

Application of stochastic simulation to CO₂ flux from soil: Mapping and quantification of gas release

C. Cardellini

Dipartimento di Scienze della Terra, Università di Perugia, Perugia, Italy

G. Chiodini

Osservatorio Vesuviano, Istituto Nazionale di Geofisica e Vulcanologia, Naples, Italy

F. Frondini

Dipartimento di Scienze della Terra, Università di Perugia, Perugia, Italy

Received 22 August 2002; revised 1 April 2003; accepted 29 April 2003; published 13 September 2003.

[1] Conditional sequential Gaussian simulations (sGs) have been applied for the first time to the study of soil diffuse degassing from different volcanic and nonvolcanic systems. The application regards five data sets of soil CO₂ fluxes measured with the accumulation chamber methodology at the volcanic areas of Solfatara of Pozzuoli (Italy), Vesuvio cone (Italy), Nisyros (Greece), and Horseshoe Lake (California) and at the nonvolcanic degassing area of Poggio dell'Olivo (Italy). The sGs algorithm was used to generate 100 realizations of CO₂ flux for each area. Probabilistic summaries of these simulations, together with the information given by probability plots, were used (1) to draw maps of the probability that CO₂ fluxes exceed thresholds specific for a background flux, i.e., to define the probable extension of the degassing structures, (2) to calculate the total CO₂ output, and (3) to quantify the uncertainty of the estimation. The results show that the sGs is a suitable tool to model soil diffuse degassing, producing realistic images of the distribution of the CO₂ fluxes that honor the histogram and variogram of the original data. Moreover, the relation between the sample design and the uncertainty of estimation was investigated leading to an empirical relation between uncertainty and the sampling density that can be useful for the planning of future CO₂ flux surveys. *INDEX TERMS:* 1099

Geochemistry: General or miscellaneous; 1094 Geochemistry: Instruments and techniques; 8424 Volcanology: Hydrothermal systems (8135); 8494 Volcanology: Instruments and techniques; *KEYWORDS:* CO₂ flux, stochastic simulation, volcanic degassing, diffuse degassing

Citation: Cardellini, C., G. Chiodini, and F. Frondini, Application of stochastic simulation to CO₂ flux from soil: Mapping and quantification of gas release, *J. Geophys. Res.*, 108(B9), 2425, doi:10.1029/2002JB002165, 2003.

1. Introduction

[2] Recently, a great interest has been addressed to the study of CO₂ Earth degassing. There are many objectives to such studies, for example, the definition of the relations between the flux and the tectonic structures [Etioppe *et al.*, 1999; Etioppe *et al.*, 1999; Lewicki and Brantley, 2000], the quantification of deeply derived CO₂ release to the atmosphere in the framework of the carbon global budget [Allard *et al.*, 1991; Brantley and Koepnick, 1995; Seward and Kerrick, 1996; Kerrick *et al.*, 1995; Marty and Tolstikhin, 1998; Williams *et al.*, 1992], and the study of volcanic degassing. In particular, numerous studies have been focused on the CO₂ soil diffuse degassing from quiescent active volcanoes [Brombach *et al.*, 2001; Chiodini *et al.*, 2001, 1996, 1998; Hernandez *et al.*, 1998; Gerlach *et al.*, 2001; Salazar *et al.*, 2001; Farrar *et al.*, 1995]. Most of

these studies showed that gas is not released uniformly from the whole volcanic apparatus, but rather from relatively restricted regions, which were named diffuse degassing structures (DDS) [Chiodini *et al.*, 2001]. Moreover, quantitative estimates of hydrothermal-volcanic gas released from DDS highlighted the importance of gas and thermal energy released by DDS in the mass and energetic balance of quiescent volcanic systems [Chiodini *et al.*, 2001]. Independently from the specific aims of investigations, the mapping of DDS and the quantification of the amount of released CO₂ can be considered common objectives of all these studies.

