

# The SoilExp software: An open-source Graphical User Interface (GUI) for post-processing spatial and temporal soil surveys

G. Boudoire, M. Liuzzo, S. Cappuzzo, G. Giuffrida, P. Cosenza, A. Derrien, E.E. Falcone

▶ **To cite this version:**

G. Boudoire, M. Liuzzo, S. Cappuzzo, G. Giuffrida, P. Cosenza, et al.. The SoilExp software: An open-source Graphical User Interface (GUI) for post-processing spatial and temporal soil surveys. Computers & Geosciences, 2020, 142, pp.104553. 10.1016/j.cageo.2020.104553 . hal-02901361

**HAL Id: hal-02901361**
**https://uca.hal.science/hal-02901361**

Submitted on 4 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The SoilExp software: an open-source Graphical User Interface (GUI) for post-processing spatial and temporal soil surveys

**G. BOUDOIRE[1,2]***, M. LIUZZO[1], S. CAPPUZZO[1], G. GIUFFRIDA[1], P. COSENZA[1], A. DERRIEN[3], E.E. FALCONE[1]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Palermo, Via Ugo La Malfa 153, 90146 Palermo, Italy

[2]Université Clermont Auvergne, CNRS, IRD, OPGC, Laboratoire Magmas et Volcans, 6 avenue Blaise Pascal, 63178 Aubière, France

[3]Observatoire Volcanologique du Piton de la Fournaise (OVPF), Institut de Physique du Globe de Paris (IPGP), Sorbonne Paris-Cité, UMR 7154 CNRS, Université Paris Diderot, Bourg Murat, France

* Corresponding author. Present address: Laboratoire Magmas et Volcans – 6 avenue Blaise Pascal – 63170 Aubière (France). Telephone: +33 07 85 22 19 11. E-mail: guillaume.boudoire@uca.fr

G. Boudoire has developed the SoilExp software, and performed field and laboratory tests. M. Liuzzo has conceived and designed the MEGA instrument and performed field and laboratory tests. S. Cappuzzo has conceived and designed the MEGA instrument. G. Giuffrida has performed field and laboratory tests. P. Cosenza has contributed to the design of MEGA instrument. A. Derrien has performed field tests and E.E. Falcone contributed to the writing of the manuscript.

# Abstract

Preliminary interpretation of geological processes during field measurement campaigns require fast data analysis to adapt ongoing target strategies. It is the case of soil investigations where coupling geochemical and geophysical records favor a better understanding of subsurface processes. This task requires (i) statistical analysis ~~is needed~~ to identify areas of interest during spatial surveys and (ii) signal processing ~~is required~~ to analyze temporal series.

Here we present SoilExp, an open-source Python-based Graphical User Interface (GUI) that permits to process spatial and temporal surveys of soil gases (e.g. soil $CO_2$ flux) combined with common physical parameters (e.g. self-potential, temperature) that are synchronously recorded on the field. SoilExp mixes innovative algorithms with the more common tools used for the analysis of both spatial surveys or temporal series. It offers the possibility to display distribution plots, maps, comparative plots, spectra and spectrograms, as well as data statistical analysis, in order to deal efficiently with datasets acquired on the field. Field measurements performed at Stromboli (Italy) supports that such software solution facilitates a quick visualization of the data output and is a powerful tool on the geochemical and geophysical analysis.



# Keywords

Geo-spatial survey, time series, soil $CO_2$ degassing, self-potential, Python


# 1. Introduction

Identifying hidden geologic structures and studying gas and hydrothermal fluid circulation within the ground is of first interest in many disciplines as agriculture (Kucera & Kirkham, 1971), mineral resources (Hinkle & Dilbert, 1984; Lovell *et al.*,

1983), geothermy (Chiodini *et al.*, 2001, 2005), geological storage (Sandig *et al.*, 2014) and natural hazards (Allard *et al.*, 1991; Finizola *et al.*, 2002; Hernandez *et al.*, 2001; Irwin & Barnes, 1980). Coupling geochemical and geophysical records has demonstrated a real complementarity to characterize soil heterogeneities and related fluid circulations (Aubert *et al.*, 1984; Boudoire *et al.*, 2018; Elskens *et al.*, 1964; Finizola et *al.*, 2003; Gaudin *et al.*, 2015; Giammanco *et al.*, 1997). In particular, diffusive $CO_2$ degassing ($CO_2$), self-potential (*SP*) and temperature (*T*) measurements are among the most common methods used by the scientific and industrial community to perform both spatial surveys or temporal records for monitoring purposes (Boudoire *et al.*, 2018; Byrdina *et al.*, 2012; Finizola *et al.*, 2003; Gresse *et al.*, 2016; Pearson *et al.*, 2008).

These measurements often require (i) the use of self-alone instruments and (ii) a preliminary data treatment to be used reliably. For instance, (i) measurements of diffusive $CO_2$ degassing ($CO_2$) may require the use of a stainless steel probe (active method) or an accumulation chamber (passive method) connected to infrared spectrometers, self-potential may require the use of non-polarizable $Cu/CuSO_4$ electrodes coupled with a high impedance voltmeter and temperature (*T*) measurements may be performed with K-type thermal probes and a digital thermometer or with a pyrometer (Finizola *et al.*, 2010). Additionally (ii), spatial surveys often need a quick first idea of results during the daily performed acquisitions in order to identify the main areas of interests and eventually adapt or correct the ongoing fieldwork strategy (Chatterjee *et al.*, 2019). Meanwhile, temporal series are often subjected to an environmental influence that needs to be corrected before an accurate use of the signals as regression, moving average or cut-band filter for the most common ones (Boudoire *et al.*, 2017a; Liuzzo *et al.*, 2013; Padron *et al.*, 2008; Viveiros *et al.*, 2008). Many industrial software packages or homemade codes are able to deal efficiently with this

kind of data but often required to be used additionally to cover the whole range of expected common data treatment tools (with various file formatting). It is often time consuming and limit a fast and efficient evaluation of the datasets.

Here we present a new user-friendly Python-based GUI (Graphical User Interface) software: Soil Exploration (SoilExp). SoilExp is able to analyze both spatial and temporal datasets obtained on the field and respecting some file formatting rules. The final aim of SoilExp is to provide to the geologic-environmental researchers community both innovative and classical tools for a first data processing: (i) data correction (linear regression, moving average, cut-band filter), (ii) data analysis (statistical analysis, populations identification), (iii) data comparison (correlations, cross-correlations) and, (iv) graphical representation (distribution plots, comparative plots, spectra, spectrograms, maps). To illustrate the potentiality of SoilExp to sustain field surveys and to address scientific issues, both soil $CO_2$ flux and self-potential measurements were performed at Stromboli (Italy). Results are presented in a final section and discussed with respect to those obtained from previous field surveys.

## 2. Overview on the SoilExp software

The SoilExp 1.0 software distribution is written in Python 2.7 (Fig. 1). The Graphical User Interface (GUI) is based on the Tkinter library. It requires the following libraries: Pandas, Numpy, SciPy, Matplotlib, Scikit-Learn, PySerial is required. Thus, as processed during SoilExp 1.0 development, we recommend to the user to install the Anaconda distribution on their machines in order to benefit of the Spyder open source cross-platform integrated development environment (IDE) with scientific libraries. Full details are provided in the user guide. Information related to the installation of the software distribution (SoilExp 1.0), to its step-by-step use and to potential script modifications are reported in the associated user manual.

Indeed, in this study, we focus on the main functionalities provided by the SoilExp 1.0 distribution. These main functionalities are exposed through three independent scripts described in the following parts (Fig. 2). The first script is dedicated to save/reset field data from the MEGA (Multisensors Electrical and Gas Analyzer) instrument and calibrate its sensors using an USB-Serial connection (so-called "Serial" option in the following parts) (Fig. 3a). The second script is dedicated to the analysis of spatial surveys (so-called "Space" option in the following parts) (Fig. 3b). The third script is dedicated to time series processing (so-called "Time" option in the following parts) (Fig. 3c).

The "Serial" option is dedicated for applications based on the use of the MEGA instrument that has been entirely conceived and designed at the INGV of Palermo by two of the current authors (Liuzzo & Cappuzzo). The MEGA instrument is not the focus of this paper and will be better presented to the community in future specific contributions. The other two options are composed of four panels (Fig. 3): (i) the first one (e.g. '1. File Treatment') is used to treat raw data files and create intermediate formatted files; (ii) the second one (e.g. '2. Data Processing') is used to select a parameter of interest from an intermediate formatted file, modify its corresponding series (correction, filtering, average) and, display resulted plots (for the "Time" option); (iii) the third one (e.g. '3. Data Analysis') is dedicated to show the results of the data analysis as correlations, cross-correlations, statistics, populations identification and plots (for the "Space" option); (iv) the last one (e.g. '4. Save') is used to save final processed datasets and related information (as populations) in .csv files.

# 3. Main functionalities

## 3.1. Initialization

The initialization step (first panel of the “Time” and “Space” options) aims at converting raw data files to intermediate files that may be manipulated in the other panels (Fig. 3b, c). Raw data files are “.csv” files downloaded from the MEGA instrument or created by the user in a compatible format to be correctly processed (see the user manual).

