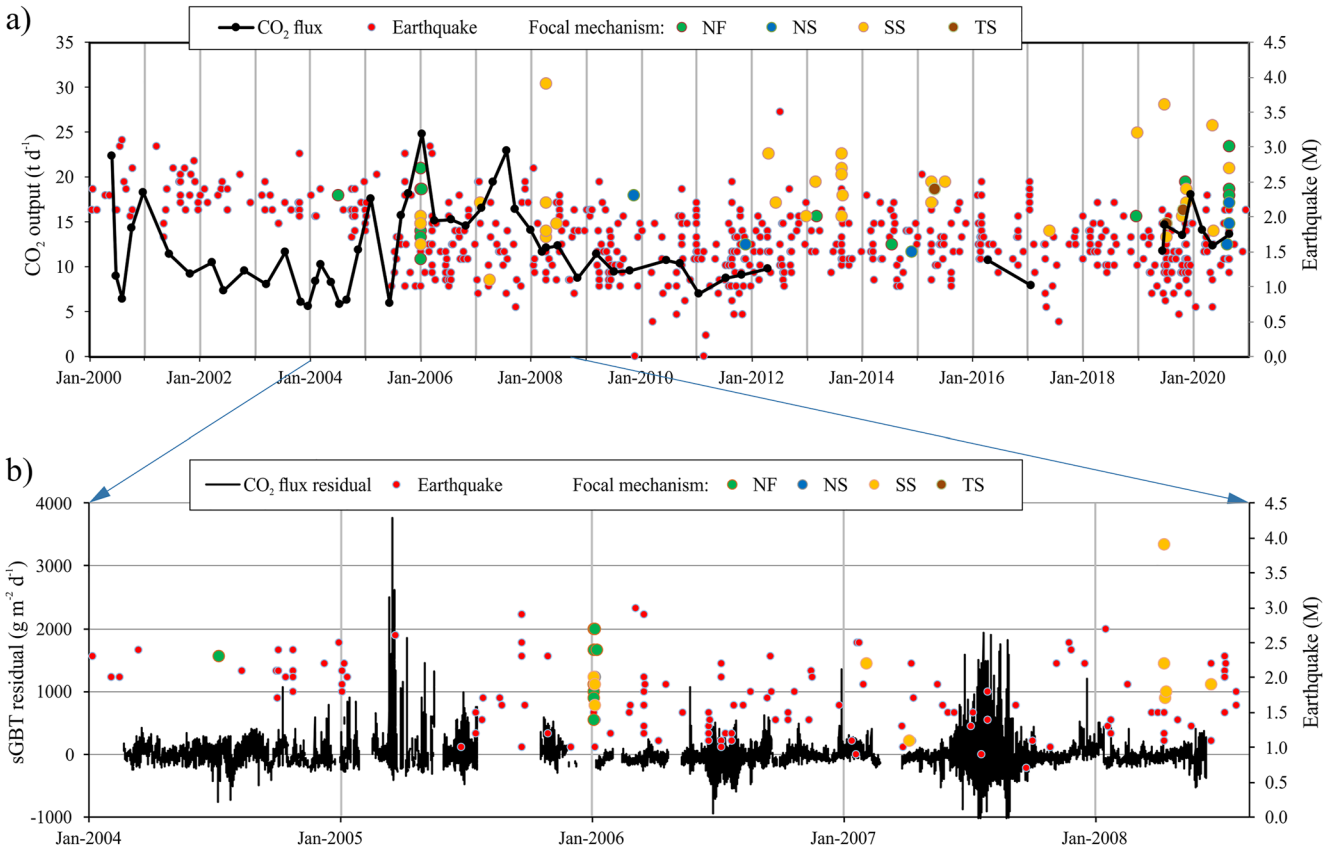

Figure 10. Comparison of soil CO_2 flux with local seismicity. (a) Total CO_2 output from Cava dei Selci (CdS) target area estimated by the 2000–2020 surveys. (b) Residual values of soil CO_2 flux continuously recorded in 2004–2008 obtained by filtering with sGBT method (see Figure 5a″). Focal mechanism of earthquakes: NF = Normal Faulting, NS = Transtension, SS = Strike-Slip Faulting, TS = Transpression (see Table S7 in Supporting Information S1).