[3] Mapping of CO₂ fluxes from soil was mainly performed by use of interpolation algorithms, generally kriging [e.g., Bergfeld *et al.*, 2001; Chiodini *et al.*, 1996, 1998, 2001; Rogie *et al.*, 2001; Gerlach *et al.*, 2001]. The kriging algorithm is focused in providing the “best,” defined in a minimized least squares sense, hence unique, local estimate of a variable without specific regard to the resulting spatial statistics of the all estimates taken together [Deutsch and

Table 1. Summary of Sampling Design and CO₂ Flux Statistics

Study Area	Surveyed Surface, m ²	Number of Measurements	Sampling Design	Minimum and Maximum, g m ⁻² d ⁻¹	Mean CO ₂ Flux, g m ⁻² d ⁻¹
Solfatara of Pozzuoli	1.4×10^6	414	random	3.0–30987	1300.0
Vesuvio cone	3.4×10^4	110	20 m spaced grid	0.1–252	20.0
Poggio dell'Olivo	8.2×10^5	196	random	0.5–8797	489.0
Nisyros caldera	2.0×10^6	2883	20–25 m spaced grid	0.01–6175	39.7
Horseshoe Lake	1.3×10^5	313	20 m spaced grid	5.7–8670	800.0

Journal, 1998], producing a set of estimated values at the unsampled locations whose variogram does not match that of the original data. Moreover, kriging smoothes out the extreme extrapolated values, with small values being overestimated while large values are underestimated, hiding the pattern of high values, which is important in our applications to define degassing structures.

[4] Total CO₂ releases were usually calculated either by multiplying the arithmetic mean value of CO₂ fluxes by the surveyed areas, or by applying volume and area integration algorithms to the grids produced to contour the CO₂ flux, or by a graphical statistical approach (GSA) independent from any interpolation technique (described by *Chiodini et al.* [1998]). Together with the quantification of total CO₂ release, the definition of the related uncertainty is essential, especially in volcanic surveillance for the recognition of anomalous states. Kriging provides only an incomplete measure of local accuracy, except if a Gaussian model for errors is assumed, and no indication of the accuracy if several points are considered together [*Deutsch and Journal*, 1998]. On the contrary GSA approach allows the definition of a confidence interval for the estimation, but this calculation does not take into account the spatial correlation between the data, resulting generally in an overestimation of the uncertainty. This overestimation of the uncertainty is particularly unsuitable for monitoring purposes, since smaller uncertainty makes it possible to detect smaller variations.

[5] The aim of this work is the application of stochastic simulation algorithms to soil gas flux data. This approach is becoming common and preferred to traditional interpolation algorithms in the soil science, where the spatial variability of the measured attributes has to be preserved [*Goovaerts*, 2000], for example, in the definition and characterization of contaminated soils and groundwater [*Goovaerts*, 1997, 1999b, 2001, and references therein; *Lin and Chang*, 2000; *Lin et al.*, 2001; *Istok and Rautman*, 1996] and for assessment of corn yield risks connected to soil strength/compaction [*Lapen et al.*, 2001].

[6] The basic idea of stochastic simulation is to generate a set of equiprobable representations (realizations) of the spatial distribution of the attribute, all reproducing reasonably the global statistic and spatial features of data samples (i.e., sample histogram and semivariogram model), instead of producing a single representation that yields the minimum error variance at each location. The ensemble of these realizations is thus an explicit representation of the uncertainty associated with our conceptual understanding of the single, but unknown reality [*Rautman and Istok*, 1996]. According to *Goovaerts* [2001], differences among many simulated maps have been used as a measure of the uncertainty.

[7] In this paper a stochastic simulation algorithm is applied to 5 data set of soil CO₂ fluxes measured in different volcanic and nonvolcanic degassing areas, with the objective of mapping the degassing areas, i.e., defining the diffuse degassing structures, and evaluating the total emitted CO₂ with the associated uncertainty. Moreover, we try to define a reasonable criterion for the definition of an “opportunistic” sampling design.