In the “Space” and “Time” options (Fig. 3b, c) the user can choose either the instrumental calibration by default or new calibration parameters to recalculate data series. Data reduction then performed in order to identify internal errors (typographical errors) or unreliable measurements, i.e. out of the range of values defined for the battery voltage, the pump flux and the horizontal dilution of precision (HDOP) of the global positioning system (GPS). These limit values used to define outliers are set by default but may be changed by the user directly in the GUI. Rows containing bad values are either linearly interpolated with the “Time” option in order to correctly apply further time series analysis (in this case interpolated rows are kept in memory in order to be removed in final .csv files) or left as empty rows with the “Space” option. Additionally, as soil $CO_2$ measurements may be acquired with the accumulation chamber method in the “Space” option, an additional filter is applied on the r-squared value of computed soil $CO_2$ flux. In this case, values of soil $CO_2$ flux out of the r-squared range will be considered as outliers values and thus set at 0 $gm^{-2}d^{-1}$ (no flux). In the case where soil $CO_2$ flux has to be calculated from the dynamic concentration method, the software integrates the possibility in both options to convert the $CO_2$ %molar contents in flux (Camarda *et al.*, 2006; Gurrieri and Valenza, 1988; Liuzzo *et al.*, 2015). The conversion is made by the use of the equation of Camarda *et al.* (2006) that takes into account the soil permeability value defined by the user.

Once the raw data file is cleaned, the last step will generate intermediate files that are of first interests to keep processing the data thanks to the other panels of the

GUI. The definition of the time-lag period (in seconds) is here fundamental. We define the time-lag as the period in seconds separating two independent series of measurements (i.e. the period during which the used instrument will be in stand-by). With the "Time" option, a new intermediate file will be created each time that two consecutive rows are separated by a duration greater than the time-lag. With the "Space" option, an unique intermediate file will be created with the median values of each independent series (the median value being here considered as representative of the acquisition to avoid the effect of potential spikes on the average). The time-lag must not be confused with the sampling rate that is necessary shorter and defines the period (in seconds) between two measurements within the same series. The sampling rate is used to correctly adapt the scale and the legend of plots (spectra, spectrogram) by converting a range of measurements (number of records) into a time range (number of seconds).

### 3.2. The "Space" option: dealing with spatial surveys

The "Space" option aims to propose innovative and classical tools to deal with spatial surveys, i.e. with datasets where each point of measurements is defined by distinct geographical coordinates (Fig. 3b).

#### *3.2.1. Correlations and preliminary data correction*

The "Space" option allows coefficients of linear correlations to be identified (slope, offset, r-squared) between the parameter of interest (e.g. $CO_2$) and other records (e.g. temperature, pressure, wind speed) obtained by the user (Fig. 3b). In particular, these coefficients are often useful during soil surveys, where records can be affected by external parameters. For instance, soil $CO_2$ flux may be slightly dependent on pressure

178 (Barde-Cabusson *et al.*, 2009; Liuzzo *et al.*, 2013; Viveiros *et al.*, 2008). The equation
179 used for the linear regression is the following:
180
181 $$Signal_{Clr}(x,y) = Signal_{Raw}(x,y) - (a \times (Param_{Raw}(x,y) - Param_{Avg})$$
182
183 where $Signal_{Clr}\ (x,y)$ is the value of the parameter of interest at the geographical position
184 $(x,y)$ after the correction by linear regression, $Signal_{Raw}\ (x,y)$ is the value before the
185 correction, $a$ is the slope of the linear correlation, $Param_{Raw}\ (x,y)$ is the parameter used
186 to performed the regression at the geographical position $(x,y)$ and, $Param_{Avg}$ is the
187 average of this parameter for the whole dataset in order to correct the offset linked to
188 the correction. After the correction, coefficients of linear correlations are reprocessed
189 and automatically updated in order to verify the efficiency of the correction and identify
190 potential needs of further steps of correction.
191
192 ***3.2.2. Statistical analysis***
193 In order to better constrain the data distribution of the parameter of interest, the "Space"
194 option allows the user to display also a probability histogram together with the best fit
195 line of the potential normal distribution (see Supplementary Material). The normality
196 of the data series is tested via the Anderson-Darling normality test (SciPy library;
197 Anderson and Darling, 1952; Stephens, 1974, 1976). In our case, i.e. making the
198 assumption that both mean and variance are initially unknown, the Anderson-Darling
199 normality test rejects the hypothesis of normality with a 95% significance level if $A^2$
200 (the squared of the test statistic A) exceeds 0.752 for data series owning more than 8
201 samples (D'Agostino, 1986).
202 Together with the Anderson-Darling normality test, the "Space" option gives
203 the opportunity to calculate some classical statistical values: mean, standard deviation,

median, minimum, maximum, kurtosis, skewness (Fig. 3b). Here we focus on the last two indicators less common for non-regular users of statistical tools. The kurtosis is a measure of the tailedness of the probability distribution of a random variable, i.e. describing the shape of the probability distribution (Zwillinger and Kokoska, 2000). Using the Fisher's definition, normally distributed data should provide a result of 0. Skewness is the measure of the asymmetry of the probability distribution of a random variable with respect to its mean. The skewness, which could be either positive or negative, should be about 0 for normally distributed data (Zwillinger and Kokoska, 2000).

If the indicators described above (skewness, kurtosis, Anderson-Darling normality test) do not argue in favor of a normal distribution, it may be due to the presence of more than one population in the data series. Indeed, during spatial survey, one subject of major interest is often to discriminate the different populations that contribute to the data series. This is crucial in order to recognize the existence of distinct sources as, for example, biogenic or magmatic ones for soil $CO_2$ flux in volcanic context (Boudoire *et al.*, 2017b; Liuzzo *et al.*, 2015; Viveiros *et al.*, 2008). In order to address this specific issue, we have developed a new algorithm in SoilExp able to combine the two statistical methods more used in environmental scientific research, which can distinguish various populations from log-normally distributed data (Fig. 4a, b). The first is the graphical method based on probability plots known as Sinclair method (Chiodini *et al.*, 1998; Giammanco *et al.*, 2010; Sinclair, 1974); the second is the maximum-likelihood numerical method based on the use of Gaussian Mixture Model (GMM) implementing an expectation-maximization (EM) algorithm (Benaglia *et al.*, 2009; Boudoire *et al.*, 2018; Elio *et al.*, 2016). The Sinclair method provides an user-friendly view of the populations and mixed values, however, it has two main shortcomings. One is related to the low accuracy for datasets counting less than 100 values (Sinclair, 1974).

The second limitation is related to the difficulty to precisely estimate the confidence intervals. These problems are solved using the maximum-likelihood (ML) method that fits finite mixtures of normal distributions: we have implemented a Scikit-Learn-based algorithm that simulates such fitting with 1 to 10 populations with 1000 iterations for each simulation. The best simulation is then selected based on the value of the Bayesian Information Criterion (BIC) developed for model selection among finite set of simulations (Ghosh *et al.*, 2006) and displayed on the GUI (Fig. 3b) . Finally, for each value of the data serie, the algorithm predicts the probability that the value belongs to one of the defined populations. In our case, we have considered that if one value shows a probability to be defined by a single population greater than 95% thus it will be considered as part of this population. If not, this value is considered as an intermediate value (or mixed value) between the two neighboring populations. A the end, the algorithm allows the user to automatically see the result of this ML-based partitioning of the values on probability plots (Fig. 4a, b). Furthermore, the users can simulate different partitioning by modifying the number of inferred populations directly on the GUI (Fig. 3b), if the first step of differentiation is not satisfying.

### *3.2.3. Mapping*

After having performed data correction and statistical analysis, it is possible to obtain a first idea of the two-dimensions (2D) distribution of the data (Fig. 4c, d). Our aim is not to develop complex interpolating algorithms for which many software are already built. Here we propose a simple graphical representation of the data through two distinct maps. The first one uses a simple color gradient to show the 2D evolution of the values. The second one is more innovative (presented on Fig. 4 for SP and $CO_2$ data obtained at Stromboli), meaning that the map-builder takes into consideration the results of the population analysis described above, generating and displaying a repartition of the

values between the different populations (and related mixing values). If an internet connection and an API key are available (**https://developers.google.com/maps/documentation/javascript/get-api-key**), a background satellite map will be automatically downloaded and georeferenced from the Google Maps Platform. If not, the background will remain neutral. However, the upper left box (Fig. 4c, d) highlights the coordinates of the corners to facilitate the extraction of an adequate background map from other sources.

### 3.3. The "Time" option: processing time series

The "Time" option aims to propose classical tools to deal with time series, i.e. with datasets where the measurements have specific frequency (here defined as the sampling rate) (Fig. 3c).

#### *3.3.1. Correlations and cross-correlations*

"Time" option allows to identify coefficients of linear correlations (slope, offset, r-squared) between the parameter of interest and other records, where the control panel (Fig. 3c) is similar to the one in "Space" option. Sometimes some signals may have a time delay between them, which can be attributed either to an instrumental lag or to an effect caused by a natural phenomenon. To take into account these effects, we have implemented a SciPy-based algorithm to calculate the cross-correlations between each parameters. The algorithm couples complex-valued functions with conjugates and Fast Fourier Transform (FFT) to numerically determine both lags and r-squared values between time series. Best results are shown in the table of the "Time" option GUI (Fig. 3c).

#### *3.3.2. Signal processing*

The "Time" option gives the possibility to the user to apply three of the most common signal processing tools used in the geo-scientific community: (i) linear regression, (ii) moving average and, (iii) cut band filter.