It has to be noted that high CO_2 fluxes have been measured also in absence of earthquakes: on 31 May 2000 (22.4 t d^{-1}), on 19 December 2000 (18.3 t d^{-1}), on 3 February 2005 (17.6 t d^{-1}), on 25 July 2007 (22.9 t d^{-1}), on 26 April 2007 (19.4 t d^{-1}) and on 27 February 2020 (14.1 t d^{-1}).

No correlation was found with earthquakes of the southern Colli Albani area that occurred mostly in periods of low-medium CO_2 output (≤10 t d^{-1}). The only exceptions are: (a) the medium-high CO_2 flux measured on 28 August 2020 (13.6 t d^{-1}), the same day of Lariano seismic sequence (16 events on 28 and 29 August 2020, main shock M_L = 3.0 and cluster hypocenters with depth between 11 and 14 km) and (b) the medium-high CO_2 flux measured on 31 October 2006 (14.5 t d^{-1}), shortly after the 23 October 2006 Castel Gandolfo earthquake (M_L = 1.8, depth 9 km).

In conclusion, the relation between soil CO_2 flux and the distribution of seismicity in the Colli Albani area is complex. The different behavior between CO_2 output and the distribution of seismicity in the various districts could be due to the different stress regimes and probably depends on the low magnitude of the earthquakes recorded in the area ($M_w \leq 3.9$).

6. Conclusions

Twenty years monitoring of soil CO_2 flux at CdS provides a solid geochemical background for future early predictions of volcanic unrest of the Colli Albani. We have shown that sGs—stochastic Gaussian simulation—method is the most appropriate to process data of soil CO_2 flux surveys. Equally the sGBT—stochastic Gradient Boosting Trees regression proved to be the most effective to sift through the temporal soil CO_2 flux data and environmental parameters data set with the aim of recognizing anomalies related to endogenous processes. The sGBT regression could thus be reliably applied to the data sets of other automatic soil CO_2 flux stations operating for volcanic

surveillance in Italy and worldwide, improving the capability of discriminating anomalies related to volcanic processes. An important result has also been obtained by the analysis and relocation of 1,062 earthquakes (575 in this paper) recorded in 1997–2020 within 30 km from CdS. The Colli Albani volcanic complex experiences a moderate seismicity ($Md \leq 4$) and seismologists generally neglect most of the recorded low-magnitude events because the related hazard is very low. We showed here that the analysis of these events contributes to improved knowledge of the active faults intersecting the Colli Albani volcanic complex, some of which also likely act as preferential ascent pathways for the deep originated gas.

Some relations have been found between soil CO_2 flux and regional seismicity; in particular the relationship with deep hypocenters shows that only seismic events originating at least at 10 km depth may stress the deep gas source and increase the gas output at CdS. We have also to consider that no volcanic unrest crisis occurred in the considered time span. It is logical to infer that anomalous seismicity, characteristically associated to a volcanic unrest, will produce a significant increase of soil CO_2 release, as well as a change in the chemical and isotopic composition of the gas emitted at CdS, which remained nearly constant in the last 20 years.

Data Availability Statement

Geochemical data used in this study can be found at: http://hdl.handle.net/2122/14780. Seismic data are available at: http://hdl.handle.net/2122/15558.

Acknowledgments
We are grateful to dr. Graziella Caprarelli, editor of Earth and Space Science, whose comments and suggestions significantly improved the quality of the paper. We thank the colleagues of INGV-Palermo Labs for gas analyses and Mr. Marrocco, the owner of the site where the soil CO_2 flux automatic station is located since 20 years. This study has benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri—Dipartimento della Protezione Civile (DPC). This paper does not necessarily represent DPC official opinion and policies.