2. Materials and Methods

2.1. Experimental Data Set

[8] The CO₂ flux data used in this paper were collected during the last decade at Solfatara of Pozzuoli crater (2000, Naples, Italy), Vesuvio cone (2000, Naples, Italy), Poggio dell'Olivo gas manifestation (1999, Viterbo, Italy), Nisyros caldera (1999–2001, Nisyros Island, Greece), and Horseshoe Lake (1997, Long Valley Caldera, California). Excluding the data of Solfatara of Pozzuoli and most of the Nisyros data that are unpublished, the remaining data were already treated and interpreted with different statistical and geostatistical tools [*Brombach et al.*, 2001; *Chiodini et al.*, 1999, 2001; *Rogie et al.*, 2001; *F. Frondini et al.*, Diffuse degassing at Vesuvio, Italy, submitted to *Bulletin of Volcanology*, 2003]. In all the study areas the survey of diffuse CO₂ flux from soil was performed by the accumulation chamber methodology. This methodology allows quick direct measurements of the CO₂ flux from soil without drastically altering the natural flux in a wide range of fluxes [*Chiodini et al.*, 1996, 1998; *Evans et al.*, 2001; *Welles et al.*, 2001]. The measurement is made over a surface of about 0.03 m² that can be considered a “point support” measurement.

[9] The data sets used for the application of stochastic simulation have been chosen for their different size, ranges of measured CO₂ flux values and for the differences in the sampling design used in each study area (Table 1 and Figure 1). In the case of Horseshoe Lake a subset of data has been selected from those previously published [*Rogie et al.*, 2001], to obtain a regularly shaped study area without large border areas with no data.

2.2. GSA Method

[10] In DDSs CO₂ flux from soil is fed by multiple gas sources such as biological and volcanic. This dual origin of the gas often results in a bimodal distribution of CO₂ flux values, which plots as a curve with an inflection point on logarithmic probability plots (see, e.g., Figures 3a, 3b, 3c, and 3e). On a logarithmic probability plot, a curve with an inflection point describes in fact the theoretical distribution of two overlapping lognormal populations, while a single lognormal population would plot as a “straight” line and n overlapping lognormal populations would result on a curve characterized by $n - 1$ inflection points.