The linear regression method is the same than in the "Space" option and only require to select the parameter used for the regression and to compute the corresponding coefficients. This method is used to remove short-term environmental influence on geochemical and geophysical signals (Boudoire *et al.*, 2017a; Liuzzo *et al.*, 2013).

The moving average method is a type of finite impulse response filter used to smooth out short-term signal variations. This method performs an average on a defined subset of the data series, then shifts forward to repeat the calculations, excluding the first value of the previous subset and including the next one. Using the convolution operator of the Numpy library, we have implemented a simple moving average method, i.e. giving the same weight to each value $a_j$:

$$movav\,(a_i) = \left(\frac{1}{k}\right) \times \sum_{j=i-k/2}^{i+k/2} a_j \;\text{ for }\; i \in \left]\frac{k}{2}, \frac{n-k}{2}\right[$$

where $i$ is the position of the value $a_i$ in the data series on which the moving average is applied, $n$ the length of the data series and $k$ the size of the subset. To deal with border effects (i.e. when the number of available values to perform the moving average is lower than the size of the defined subset), we have adapted the convolution to the number of available values:

$$movav\,(a_i) = \left(\frac{1}{i}\right) \times \sum_{j=0}^{i} a_j \;\text{ for }\; i \in \left[0, \frac{k}{2}\right]$$

306 $$movav\,(a_i) = \left(\frac{1}{n-i}\right) \times \sum_{j=i}^{n} a_j \;\text{ for }\; i \in [\,\frac{n-k}{2}\,, \mathrm{n}\,]$$

307

308 To enhance the reliability of the calculations linked to correlations and cross-
309 correlations, the moving average method is applied to all data series when computed.

310 Finally, to treat long-term signal variations, we have used the Fast Fourier
311 Transform (FFT) package of the SciPy library to develop a cut (or block) band filter.
312 This filter removes from the signal spectra (cf. '*fft'*) the frequencies belonging to an
313 interval defined by the user before making the inverse operation to rebuild the signal
314 (cf. '*ifft'*).

315

316 ***3.3.3. Graphical representation***

317 When pressing the plot-related buttons of the "Time" option, the user automatically
318 applies the correction and filtering methods that has been defined previously (Fig. 2).

319 Consequently, the user may decide to perform several combination between the
320 signals which is intended to compare:

321 (i) Compare the raw signal with the new corrected and filtered signal, and
322 eventually reinitialize the signal to apply a distinct protocol (Fig. 5a). Based
323 on the same statistical algorithm used with the "Space" option to
324 characterize populations, we have implemented an option allowing the user
325 to directly show on the plot the values belonging to the "highest" population
326 (often considered as representative of anomalous values with respect to the
327 background; Boudoire *et al.*, 2017a; Liuzzo *et al.*, 2013, Liuzzo *et al.*, 2015);
328 (ii) Compare the treated signal with another signal of interest (Fig. 5b). This
329 plot may be particularly useful to investigate well correlated or cross-
330 correlated signals;

(iii) See the FFT spectrum on which are displayed the three greatest frequencies (Fig. 5c). Thanks to the labels indicating the corresponding number of measurements, the user may define the frequency interval on which applying the cut-band filter;

(iv) See the corresponding spectrogram that is a different visual representation of the FFT spectrum, extensively used in geophysical signal processing (Fig. 5d). It is particularly useful to detect periodic components and signal perturbations that may affect all frequencies. Here we use the '*specgram'* function of the Matplotlib library with a linear detrend and a magnitude mode of 256 NFFT of default (Nonequispaced Fast Fourier Transform: the number of points in each processed block) and a 128 noverlap (the number of points of overlap between processed blocks). The user is free to modify these parameters directly in the Python 2.7 script (see user manual).

The signal analysis depends on the sampling rate, therefore we cannot use an unique legend for spectrum and spectrogram axes. Consequently, we have adapted the algorithms to show both the results of the raw signal analysis (in term of number of measurements) and their meaning using more classical units. For the last one, we have coupled the number of measurements and the sampling rate to have a real temporal scale (i) in seconds (between parenthesis) on the spectrum and (ii) in hertz on the spectrogram.

### 3.4. Saving and exporting results

The SoilExp software gives the opportunity to save every graphical object with different extensions (.png, .eps …), which can be easily further modified later.

Additionally both "Space" and "Time" GUI options have dedicated buttons to save .csv files. In the "Space" option, the final .csv file is similar to the intermediate

file but takes into account the results of the linear regressions that could be applied to correct the dataset. Additionally, it is possible to save a .csv file recording the data repartition between the defined populations and mixing groups. Both files aim to be eventually further processed through software dedicated to complementary and more specific tools as e.g. data interpolation, kriging, sequential Gaussian simulation (SGS). In the "Time" option, the final .csv file is also similar to the intermediate file but (i) has one supplemental column for the corrected and filtered data series and (ii) shows empty rows for missing values, which have been interpolated for the needs of signal processing. Such final file may be then processed through other complementary software for measurements of volcanic gas in plume or other environmental applications in atmospheric measurements (Fig. 1e, f).

# 4. SoilExp application: an example at Stromboli (Italy)

In volcanic environment, two of the main goals of soil surveys are (i) the identification of volcano-tectonic structures (Giammanco et al., 1997; Finizola et al., 2002, 2010) and (ii) the characterization of hydrothermal fluid circulation (Revil et al., 2011; ; Boudoire et al., 2018). Once, because these low permeable structures may favor the ascent of magmatic fluids leading to fissural eruptions (Boudoire et al., 2017b). Moreover, such structural interfaces may raise important issues concerning soil stability and thus landslide outbreak (Neri et al., 2004). To test the ability of SoilExp to deal with such goals, we have performed a spatial soil survey at Stromboli (Sicily, Italy) by the mean of the MEGA instrumental kit (Fig. 1). Three transects were performed with a 20 m-spacing for a total of 45 measurements of soil $CO_2$ flux and self-potential (dataset available with our distribution as "intermediate" test file). Here, we focused on the first transect (14 measurements), the one on the northern flank of the volcano which is the closest to populated areas (Fig. 1c).

Data analysis performed with the "Space" option of SoilExp reveals (i) the absence of correlation between soil $CO_2$ 'dynamic' concentration (*'CO2_10'*) and the environmental parameter (pressure *'P_atm'*, temperature *'T_atm'*, humidity '*Rh'*) during the transect and (ii) an important correlation ($R^2$ = 0.79) between soil $CO_2$ 'dynamic' concentration (*'CO2_10'*) and self-potential measurements (*'SP'*). Consequently, no correction from the environmental influence was applied (Viveiros et al., 2008) and we focus on both soil $CO_2$ flux and self-potential measurements in the following parts. The analysis performed by SoilExp shows that soil $CO_2$ 'dynamic' concentration ('CO2_10') varies from 0.07 to 0.95 %. Self-potential ('SP') varies from -155 to +77 mV. The Anderson-Darling normally test gives $A^2$ equal to 14.3 and 3.3 for 'CO2_10' and 'SP', respectively. These values are well above 0.752, and testify that both datasets do not present a normal distribution (at 95% of significance level). It means that these datasets are better explained by the presence of two or more populations. Actually, the new statistical algorithm developed in SoilExp highlights the presence of two populations of values for both parameters (Fig. 4a, b). Soil $CO_2$ 'dynamic' concentration shows the presence of two populations: one with high values (>0.20 % for 7.1% of the dataset; Fig. 4a) and the other with low values (<0.20 % for 92.9% of the dataset). We applied the equation of Camarda et al. (2006) to convert soil $CO_2$ 'dynamic' concentration in soil $CO_2$ flux for a range of soil permeability between 15 and 50, i.e. the most common values for volcanic soils (Camarda et al., 2006). The calculated upper limit of the population of low soil $CO_2$ flux does not exceed 42 $gm^{-2}d^{-1}$. This value is in accordance with the definition of a "background" population characterized by low soil $CO_2$ flux and generally ascribed to the biological soil activity (Liuzzo et al., 2015; Boudoire et al., 2017b). Conversely, the population of higher soil $CO_2$ flux (up to 233 $gm^{-2}d^{-1}$) is consistent with a magmatic-hydrothermal origin of the released fluids (Giammanco et al., 1997; Liuzzo et al., 2015). Self-potential shows also

409 the presence of one population of high values (from -9 up to +77 mV for 14.3% of the
410 dataset; Fig. 4b) whereas most of the dataset is defined by a population of more negative
411 values (from -169 up to -100 mV for 85.7% of the dataset).

412 Interestingly, the map-building of the soil $CO_2$ 'dynamic' concentration (Fig.
413 4c) and self-potential (Fig. 4d), based on this population analysis, shows that the
414 population of high soil $CO_2$ 'dynamic' concentration spatially correlates with the high
415 self-potential measurements. This positive correlation between the two parameters is
416 consistent with an upward migration of hydrothermal fluids in a restricted part of the
417 transect (<40 m-wide) as documented for other volcanic systems (Barde-Cabusson et
418 al., 2009; Bennati et al., 2011). Actually, this restricted part of the transect is cut by the
419 Nel Cannestrà eruptive fissure that is known representing a low permeability structure,
420 in relation with N41° inferred regional faults, (Finizola et al., 2002, 2010; Carapezza et
421 al., 2009). The identification and characterization of such structure that favors the
422 ascent of magmatic fluids raise important civil protection issues (Boudoire et al.,
423 2017b). Current monitoring is performed in this area by the Istituto Nationale di
424 Geofisica e Vulcanologia (INGV) (Carapezza et al., 2009).