References

Acocella, V., & Funiciello, R. (2006). Transverse systems along the extensional Tyrrhenian margin of central Italy and their influence on volcanism. *Tectonics*, *25*, TC2003. https://doi.org/10.1029/2005TC001845

Amato, A., Chiarabba, C., Cocco, M., di Bona, M., & Selvaggi, G. (1994). The 1989–1990 seismic swarm in the Alban Hills volcanic area, central Italy. *Journal of Volcanology and Geothermal Research*, *61*(3–4), 225–237. https://doi.org/10.1016/0377-0273(94)90005-1

Anzidei, M., Carapezza, M. L., Esposito, A., Giordano, G., Lelli, M., & Tarchini, L. (2008). The Albano maar lake high resolution bathymetry and dissolved CO_2 budget (Colli Albani volcano, Italy): Constrains to hazard evaluation. *Journal of Volcanology and Geothermal Research*, *171*(3–4), 258–268. https://doi.org/10.1016/j.jvolgeores.2007.11.024

Badalamenti, B., Capasso, G., Carapezza, M. L., D'Alessandro, W., Di Gangi, F., Diliberto, I. S., et al. (1994). Soil gas investigations during the 1991–1993 Etna eruption. *Acta Vulcanologica*, *4*, 135–141.

Barberi, F., Carapezza, M. L., Ranaldi, M., & Tarchini, L. (2007). Gas blowout from shallow boreholes at Fiumicino (Rome): Induced hazard and evidence of deep CO_2 degassing on the Tyrrhenian margin of central Italy. *Journal of Volcanology and Geothermal Research*, *165*(1–2), 17–31. https://doi.org/10.1016/j.jvolgeores.2007.04.009

Burton, M. R., Sawyer, G. M., & Granieri, D. (2013). Deep carbon emissions from volcanoesThe monitoring of natural soil CO_2 emissions: Issue and perspectives. *Reviews in Mineralogy and Geochemistry*, *75*(1), 323–354. https://doi.org/10.2138/rmg.2013.75.11

Camarda, M., De Gregorio, S., Capasso, G., Di Martino, R. M. R., Gurrieri, S., & Prano, V. (2019). The monitoring of natural soil CO_2 emissions: Issue and perspectives. *Earth Science Review*, *198*, 102928.

Carapezza, M. L., Badalamenti, B., Cavarra, L., & Scalzo, A. (2003). Gas hazard assessment in a densely inhabited area of Colli Albani Volcano (Cava dei Selci, Roma). *Journal of Volcanology and Geothermal Research*, *123*(1–2), 81–94. https://doi.org/10.1016/s0377-0273(03)00029-5

Carapezza, M. L., Barberi, F., Ranaldi, M., Ricci, T., Tarchini, L., Barrancos, J., et al. (2012). Hazardous gas emissions from the flanks of the quiescent Colli Albani volcano (Rome, Italy). *Applied Geochemistry*, *27*(9), 1767–1782. https://doi.org/10.1016/j.apgeochem.2012.02.012

Carapezza, M. L., Barberi, F., Ranaldi, M., Tarchini, L., & Pagliuca, N. M. (2019). Faulting and gas discharge in the Rome area (Central Italy) and associated hazards. *Tectonics*, *38*(3), 941–959. https://doi.org/10.1029/2018TC005247

Carapezza, M. L., Barberi, F., Tarchini, L., & Ranaldi, M. (2021). Hazard from endogenous gas emissions and phreatic explosions in Rome City. *Geophysical Research Letters*, *48*, e2020GL089797. https://doi.org/10.1029/2020GL089797

Carapezza, M. L., Barberi, F., Tarchini, L., Ranaldi, M., & Ricci, T. (2010). Volcanic hazards of the Colli Albani. In Funiciello, R., & Giordano, G. (Eds.), *The Colli Albani volcano, Special Publications of IAVCEI* (Vol. 3, pp. 279–297). Geological Society of London.

Carapezza, M. L., & Diliberto, I. S. (1994). Helium and CO_2 soil gas concentration (Vulcano). *Acta Vulcanologica*, *6*, 36–39.

Carapezza, M. L., & Federico, C. (2000). The contribution of fluid geochemistry to the volcano monitoring of Stromboli. *Journal of Volcanology and Geothermal Research*, *95*, 227–245. https://doi.org/10.1016/s0377-0273(99)00128-6

Carapezza, M. L., Inguaggiato, S., Brusca, L., & Longo, M. (2004). Geochemical precursors of the activity of an open conduit volcano: The Stromboli 2002–2003 eruptive events. *Geophysical Research Letters*, *31*, L07620. https://doi.org/10.1029/2004GL019614

Carapezza, M. L., Ranaldi, M., Gattuso, A., Pagliuca, N. M., & Tarchini, L. (2015). The sealing capacity of the cap rock above the Torre Alfina geothermal reservoir (Central Italy) revealed by soil CO_2 flux investigations. *Journal of Volcanology and Geothermal Research*, *291*, 25–34. https://doi.org/10.1016/j.jvolgeores.2014.12.011