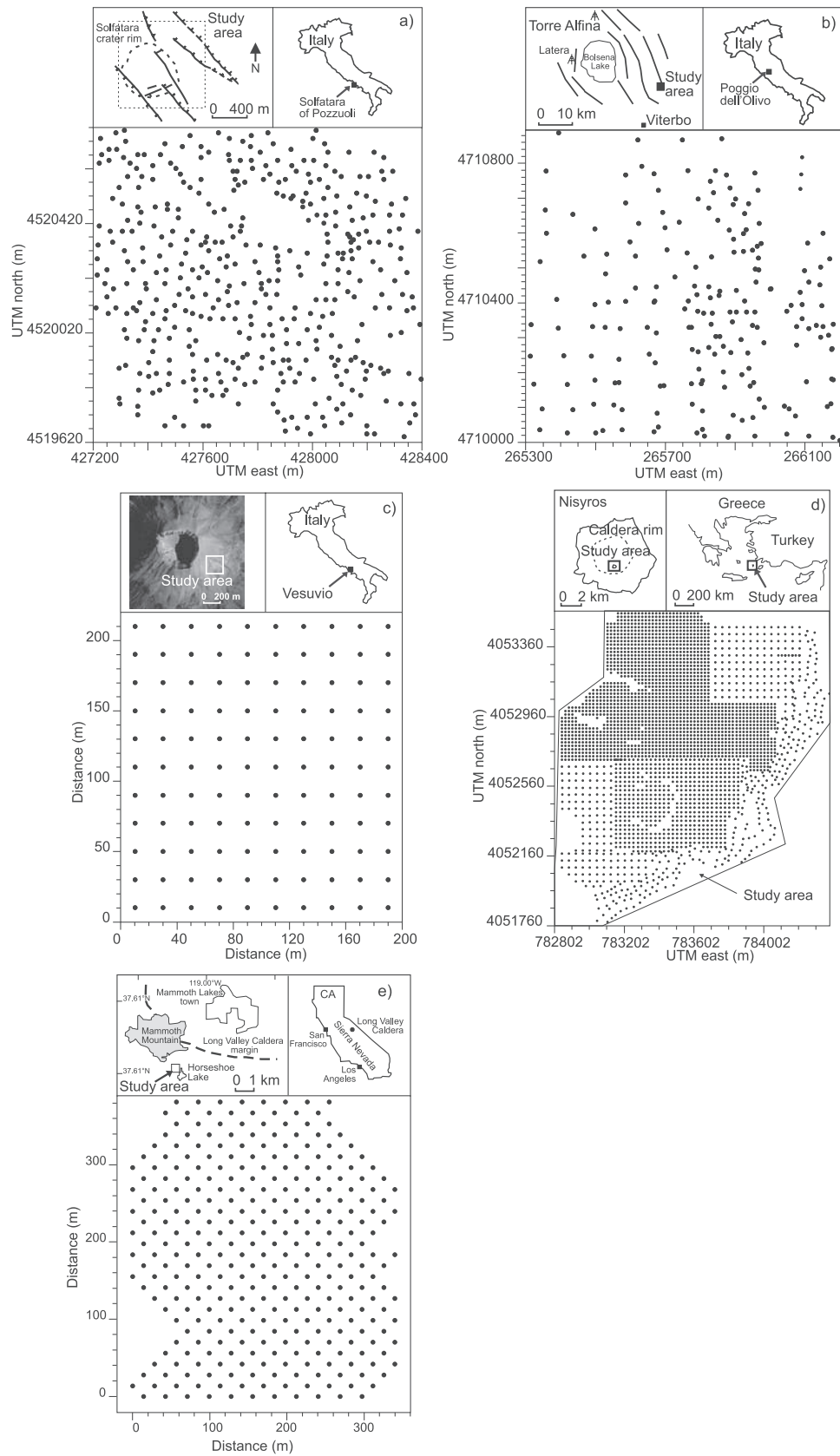


Figure 1. Location map of study areas and measurement point locations: (a) Solfatara of Pozzuoli; (b) Poggio dell'Olivio, (c) Vesuvio cone, (d) Nisyros caldera, and (e) Horseshoe Lake.