425

426 # 5. Conclusion

427 In this work we presented an open-source Graphical User Interface (GUI) software,
428 SoilExp, which is written in Python language and is able to provide statistical and
429 spectrum analysis as well as several options on filtering and correcting analysis on
430 records acquired during spatial/temporal surveys. The software is based on two main
431 options. Firstly, the "Space" option, aims to display the main statistical indicators used
432 to study spatial surveys, to test the normality of data series, to identify and define the
433 populations constituting the dataset through an innovative algorithm, and to show
434 results on satellite maps. The second one, the "Time" option, aims to process time series

435 through classical tools used in signal processing (linear regression, moving average,
436 cut-band filter, cross-correlations) and in signal representations (scatter plots, spectra,
437 spectrogram). Beyond facilitating the fast outcome from field surveys by offering
438 filtering tools, graphical results and statistical analyses, SoilExp gives to the users the
439 possibility to integrate all the results in a unique tool of elaboration, improving the
440 research potential of the scientific community dealing with spatial and temporal soil
441 surveys.

442

## 443 Acknowledgments

444 S. Gurrieri is acknowledged for constructive discussions and advices on the script. P.
445 Boudoire, A. Finizola and T. Ricci are acknowledged for providing technical support
446 crucial for instrumental tests. People proposing or commenting open source Python
447 codes and solutions on forums are gratefully acknowledged. We thank the Associate
448 Editor and Jean Vandemeulebrouck for their suggestions that greatly improved the
449 clarity and quality of the paper. This work has been funded by the Fondo Sociale
450 Europeo (PO FSE 2014-2020) in the frame of the project "Metodi di controllo
451 geochimico e geofisico dei fenomeni naturali sul campo ed in laboratorio". We also
452 acknowledge the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25).

453

## 454 Computer code availability

455 The SoilExp software including full scripts, user manual and example files may be
456 freely downloaded from https://github.com/FreeMindsObservatory/SoilExp (main
457 developer: Dr. Guillaume Boudoire; corresponding author). The first distribution
458 started to be developed in 2017 (SoilExp 1.0 based on Python 2.7; 16.6 Mo), made
459 available for MacOSX and Windows platforms. It may be modified by any contributor

according his needs thanks to a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0) (https://creativecommons.org/licenses/by-nc-sa/4.0/). The software requires the installation of the Anaconda distribution to use the Spyder open source cross-platform integrated development environment (IDE) integrated all required scientific libraries. Please contact the corresponding author for any support regarding the SoilExp software.

## References


Allard, P., Carbonnelle, J., Dajlevic, D., Le Bronec, J., Morel, P., Robe, M.C., Maurenas, J.M., Faivre-Pierret, R., Martin, D., Sabroux, J.C., Zettwoog, P., 1991. Eruptive and diffuse emissions of $CO_2$ from Mount Etna. Nature 351(6325), 387-391.

Anderson, T.W., Darling, D. A., 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics 23, 193–212, doi:10.1214/aoms/1177729437.

Aubert, M., Camus, G., Fournier, C. 1984. Resistivity and magnetic surveys in groundwater prospecting in volcanic areas—case history maar of Beaunit, Puy de Dome, France. Geophysical prospecting 32(4), 554-563.

Barde-Cabusson, S., Finizola, A., Revil, A., Ricci, T., Piscitelli, S., Rizzo, E., Angeletti, B., Balasco, M., Bennati, L., Byrdina, S., Carzaniga, N., Crespy, A., Di Gangi, F., Morin, J., Perrone, A., Rossi, M., Roulleau, E., Suski, B., Villeneuve, N., 2009. New geological insights and structural control on fluid circulation in La Fossa cone (Vulcano, Aeolian Islands, Italy). J. Volcanol. Geotherm. Res. 185, 231–245, doi: 10.1016/j.jvolgeores.2009.06.002.

Benaglia, T., Chauveau, D., Hunter, D., Young, D., 2009. mixtools : An R package for analyzing finite mixture models. Journal of Statistical Software 32(6), 1-29.

Bennati, L., Finizola, A., Walker, J.A., Lopez, D.L., Higuera-Diaz, I.C., Schütze, C., Barahona, F., Cartagena, R., Conde, V., Funes, R., 2011. Fluid circulation in a complex volcano-tectonic setting, inferred from self-potential and soil CO2 flux surveys: the Santa María–Cerro Quemado–Zunil volcanoes and Xela caldera (Northwestern Guatemala). J. Volcanol. Geotherm. Res. 199 (3–4), 216–229. http://dx.doi.org/10.1016/j.jvolgeores.2010.11. 008.

Boudoire, G., Di Muro, A., Liuzzo, M., Ferrazzini, V., Peltier, A., Gurrieri, S., Michon, L., Giudice, G., Kowalski, P., Boissier, P., 2017a. New perspectives on volcano monitoring in a tropical environment: continuous measurements of soil $CO_2$ flux at Piton de la Fournaise (La Réunion Island, France). Geophysical Research Letters 44(16), 8244-8253.

Boudoire, G., Finizola, A., Di Muro, A., Peltier, A., Liuzzo, M., Grassa, F., Delcher, E., Brunet, C., Boissier, P., Chaput, M., Ferrazzini, V., Gurrieri, S., 2018. Small-scale spatial variability of soil $CO_2$ flux: Implication for monitoring strategy. Journal of Volcanology and Geothermal Research 366, 13-26.

Boudoire, G., Liuzzo, M., Di Muro, A., Ferrazzini, V., Michon, L., Grassa, F., Derrien, A., Villeneuve, N., Bourdeu, A., Brunet, C. Giudice, G., Gurrieri, S., 2017b. Investigating the deepest part of a volcano plumbing system: evidence for an active

511 magma path below the western flank of Piton de la Fournaise (La Réunion Island).
512 Journal of Volcanology and Geothermal Research 341, 193-207.
513
514 Byrdina, S., Rücker, C., Zimmer, M., Friedel, S., Serfling, U., 2012. Self potential
515 signals preceding variations of fumarole activity at Merapi volcano, Central Java.
516 Journal of Volcanology and Geothermal Research 215, 40-47.
517
518 Camarda, M., Gurrieri, S., Valenza, M., 2006. In situ permeability measurements based
519 on a radial gas advection model : Relationships between soil permeability and diffuse
520 $CO_2$ degassing in volcanic areas. Pure and applied geophysics 163(4), 897-914.
521
522 Carapezza, M.L., Ricci, T., Ranaldi, M., Tarchini, L., 2009. Active degassing structures
523 of Stromboli and variations in diffuse CO2 output related to the volcanic activity.
524 Journal of Volcanology and Geothermal Research 182(3-4), 231-245.
525
526 Chatterjee, S., Deering, C.D., Waite, G.P., Prandi, C., Lin, P., 2019. An adaptive
527 sampling strategy developed for studies of diffuse volcanic soil gas emissions. Journal
528 of Volcanology and Geothermal Research.
529
530 Chiodini, G., Cioni, R., Guidi, M., Raco, B., Marini, L., 1998. Soil $CO_2$ flux
531 measurements in volcanic and geothermal areas. Applied Geochemistry 13(5), 543-
532 552.
533
534 Chiodini, G., Frondini, F., Cardellini, C., Granieri, D., Marini, L., Ventura, G., 2001.
535 $CO_2$ degassing and energy release at Solfatara volcano, Campi Flegrei, Italy. Journal of
536 Geophysical Research : Solid Earth 106(B8), 16213-16221.

537

538 Chiodini, G., Granieri, D., Avino, R., Caliro, S., Costa, A., Werner, C., 2005. Carbon
539 dioxide diffuse degassing and estimation of heat release from volcanic and
540 hydrothermal systems. Journal of Geophysical Research : Solid Earth 110(B8).

541

542 D'Agostino, R.B., 1986. Tests for the Normal Distribution. In Goodness-of-Fit
543 Techniques. New York, Marcel Dekker, ISBN 0-8247-7487-6.

544

545 Elio, J., Ortega, M.F., Nisi, B., Mazadiego, L.F., Vaselli, O., Caballero, J., Chacon, E.,
546 2016. A multi-statistical approach for estimating the total output of $CO_2$ from diffusive
547 soil degassing by the accumulation chamber method. International Journal of
548 Greenhouse Gas Control 47, 351-363.

549

550 Elskens, I., Tazieff, H., Tonani, F., 1964. A new method for volcanic gas analyses in
551 the field. Bulletin Volcanologique 27(1), 347-350.

552

553 Finizola, A., Ricci, T., Deiana, R., Cabusson, S.B., Rossi, M., Praticelli, N., Giocoli,
554 A., Romano, G., Delcher, E., Suski, B., Revil, A., Menny, P., Di Gangi, F., Letort, J.,
555 Peltier, A., Villasante-Marcos, V., Douillet, G., Avard, G., Lelli, M., 2010. Adventive
556 hydrothermal circulation on Stromboli volcano (Aeolian Islands, Italy) revealed by
557 geophysical and geochemical approaches: implications for general fluid flow models
558 on volcanoes. Journal of Volcanology and Geothermal Research 196(1-2), 111-119.