Carapezza, M. L., Ranaldi, M., Tarchini, L., Gattuso, A., Pagliuca, N. M., Vinci, M., & Barberi, F. (2020). Dangerous emissions of endogenous CO_2 and H_2S from gas blowouts of shallow wells in the Rome Municipality (Italy). *Applied Geochemistry*, *123*, 104769. https://doi.org/10.1016/j.apgeochem.2020.104769

Carapezza, M. L., Ricci, T., Barberi, F., Ranaldi, M., & Tarchini, L. (2010). *Hazardous gas blowouts from shallow wells in the Colli Albani volcanic complex (Rome, Italy)* (pp. 913–916). Water-Rock Interaction.

Carapezza, M. L., Ricci, T., Ranaldi, M., & Tarchini, L. (2009). Active degassing structures of Stromboli and variations in diffuse CO_2 output related to the volcanic activity. *Journal of Volcanology and Geothermal Research*, *182*(3–4), 231–245. https://doi.org/10.1016/j.jvolgeores.2008.08.006

Carapezza, M. L., & Tarchini, L. (2007). Accidental gas emission from shallow pressurized aquifers at Alban Hills volcano (Rome, Italy): Geochemical evidence of magmatic degassing? *Journal of Volcanology and Geothermal Research*, *165*(1–2), 5–16. https://doi.org/10.1016/j.jvolgeores.2007.04.008

Cardellini, C., Chiodini, G., & Frondini, F. (2003). Application of stochastic simulation to CO_2 flux from soil: Mapping and quantification of gas release. *Journal of Geophysical Research*, *108*(B9), 2425. https://doi.org/10.1029/2002JB002165

Cardellini, C., Chiodini, G., Frondini, F., Avino, R., Bagnato, E., Caliro, S., et al. (2017). Monitoring diffuse volcanic degassing during volcanic unrests: The case of Campi Flegrei (Italy). *Scientific Reports*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-06941-2

Cardellini, C., Frondini, F., Morgantini, N., Chiodini, G., Avino, R., Caliro, S., et al. (2004). Diffuse soil CO_2 degassing at Salcheto (Tuscany) and Lipary (Sicily): A probabilistic approach to mapping and quantification of gas release. In *Water-rock interaction, two volume set: Proceedings of the eleventh international symposium on water-rock interaction* (Vol. 27, p. 101). CRC Press.

Cesi, C., Eulilli, V., & Ferri, F. (2008). Analisi ed interpretazione dei valori delle anomalie di gravità del territorio dell'area romana: Correlazione con gli elementi geologici di superficie e la struttura profonda. *La Geologia di Roma. Memorie Descrittive della Carta Geologica d'Italia*, *80*, 97–114.

Chatelain, J. L. (1978). *Etude fine de la sismicité en zone de collision continentale à l'aide d'un réseau de stations portables: La region Hindu-Kush-Pamir, thèse de 3 éme cycle*. Univ. Paul Sabatier.

Chiodini, G., Avino, R., Brombach, T., Caliro, S., Cardellini, C., De Vita, S., et al. (2004). Fumarolic and diffuse soil degassing west of Mount Epomeo, Ischia, Italy. *Journal of Volcanology and Geothermal Research*, *133*(1–4), 291–309. https://doi.org/10.1016/s0377-0273(03)00403-7

Chiodini, G., Baldini, A., Barberi, F., Carapezza, M. L., Cardellini, C., Frondini, F., et al. (2007). Carbon dioxide degassing at Latera caldera (Italy): Evidence of geothermal reservoir and evaluation of its potential energy. *Journal of Geophysical Research*, *112*,B12204. https://doi.org/10.1029/2006JB004896

Chiodini, G., Caliro, S., Cardellini, C., Granieri, D., Avino, R., Baldini, A., & Minopoli, C. (2010). Long-term variations of the Campi Flegrei, Italy, volcanic system as revealed by the monitoring of hydrothermal activity. *Journal of Geophysical Research*, *115*, B03205. https://doi.org/10.1029/2008JB006258

Chiodini, G., Cardellini, C., Caliro, S., Avino, R., Donnini, M., Granieri, D., et al. (2020). The hydrothermal system of Bagni san Filippo (Italy): Fluids circulation and CO_2 degassing. *Italian Journal of Geosciences*, *139*(3), 383–397.