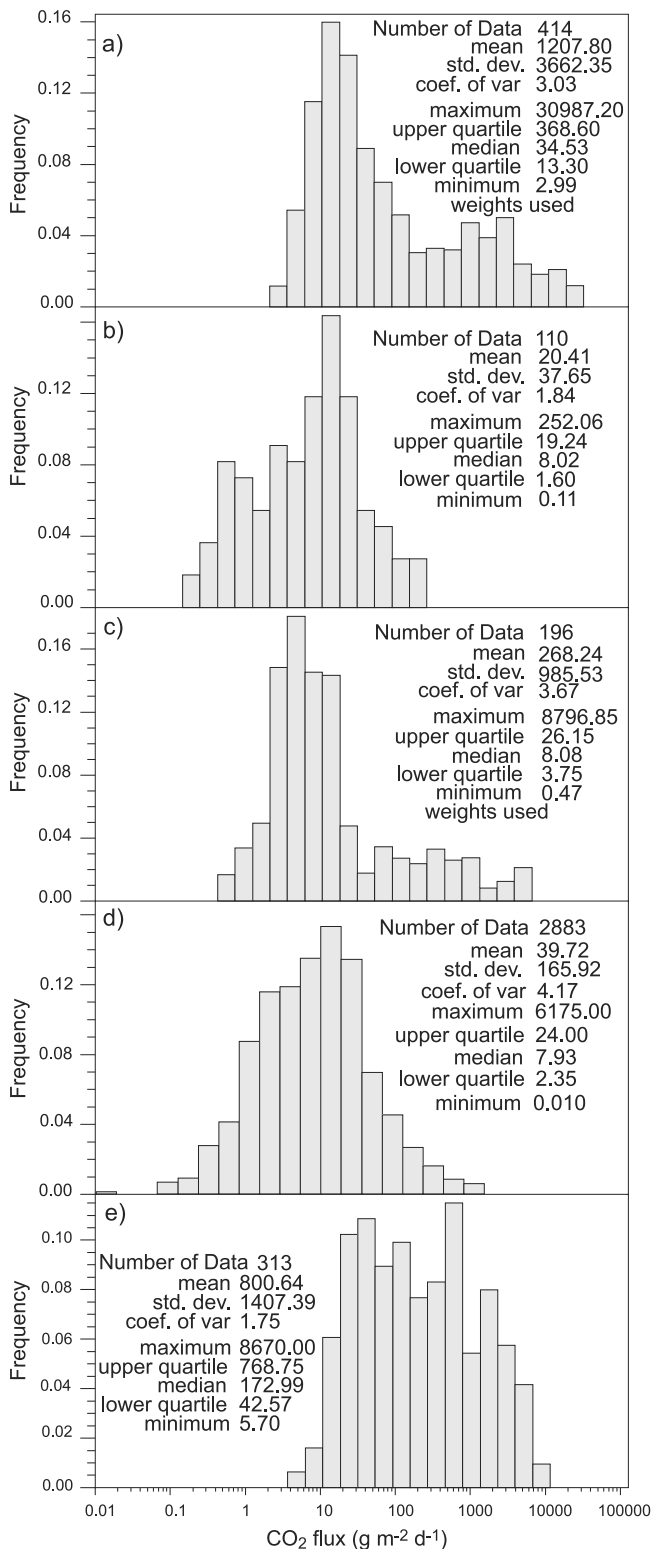


Figure 2. Histogram plot of CO₂ flux and summary of data statistics: (a) Solfatara of Pozzuoli; (b) Vesuvio cone, (c) Poggio dell'Olivo, (d) Nisyros caldera and (e) Horseshoe Lake.

[11] GSA method [Chiodini *et al.*, 1998] consists in the partition of these complex distributions into different log-normal populations and in the estimation of the proportion (f_i), the mean (M_i), and the standard deviation of each

population following the graphical procedure by Sinclair [1974]. Since the computed statistical parameters (i.e., mean, standard deviation and proportion) refer to the logarithm of values, the mean value of CO₂ flux and the central 90% confidence interval of the mean are estimated