559

560 Finizola, A., Sortino, F., Lénat, J.F., Valenza, M., 2002. Fluid circulation at Stromboli
561 volcano (Aeolian Islands, Italy) from self-potential and $CO_2$ surveys. Journal of
562 Volcanology and Geothermal Research 116(1), 1-18.

Gaudin, D., Finizola, A., Delcher, E., Beauducel, F., Allemand, P., Delacourt, C., Brothelande, E., Peltier, A., Di Gangi, F., 2015. Influence of rainfalls on heat and steam fluxes of fumarolic zones: Six months records along the Ty fault (Soufrière of Guadeloupe, Lesser Antilles). Journal of Volcanology and Geothermal Research 302, 273-285.

Ghosh, J.K., Delampady, M., Samanta, T., 2006. An Introduction to Bayesian Analysis: Theory and Methods. Springer-Verlag, New York.

Giammanco, S., Gurrieri, S., Valenza, M., 1997. Soil $CO_2$ degassing along tectonic structures of Mount Etna (Sicily) : the Pernicana fault. Applied Geochemistry 12(4), 429-436.

Giammanco, S., Bellotti, F., Groppelli, G., Pinton, A., 2010. Statistical analysis reveals spatial and temporal anomalies of soil $CO_2$ efflux on Mount Etna volcano (Italy). Journal of Volcanology and Geothermal Research 194(1), 1-14.

Gresse, M., Vandemeulebrouck, J., Byrdina, S., Chiodini, G., Bruno, P. P., 2016. Changes in $CO_2$ diffuse degassing induced by the passing of seismic waves. Journal Of Volcanology And Geothermal Research 320, 12–18.

Gurrieri, S., Valenza, M., 1988. Gas transport in natural porous mediums : a method for measuring $CO_2$ flows from the ground in volcanic and geothermal areas. Rend. Soc. Ital. Mineral. Petrol 43, 1151-1158.

589 Hernández, P.A., Salazar, J.M., Shimoike, Y., Mori, T., Notsu, K., Pérez, N., 2001.
590 Diffuse emission of $CO_2$ from Miyakejima volcano, Japan. Chemical Geology 177(1),
591 175-185.
592
593 Hinkle, M.E., Dilbert, C.A., 1984. Gases and trace elements in soils at the North Silver
594 Bell deposit, Pima County, Arizona. Journal of Geochemical Exploration 20(3), 323-
595 336.
596
597 Irwin, W.P., Barnes, I., 1980. Tectonic relations of carbon dioxide discharges and
598 earthquakes. Journal of Geophysical Research : Solid Earth 85(B6), 3115-3121.
599
600 Kucera, C., Kirkham, D.R., 1971. Soil respiration studies in tallgrass prairie in
601 Missouri. Ecology 52(5), 912-915.
602
603 Liuzzo, M., Di Muro, A., Giudice, G., Michon, L., Ferrazzini, V., Gurrieri, S., 2015.
604 New evidence of $CO_2$ soil degassing anomalies on Piton de la Fournaise volcano and
605 the link with volcano tectonic structures. Geochemistry, Geophysics, Geosystems
606 16(12), 4388-4404.
607
608 Liuzzo, M., Gurrieri, S., Giudice, G., Giuffrida, G., 2013. Ten years of soil $CO_2$
609 continuous monitoring on Mt. Etna : Exploring the relationship between processes of
610 soil degassing and volcanic activity. Geochemistry, Geophysics, Geosystems 14(8),
611 2886-2899.
612

Lovell, J.S., Hale, M., Webb, J.S., 1983. Soil air carbon dioxide and oxygen measurements as a guide to concealed mineralization in semi-arid and arid regions. Journal of Geochemical Exploration 19(1-3), 305-317.

Neri, M., Acocella, V., Behncke, B., 2004. The role of the Pernicana Fault System in the spreading of Mt. Etna (Italy) during the 2002–2003 eruption. Bulletin of Volcanology 66(5), 417-430.

Padrón, E., Melián, G., Marrero, R., Nolasco, D., Barrancos, J., Padilla, G., Hernández, P.A., Pérez, N.M., 2008. Changes in the diffuse $CO_2$ emission and relation to seismic activity in and around El Hierro, Canary Islands. In Terrestrial Fluids, Earthquakes and Volcanoes : The Hiroshi Wakita Volume III, 95-114, Birkhäuser Basel.

Pearson, S.C.P., Connor, C.B., Sanford, W.E., 2008. Rapid response of a hydrologic system to volcanic activity: Masaya volcano, Nicaragua. Geology 36(12), 951-954.

Revil, A., Finizola, A., Ricci, T., Delcher, E., Peltier, A., Barde-Cabusson, S., Avard, G., Bailly, T., Bennati, L., Byrdina, S., Colonge, J., Di Gangi, F., Douillet, G., Lupi, M., Letort, J., Tsang Hin Sun, E., 2011. Hydrogeology of Stromboli volcano, Aeolian Islands (Italy) from the interpretation of resistivity tomograms, self-potential, soil temperature and soil $CO_2$ concentration measurements. Geophysical Journal International 186, 1078–1094, https://doi.org/10.1111/j.1365-246X.2011.05112.x.

Sandig, C., Sauer, U., Bräuer, K., Serfling, U., Schütze, C., 2014. Comparative study of geophysical and soil–gas investigations at the Hartoušov (Czech Republic) natural $CO_2$ degassing site. Environmental earth sciences 72(5), 1421-1434.

Sinclair, A.J., 1974. Selection of threshold values in geochemical data using probability graphs. Journal of Geochemical Exploration 3(2), 129-149.

Stephens, M.A., 1974. EDF Statistics for Goodness of Fit and Some Comparisons. Journal of the American Statistical Association 69, 730–737, doi:10.2307/2286009.

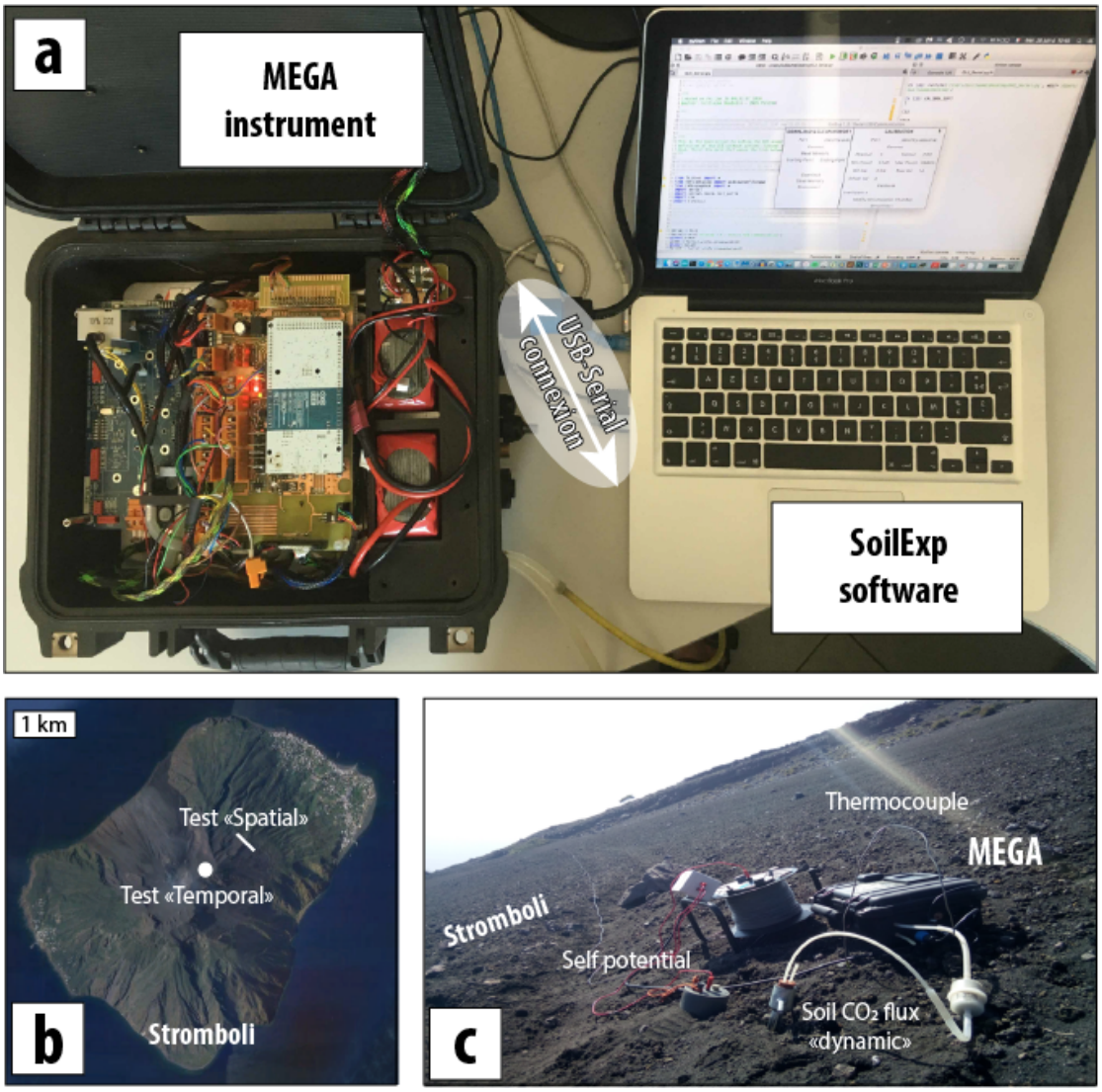
Stephens, M.A., 1976. Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters. Annals of Statistics 4, 357-369.