Chiodini, G., Cioni, R., Guidi, M., Marini, L., & Raco, B. (1998). Soil CO_2 flux measurements in volcanic and geothermal areas. *Applied Geochemistry*, *13*, 543–552. https://doi.org/10.1016/s0883-2927(97)00076-0

Chiodini, G., Granieri, D., Avino, R., Caliro, S., Costa, A., & Werner, C. (2005). Carbon dioxide diffuse degassing and estimation of heat release from volcanic and hydrothermal systems. *Journal of Geophysical Research*, *110*, B08204. https://doi.org/10.1029/2004JB003542

Cinti, D., Procesi, M., Tassi, F., Montegrossi, G., Sciarra, A., Vaselli, O., & Quattrocchi, F. (2011). Fluid geochemistry and geothermometry in the Western sector of the Sabatini volcanic district and the Tolfa Mountains (central Italy). *Chemical Geology*, *284*(1–2), 160–181. https://doi.org/10.1016/j.chemgeo.2011.02.017

Costa, A., Chiodini, G., Granieri, D., Folch, A., Hankin, R. K. S., Caliro, S., et al. (2008). A shallow layer model for heavy gas dispersion from natural sources: Application and hazard assessment at Caldara di Manziana, Italy. *Geochemistry, Geophysics, Geosystems*, *9*, Q03002. https://doi.org/10.1029/2007GC001762

De Bortoli Teixeira, D., Panosso, A. R., Cerri, P. C. E., Pereira, G. T., & La Scala, N., Jr. (2011). Soil CO_2 emission estimated by different interpolation techniques. *Plant and Soil*, *345*, 187–194. https://doi.org/10.1007/s11104-011-0770-6

De Gregorio, S., & Camarda, M. (2016). A novel approach to estimate the eruptive potential and probability in open conduit volcanoes. *Scientific Report*, *6*, 3047. https://doi.org/10.1038/srep30471

De Luca, G., Cattaneo, M., Monachesi, G., & Amato, A. (2009). Seismicity in Central and Northern Apennines integrating the Italian national and regional networks. *Tectonophysics*, *476*, 121–135. https://doi.org/10.1016/j.tecto.2008.11.032

Deutsch, C. V., & Journel, A. G. (1992). Geostatistical software library and user's guide (Vol. *119*, pp. 147). New York: Oxford University Press.

Di Martino, R. M. R., Capasso, G., Camarda, M., De Gregorio, S., & Prano, V. (2020). Deep CO_2 release revealed by stable isotope and diffuse degassing surveys at Vulcano (Aeolian Islands) in 2015–2018. *Journal of Volcanology and Geothermal Research*, *401*, 1–14. https://doi.org/10.1016/j.jvolgeores.2020.106972

Diliberto, I., Gurrieri, S., & Valenza, M. (2002). Relationships between diffuse CO_2 emissions and volcanic activity on the island of Vulcano (Aeolian Islands, Italy) during the period 1984–1994. *Bulletin of Volcanology*, *64*(3), 219–228. https://doi.org/10.1007/s00445-001-0198-6

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Federico, C., Brusca, L., Carapezza, M. L., Cigolini, C., Inguaggiato, S., Rizzo, A., & Rouwet, D. (2008). *Geochemical prediction of the 2002–2003 Stromboli eruption from variations in CO_2 and Rn emissions and in helium and carbon isotopes*. *Geophysical Monograph Series* (Vol. 182, pp. 117–128). American Geophysical Union.