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.

Shinohara, H., 2005. A new technique to estimate volcanic gas composition: Plume measurements with a portable multi-sensor system. J. Volcanol. Geotherm. Res. 143, 319–333.

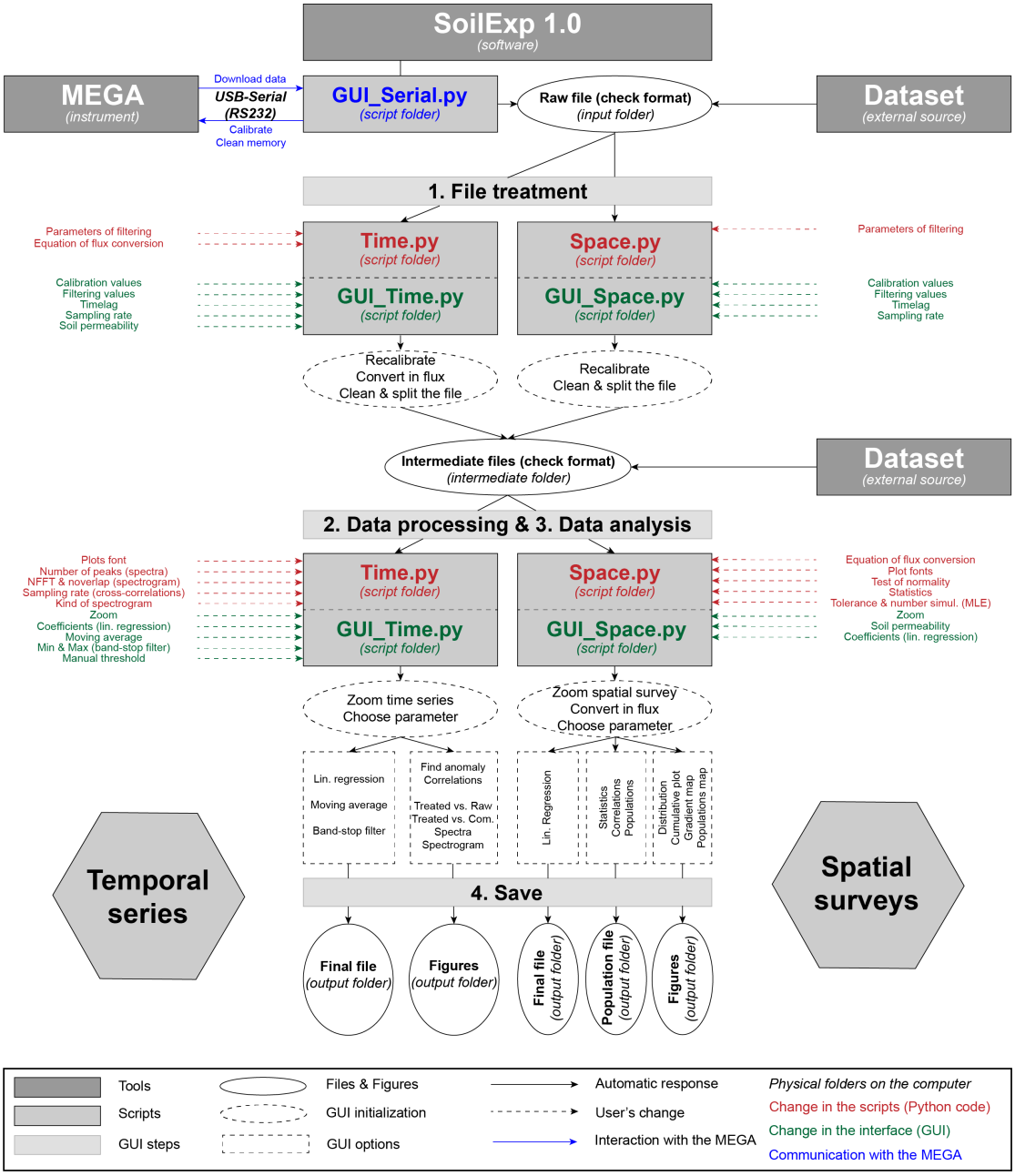
Zwillinger, D., Kokoska, S., 2000. CRC Standard Probability and Statistics Tables and Formulae. Chapman & Hall, New York.

# Figures




662

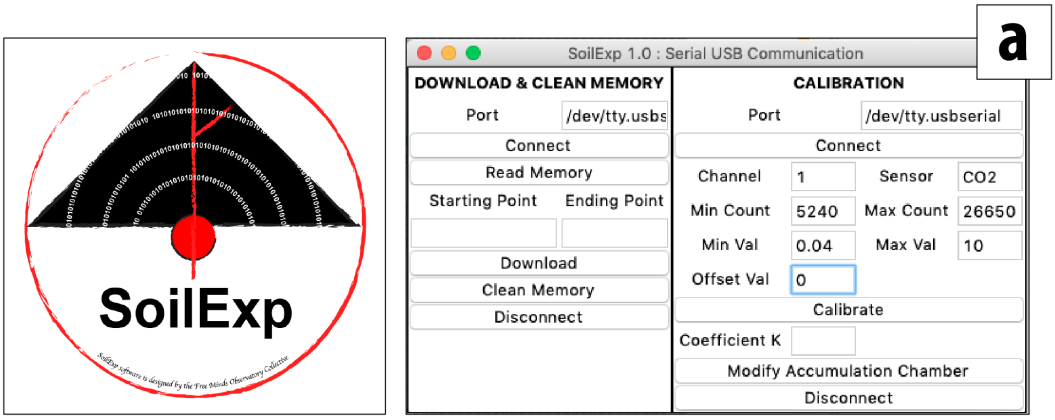
663 **Fig. 1. (a) The MEGA instrument and the SoilExp software, in evidence the USB-Serial**
664 **communication between the instrument and the software. (b) Tests of soil surveys at**
665 **Stromboli (Sicily, Italy) (c) based on soil $CO_2$ flux, self-potential and ground temperature.**
666 **These tests aim to illustrate the use of the SoilExp software in this study.**

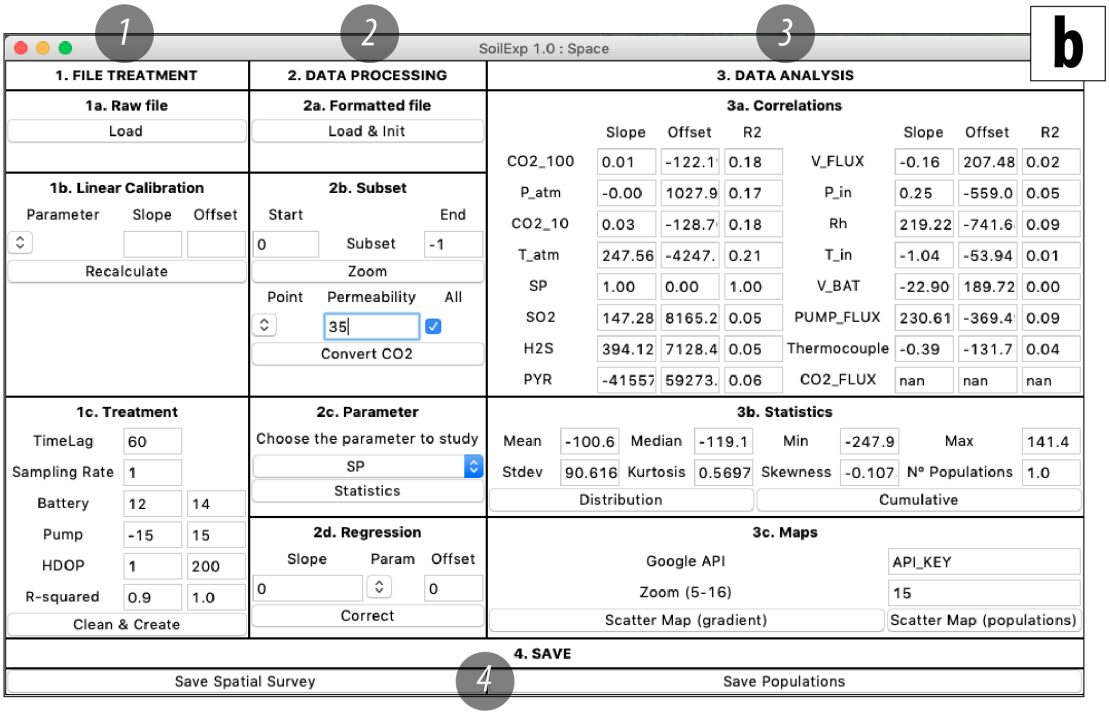



667

668 **Fig. 2. General scheme of use of the SoilExp software applied either to dataset acquired**

669 **with the MEGA instrument or though external sources.**


a
SoilExp
SoilExp 1.0 : Serial USB Communication
DOWNLOAD & CLEAN MEMORY
Port
/dev/tty.usbs
Connect
Read Memory
Starting Point
Ending Point
Download
Clean Memory
Disconnect
CALIBRATION
Port
/dev/tty.usbserial
Connect
Channel
1
Sensor
CO2
Min Count
5240
Max Count
26650
Min Val
0.04
Max Val
10
Offset Val
0
Calibrate
Coefficient K
Modify Accumulation Chamber
Disconnect


b
1
2
3
SoilExp 1.0 : Space
1. FILE TREATMENT
1a. Raw file
Load
1b. Linear Calibration
Parameter
Slope
Offset
Recalculate
1c. Treatment
TimeLag
60
Sampling Rate
1
Battery
12
14
Pump
-15
15
HDOP
1
200
R-squared
0.9
1.0
Clean & Create
2. DATA PROCESSING
2a. Formatted file
Load & Init
2b. Subset
Start
End
0
Subset
-1
Zoom
Point
Permeability
All
35
Convert CO2
2c. Parameter
Choose the parameter to study
SP
Statistics
2d. Regression
Slope
Param
Offset
0
0
Correct
3. DATA ANALYSIS
3a. Correlations
Slope
Offset
R2
CO2_100
0.01
-122.1
0.18
P_atm
-0.00
1027.9
0.17
CO2_10
0.03
-128.7
0.18
T_atm
247.56
-4247.
0.21
SP
1.00
0.00
1.00
SO2
147.28
8165.2
0.05
H2S
394.12
7128.4
0.05
PYR
-41557
59273.
0.06
V_FLUX
-0.16
207.48
0.02
P_in
0.25
-559.0
0.05
Rh
219.22
-741.6
0.09
T_in
-1.04
-53.94
0.01
V_BAT
-22.90
189.72
0.00
PUMP_FLUX
230.61
-369.4
0.09
Thermocouple
-0.39
-131.7
0.04
CO2_FLUX
nan
nan
nan
3b. Statistics
Mean
-100.6
Median
-119.1
Min
-247.9
Max
141.4
Stdev
90.616
Kurtosis
0.5697
Skewness
-0.107
N° Populations
1.0
Distribution
Cumulative
3c. Maps
Google API
API_KEY
Zoom (5-16)
15
Scatter Map (gradient)
Scatter Map (populations)
4. SAVE
4
Save Spatial Survey
Save Populations

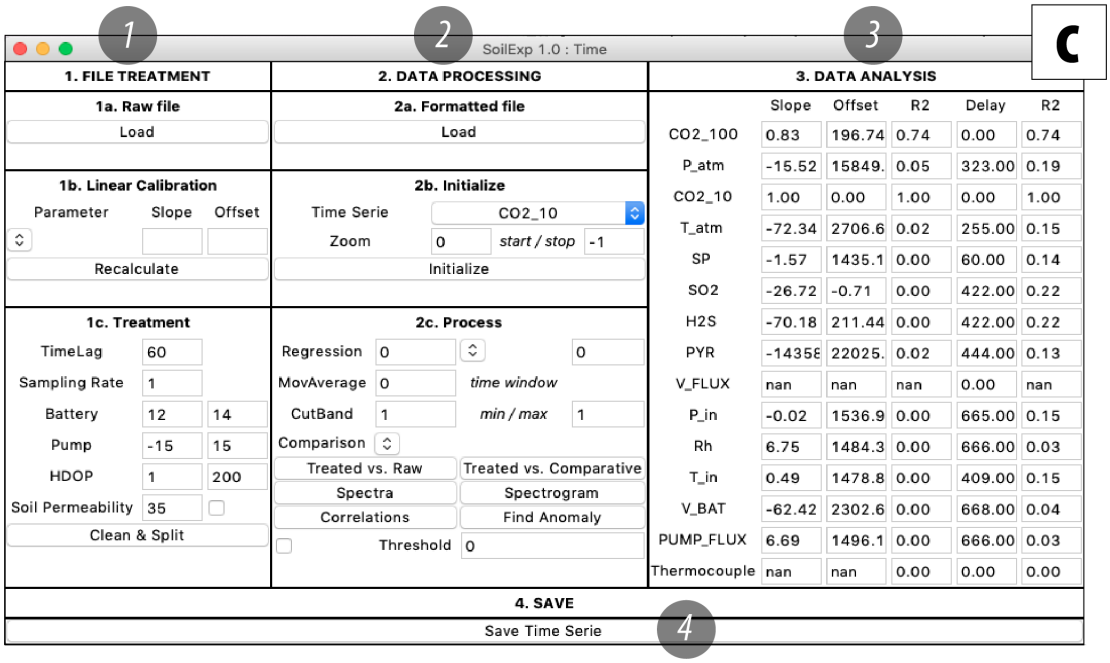

c
1
2
3
SoilExp 1.0 : Time
1. FILE TREATMENT
1a. Raw file
Load
1b. Linear Calibration
Parameter
Slope
Offset
Recalculate
1c. Treatment
TimeLag
60
Sampling Rate
1
Battery
12
14
Pump
-15
15
HDOP
1
200
Soil Permeability
35
Clean & Split
2. DATA PROCESSING
2a. Formatted file
Load
2b. Initialize
Time Serie
CO2_10
Zoom
0
start / stop
-1
Initialize
2c. Process
Regression
0
0
MovAverage
0
time window
CutBand
1
min / max
1
Comparison
Treated vs. Raw
Treated vs. Comparative
Spectra
Spectrogram
Correlations
Find Anomaly
Threshold
0
3. DATA ANALYSIS
Slope
Offset
R2
Delay
R2
CO2_100
0.83
196.74
0.74
0.00
0.74
P_atm
-15.52
15849.
0.05
323.00
0.19
CO2_10
1.00
0.00
1.00
0.00
1.00
T_atm
-72.34
2706.6
0.02
255.00
0.15
SP
-1.57
1435.1
0.00
60.00
0.14
SO2
-26.72
-0.71
0.00
422.00
0.22
H2S
-70.18
211.44
0.00
422.00
0.22
PYR
-14358
22025.
0.02
444.00
0.13
V_FLUX
nan
nan
nan
0.00
nan
P_in
-0.02
1536.9
0.00
665.00
0.15
Rh
6.75
1484.3
0.00
666.00
0.03
T_in
0.49
1478.8
0.00
409.00
0.15
V_BAT
-62.42
2302.6
0.00
668.00
0.04
PUMP_FLUX
6.69
1496.1
0.00
666.00
0.03
Thermocouple
nan
nan
0.00
0.00
0.00
4. SAVE
4
Save Time Serie

**Fig. 3. Graphical User Interface (GUI) of the "Serial" (a), "Space" (b) and "Time" (c) option of the SoilExp software. The GUI is divided in 4 panels. Panel (1) is dedicated to format the raw file in intermediate formatted files after applying potential distinct calibrations and conversions, and cleaning the dataset. Panel (2) aimed to process the data obtained from the intermediate formatted files either from the previous step or formatted independently by the user (conversion, moving average, linear regression, cut band filter). Panel (3) shows the result of the datasets processing and analysis (correlations, cross-correlations, statistics, analysis of populations, distribution, maps). Panel (4) allows to save the dataset transformed with the above operations in final .csv file.**