Fornaseri, M., & Turi, B. (1969). Carbon and oxygen isotopic composition of carbonates in lavas and ejectites from the Alban Hills, Italy. *Contributions to Mineralogy and Petrology*, *23*(3), 244–256. https://doi.org/10.1007/bf00371535

Frepoli, A., Cimini, G. B., De Gori, P., De Luca, G., Marchetti, A., Monna, S., et al. (2017). Seismic sequences and swarms in the Latium-Abruzzo-Molise Apennines (central Italy): New observations and analysis from a dense monitoring of the recent activity. *Tectonophysics*, *712*, 312–329. https://doi.org/10.1016/j.tecto.2017.05.026

Frepoli, A., Marra, F., Maggi, C., Marchetti, A., Nardi, A., Pagliuca, N. M., & Pirro, M. (2010). Seismicity, seismogenic structures, and crustal stress fields in the greater Rome area (central Italy). *Journal of Geophysical Research*, *115*, B12303. https://doi.org/10.1029/2009JB006322

Frepoli, A., & Pagliuca, N. M. (2022). *Dataset on seismic data recorded at Colli Albani (Rome) in 2000–2020*EarthPrints. Retrieved from http://hdl.handle.net/2122/15558

Friedman, J. (1999). *Greedy function approximation: A stochastic boosting machine (Tech. Rep.)*. Department of Statistics, Stanford University.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics*, *28*(2), 337–407. https://doi.org/10.1214/aos/1016218223

Funiciello, R., Heiken, G., De Benedetti, A. A., & Giordano, G. (2010). Volcanic activity of the lake Albano maar in Roman history. In *The Colli Albani Volcano*, (Vol. 331–338). The Geological Society of London.

Gerlach, T. M., & Graeber, E. J. (1985). Volatile budget of Kilauea volcano. *Nature*, *313*, 273–277. https://doi.org/10.1038/313273a0

Ghojogh, B., & Crowley, M. (2019). *The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial*. arXiv preprint arXiv:1905.12787

Giordano, G., De Benedetti, A. A., Bonamico, A., Ramazzotti, P., & Mattei, M. (2014). Incorporating surface indicators of reservoir permeability into reservoir volume calculations: Application to the Colli Albani caldera and the Central Italy Geothermal Province. *Earth-Science Reviews*, *128*, 75–92. https://doi.org/10.1016/j.earscirev.2013.10.010

Granieri, D., Avino, R., & Chiodini, G. (2010). Carbon dioxide diffuse emission from the soil: Ten years of observations at Vesuvio and Campi Flegrei (Pozzuoli), and linkages with volcanic activity. *Bulletin of Volcanology*, *72*(1), 103–118. https://doi.org/10.1007/s00445-009-0304-8

Granieri, D., Carapezza, M. L., Avino, R., Caliro, S., Cardellini, C., Chiodini, G., et al. (2013). Level of carbon dioxide diffuse degassing from the ground of Vesuvio: Comparison between extensive surveys and inferences on the gas source. *Annals of Geophysics*, *56*(4), 0449. https://doi.org/10.4401/ag-6455

Granieri, D., Carapezza, M. L., Chiodini, G., Avino, R., Caliro, S., Ranaldi, M., et al. (2006). Correlated increase in CO_2 fumarolic content and diffuse emission from La Fossa crater (Vulcano, Italy): Evidence of volcanic unrest or increasing gas release from a stationary deep magma body? *Geophysical Research Letters*, *33*, L13316. https://doi.org/10.1029/2006GL026460

Granieri, D., Chiodini, G., Avino, R., & Caliro, S. (2014). Carbon dioxide emission and heat release estimation for Pantelleria Island (Sicily, Italy). *Journal of Volcanology and Geothermal Research*, *275*, 22–33. https://doi.org/10.1016/j.jvolgeores.2014.02.011

Granieri, D., Chiodini, G., Marzocchi, W., & Avino, R. (2003). Continuous monitoring of CO_2 soil diffuse degassing at Phlegraean fields (Italy): Influence of environmental and volcanic parameters. *Earth and Planetary Science Letters*, *212*(1–2), 167–179. https://doi.org/10.1016/s0012-821x(03)00232-2

Holland, J. H., & Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press.

INGV Seismological Data Centre. (2006). *Rete Sismica Nazionale (RSN)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/SD/X0FXNH7QFY

ISIDe Working Group. (2007). *Italian Seismological Instrumental and Parametric Database (ISIDe)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/ISIDE

Laiolo, M., Ranaldi, M., Tarchini, L., Carapezza, M. L., Coppola, D., Ricci, T., & Cigolini, C. (2016). The effects of environmental parameters on diffuse degassing at Stromboli volcano: Insights from joint monitoring of soil CO_2 flux and radon activity. *Journal of Volcanology and Geothermal Research*, *315*, 65–78. https://doi.org/10.1016/j.jvolgeores.2016.02.004