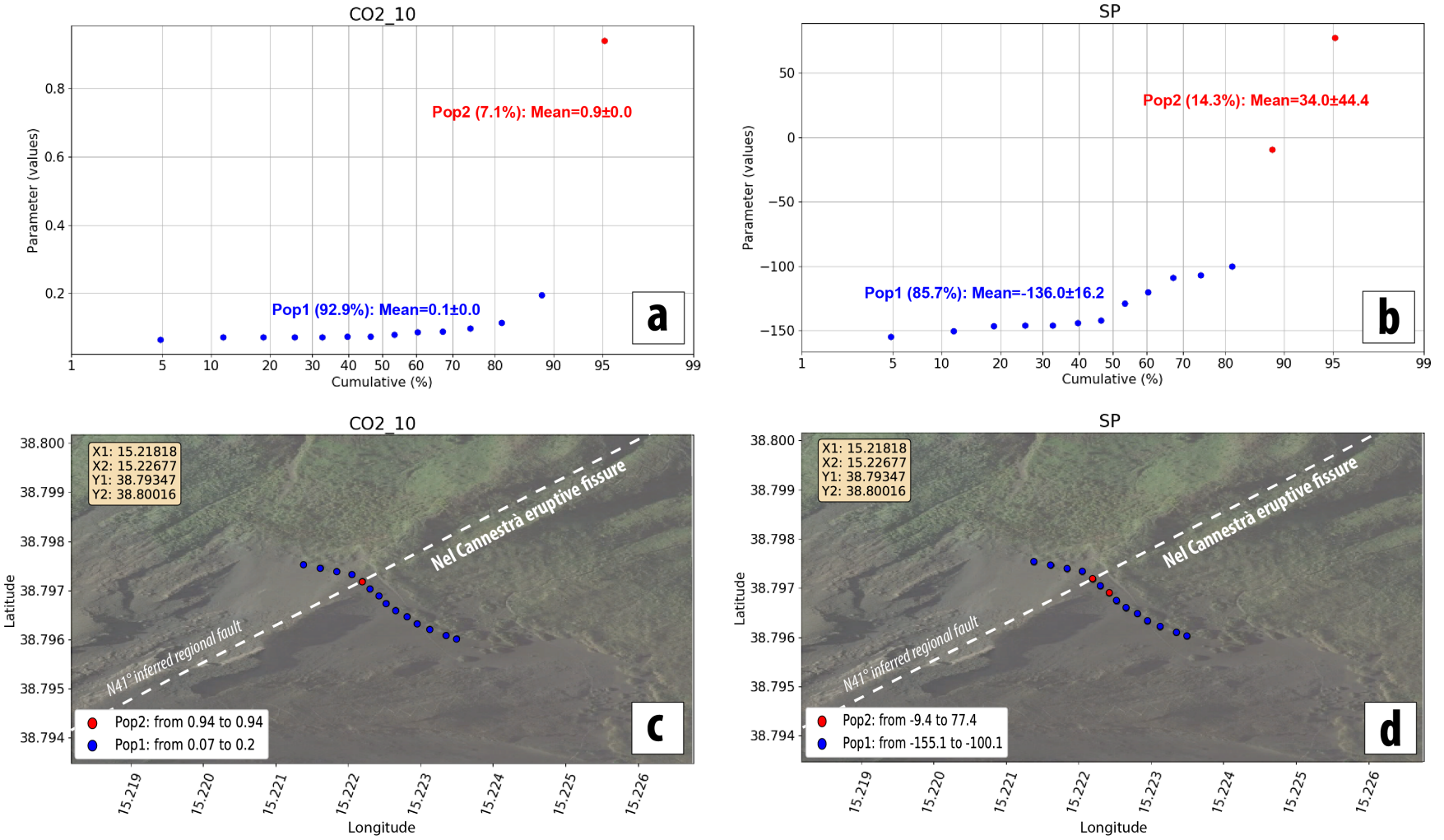
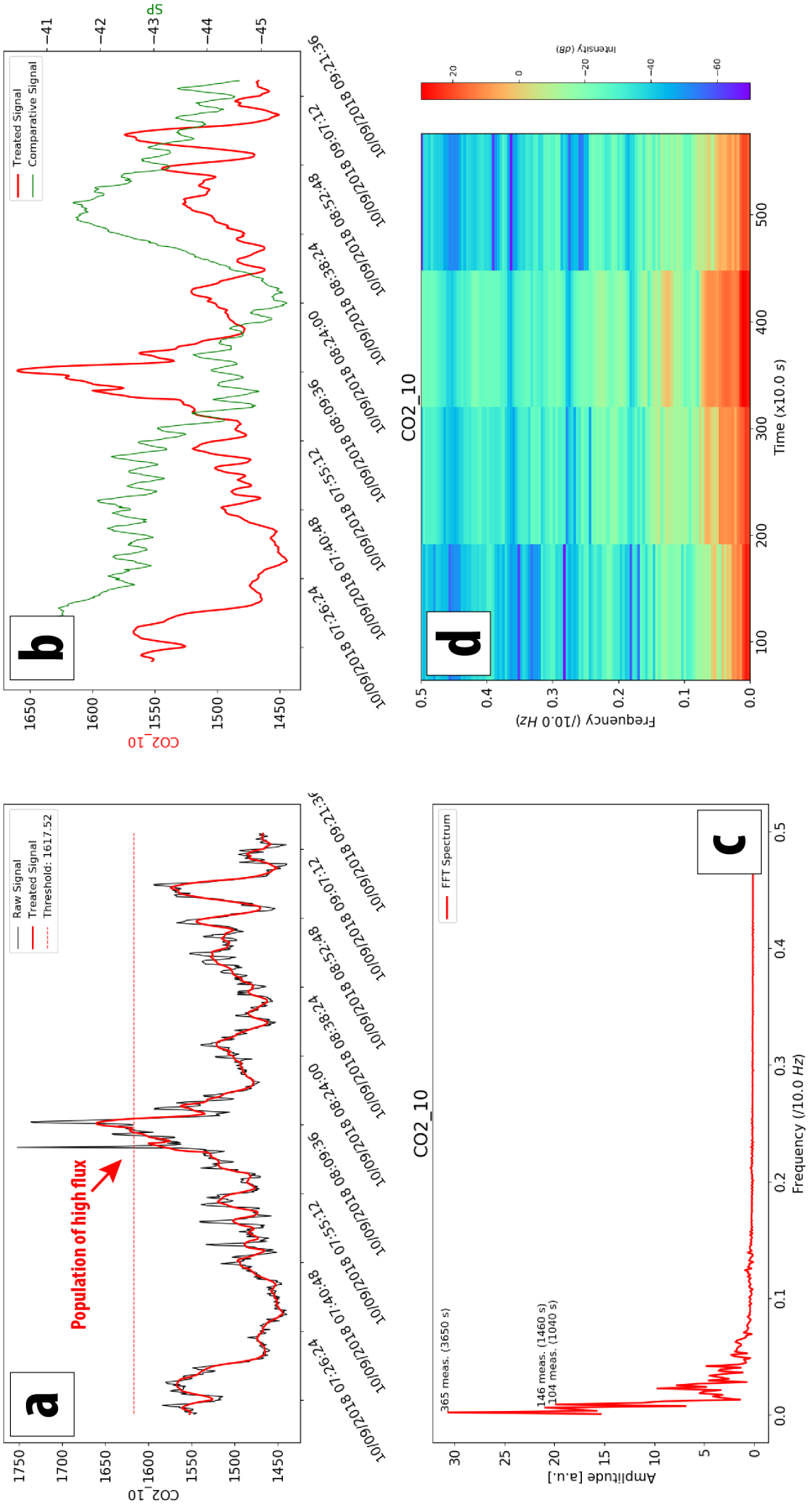

CO2_10
Parameter (values)
Cumulative (%)
Pop2 (7.1%): Mean=0.9±0.0
Pop1 (92.9%): Mean=0.1±0.0
a
SP
Parameter (values)
Cumulative (%)
Pop2 (14.3%): Mean=34.0±44.4
Pop1 (85.7%): Mean=-136.0±16.2
b
CO2_10
X1: 15.21818
X2: 15.22677
Y1: 38.79347
Y2: 38.80016
Nel Cannestrà eruptive fissure
N41° inferred regional fault
Pop2: from 0.94 to 0.94
Pop1: from 0.07 to 0.2
c
Latitude
Longitude
SP
X1: 15.21818
X2: 15.22677
Y1: 38.79347
Y2: 38.80016
Nel Cannestrà eruptive fissure
N41° inferred regional fault
Pop2: from -9.4 to 77.4
Pop1: from -155.1 to -100.1
d
Latitude
Longitude

**Fig. 4. Example of data analysis obtained by using the "Space" option at Stromboli (soil $CO_2$ flux and self-potential measurements along a transect with a 20 m-spacing; cf. Test "Spatial" on Fig. 1c, d). Probability plot of (a) soil $CO_2$ flux measurements obtained using a 0-10 %molar IR spectrometer (e.g. CO2_10 by "dynamic" concentration; Gurrieri & Valenza, 1988; Camarda *et al.*, 2006) and (b) self-potential measurements carried out with a pair of non-polarizable $Cu/CuSO_4$ electrodes (e.g. SP; Finizola *et al.*, 2010). The identification of distinct populations is based on the maximum-likelihood numerical method (see text). Map highlighting the corresponding (c) soil $CO_2$ flux and (d) self-potential transect performed at Stromboli (cf. Fig. 1). The satellite map is obtained from Google Map. In case of absence of API key (https://developers.google.com/maps/documentation/javascript/get-api-key), the background will stay white. However, the (decimal) coordinates of the corners are reported in the upper left box in order to let the user free to download a map from distinct sources. In this example, the "Space" option allows to identify a soil $CO_2$ anomaly coupled with a positive SP anomaly that highlight an upward migration of hydrothermal fluids along the Nel Cannestrà eruptive fissure. This result is in accordance with previous study (Finizola *et al.*, 2002, 2010; Carapezza *et al.*, 2009).**


a
Raw Signal
Treated Signal
Threshold: 1617.52
Population of high flux
CO2_10
b
Treated Signal
Comparative Signal
CO2_10
Sp
c
FFT Spectrum
CO2_10
365 meas. (3650 s)
146 meas. (1460 s)
104 meas. (1040 s)
Amplitude [a.u.]
Frequency (/10.0 Hz)
d
CO2_10
Frequency (/10.0 Hz)
Time (x10.0 s)
Intensity (dB)

724

**Fig. 5. Example of data analysis obtained by using the "Time" option at Stromboli (soil $CO_2$ flux measured during about 2 hours at 0.1 Hz at the same site; cf. Test "Temporal" on Fig. 1b). (a) Comparison between raw and treated data (after applying the moving average). The threshold analysis allows us to detect the highest population of values (often considered as "anomalous" values) during the acquisition. (b) Comparison between treated soil $CO_2$ flux (e.g. CO2_10) and self-potential (e.g. SP) time series. Here the detected soil $CO_2$ flux anomaly is synchronous with low self-potential records. (c) Fast Fourier Transform (FFT) spectrum of the treated soil $CO_2$ signal. The 3 greatest frequency peaks are labelled with the corresponding period that may be cut using the cut band filter. (d) Spectrogram of the treated soil $CO_2$ signal (linear detrend; magnitude mode; NFFT=256; noverlap=128). In this example, with about 1200 measurements, there are not enough data available to obtain a smoothed spectrogram considering a NFFT of 256.**