Lewicki, J. L., Bergfeld, D., Cardellini, C., Chiodini, G., Granieri, D., Varley, N., & Werner, C. (2005). Comparative soil CO_2 flux measurements and geostatistical estimation methods on Masaya volcano, Nicaragua. *Bulletin of Volcanology*, *68*, 76–90. https://doi.org/10.1007/s00445-005-0423-9

Liuzzo, M., Gurrieri, S., Giudice, G., & Giuffrida, G. (2013). Ten years of soil CO_2 continuous monitoring on Mt. Etna: Exploring the relationship between processes of soil degassing and volcanic activity. *Geochemistry, Geophysics, Geosystems*, *14*, 2886–2899. https://doi.org/10.1002/ggge.20196

Marra, F. (2001). Strike-slip faulting and block rotation: A possible triggering mechanism for lava flows in the Albani Hills? *Journal of Structural Geology*, *23*(1), 127–141. https://doi.org/10.1016/s0191-8141(00)00068-7

Marra, F., Gaeta, M., Giaccio, B., Jicha, B. R., Palladino, D. M., Polcari, M., et al. (2016). Assessing the volcanic hazard for Rome: 40Ar/39Ar and InSAR constraints on the most recent eruptive activity and present day uplift at Colli Albani volcanic district. *Geophysical Research Letters*, *43*, 6898–6906. https://doi.org/10.1002/2016GL069518

Martelli, M., Nuccio, P. M., Stuart, F. M., Burgess, R., Ellam, R. M., & Italiano, F. (2004). Helium-Strontium isotope constraints on mantle evolution beneath the Roman Comagmatic Province, Italy. *Earth and Planetary Science Letters*, *224*, 295–308. https://doi.org/10.1016/j.epsl.2004.05.025

Mazza, R., & Capelli, G. (2010). Hydrogeology of the Colli Albani volcano. In *The Colli Albani Volcano, Special Publications of IAVCEI* (Vol. 3, pp. 189–213). Geological Society of London.

Oliveira, S., Viveiros, F., Silva, C., & Pacheco, J. E. (2018). Automatic filtering of soil CO_2 flux data; Different statistical approaches applied to long time series. *Frontiers of Earth Science*, *6*, 208. https://doi.org/10.3389/feart.2018.00208

Paonita, A., Caracausi, A., Marziano, G. I., Martelli, M., & Rizzo, A. (2012). Geochemical evidence for mixing between fluids exsolved at different depths in the magmatic system of Mt Etna (Italy). *Geochimica et Cosmochimica Acta*, *84*, 380–394. https://doi.org/10.1016/j.gca.2012.01.028

Papale, P., Moretti, R., & Barbato, D. (2006). The compositional dependence of the saturation surface of $H_2O + CO_2$ fluids in silicate melts. *Chemical Geology*, *229*, 78–95. https://doi.org/10.1016/j.chemgeo.2006.01.013

Park, Y., & Ho, J. C. (2019). Tackling overfitting in boosting for noisy healthcare data. *IEEE Transactions on Knowledge and Data Engineering*, *33*(7), 2995–3006.

Ranaldi, M., Lelli, M., Tarchini, L., Carapezza, M. L., & Patera, A. (2016). Estimation of the geothermal potential of the Caldara di Manziana site in the Sabatini Volcanic District (central Italy) by integrating geochemical data and 3D-GIS modelling. *Geothermics*, *62*, 115–130. https://doi.org/10.1016/j.geothermics.2016.04.003

Reasenberg, P., & Oppenheimer, D. H. (1985). FPFIT, FPPLOT and FPPAGE: FORTRAN computer programs for calculating and displaying earthquake fault-plane solutions. *U.S. Geological Survey Open-File Report*, *85*, 739. https://doi.org/10.3133/ofr85739

Rey, A., Pegoraro, E., Tedeschi, V., De Parri, I., Jarvis, P. G., & Valentini, R. (2002). Annual variation in soil respiration and its components in a coppice oak forest in Central Italy. *Global Change Biology*, *8*(9), 851–866. https://doi.org/10.1046/j.1365-2486.2002.00521.x

Sambridge, M., & Gallagher, K. (1993). Earthquake hypocenter location using genetic algorithms. *Bulletin of the Seismological Society of America*, *83*(5), 1467–1491.

Sano, Y., & Marty, B. (1995). Origin of carbon in fumarolic gas from island arcs. *Chemical Geology*, *119*(1–4), 265–274. https://doi.org/10.1016/0009-2541(94)00097-r

Sigvaldason, G. E. (1989). International conference on lake Nyos disaster; Yaoundè, Cameroun 16–20 March, 1987. Conclusion and raccomendation. *Journal of Volcanology and Geothermal Research*, *39*, 97–107. https://doi.org/10.1016/0377-0273(89)90050-4

Tarchini, L., Carapezza, M. L., Granieri, D., & Ranaldi, M. (2021). *Dataset of soil CO_2 flux surveys and geostatistical results from 2000 to 2020 (Cava dei Selci, Colli Albani volcano, Italy)*EarthPrints. Retrieved from http://hdl.handle.net/2122/14780

Tarchini, L., Carapezza, M. L., Ranaldi, M., Sortino, F., Gattuso, A., & Acocella, V. (2019). Fluid geochemistry contribution to the interpretation of the 2011–2012 unrest of Santorini, Greece, in the frame of the dynamics of the Aegean Volcanic Arc. *Tectonics*, *38*(3), 1033–1049. https://doi.org/10.1029/2018TC005377

Teixeira, D. D. B., Panosso, A. R., Cerri, C. E. P., Pereira, G. T., & La Scala, N. (2011). Soil CO_2 emission estimated by different interpolation techniques. *Plant and Soil*, *345*(1), 187–194. https://doi.org/10.1007/s11104-011-0770-6

Trasatti, E., Marra, F., Polcari, M., Etiope, G., Ciotoli, G., Darrah, T. H., et al. (2018). Coeval uplift and subsidence reveal magma recharging near Rome (Italy). *Geochemistry, Geophysics, Geosystems*, *19*, 1484–1498. https://doi.org/10.1029/2017GC007303

Venturi, S., Tassi, F., Cabassi, J., Vaselli, O., Minardi, I., Neri, S., et al. (2019). A multi-instrumental geochemical approach to assess the environmental impact of CO_2-rich gas emissions in a densely populated area: The case of Cava dei Selci (Latium, Italy). *Applied Geochemistry*, *101*, 109–126. https://doi.org/10.1016/j.apgeochem.2019.01.003

Viveiros, F., Ferreira, T., Silva, C., & Gaspar, J. L. (2009). Meteorological factors controlling soil gases and indoor CO_2 concentration: A permanent risk in degassing areas. *The Science of the Total Environment*, *407*(4), 1362–1372. https://doi.org/10.1016/j.scitotenv.2008.10.009

Viveiros, F., Ferreira, T., Silva, C., Vieira, J. C., Gaspar, J. L., Virgili, G., & Amaral, P. (2015). Permanent monitoring of soil CO_2 degassing at Furnas and Fogo volcanoes (Sao Miguel island, Azores). *Geological Society, London, Memoirs*, *44*(1), 271–288. https://doi.org/10.1144/m44.20

Viveiros, F., Ferreira, T., Vieira, J. C., Silva, C., & Gaspar, J. L. (2008). Environmental influences on soil CO_2 degassing at Furnas and Fogo volcanoes (São Miguel island, Azores archipelago). *Journal of Volcanology and Geothermal Research*, *177*(4), 883–893. https://doi.org/10.1016/j.jvolgeores.2008.07.005

Viveiros, F., Vandemeulebrouck, J., Rinaldi, A. P., Ferreira, T., Silva, C., & Cruz, J. V. (2014). Periodic behavior of soil CO_2 emissions in diffuse degassing areas of the Azores archipelago: Application to seismovolcanic monitoring. *Journal of Geophysical Research: Solid Earth*, *119*, 7578–7597. https://doi.org/10.1002/2014JB011118

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications*. Springer Science & Business Media.

Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America*, *90*(6), 1353–1368. https://doi.org/10.1785/0120000006

Washington, H. S. (1906). *The Roman comagmatic region (No. 57)*. Carnegie Institution of Washington.

Zoback, M. L. (1992). First- and second-order patterns of stress in the lithosphere: The World stress map Project. *Journal of Geophysical Research*, *97*(B8), 11703–11728. https://doi.org/10.1029/92JB00132