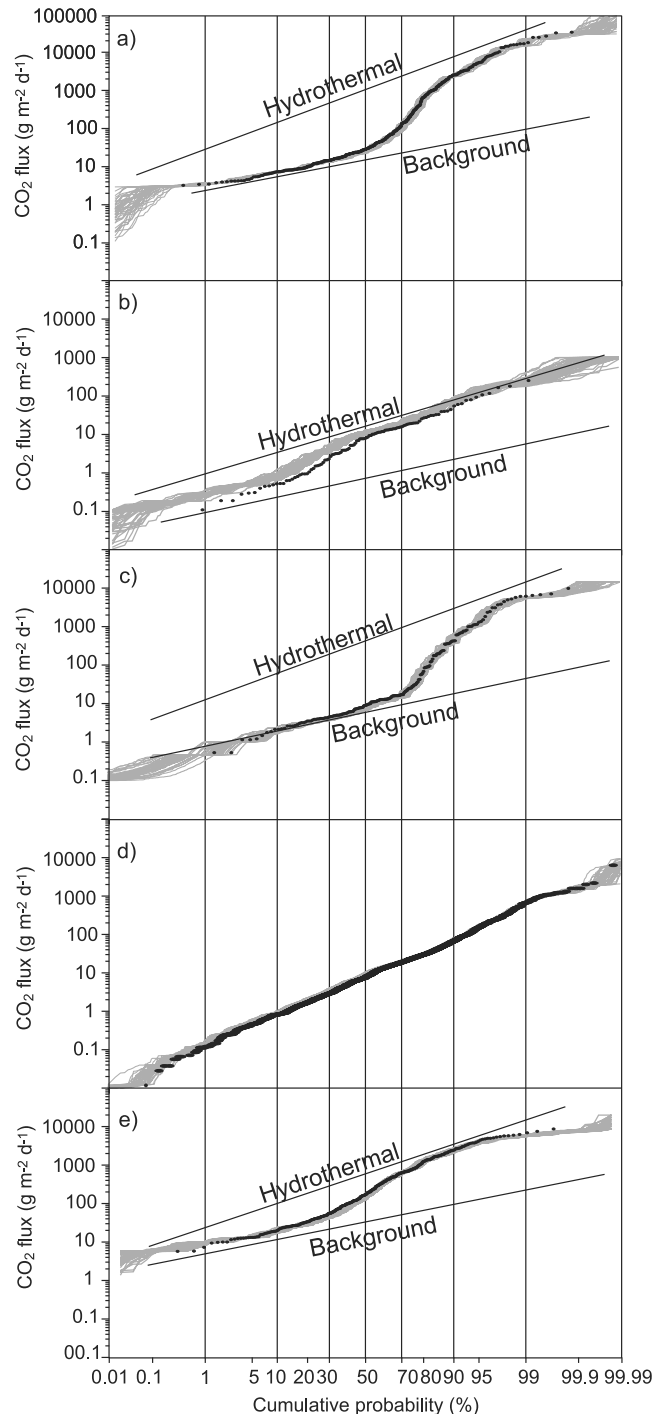


Figure 3. Probability plot of CO₂ flux: (a) Solfatara of Pozzuoli; (b) Vesuvio cone, (c) Poggio dell'Olivo, (d) Nisyros caldera, and (e) Horseshoe Lake. Figure 3 shows the original samples (dots), the theoretical partitioned populations following the procedure of Sinclair [1974] (black line), and the probability plots of simulated values relative to 50 realizations randomly selected for each area (shaded lines).

Table 2. Estimated Parameter of Partitioned Populations and Derived Total CO₂ Output

Study Area	CO ₂ Flux Population	Mean CO ₂ Flux, g m ⁻² d ⁻¹	Proportion, %	Total CO ₂ Output, t d ⁻¹	90% Confidence Interval, t d ⁻¹
Solfatara di Pozzuoli	background	23.92	61	23.4	20.7–27.3
	hydrothermal	3621.66	39	1521.1	1107.9–2308
Vesuvio cone	background	1.03	30	0.0094	0.0071–0.0174
	hydrothermal	29.50	80	0.73	0.6447–1.054
Poggio dell'Olivio	background	8.43	29	5.4	4.6–6.4
	hydrothermal	1373.23	71	247.7	153.7–544.8
Horseshoe Lake	background	46.85	62	2.3	1.6–4.2
	hydrothermal	1524.46	38	122.8	94.5–172.6

by means of the Sichel's t estimator [David, 1977]. The estimated mean flux values are used to compute the total CO₂ output associated to each population. An evaluation of the area covered by each population (S_i) is obtained by multiplying the study area (S) by the corresponding proportion of the population (i.e., $S_i = f_i S$). The total CO₂ output associated to each population is then estimated multiplying S_i by M_i . The total CO₂ release from the entire studied area can be obtained summing the contribution of each population (i.e., $\sum S_i M_i$). In the same way, the central 90% confidence interval of the mean are used to calculate the uncertainty of the total CO₂ output estimation of each population.

[12] Even if the GSA approach resulted a useful tool for the interpretation of the diffuse degassing process [e.g., Chiodini *et al.*, 1998, 2001], the results obtained by the GSA can be affected by some arbitrary choices: (1) The polymodal lognormal distribution of CO₂ flux values is a model convenient for the subsequent decomposition, but it is not a fact (natural distribution of CO₂ flux can be more complex than lognormal), and (2) the partitioning procedure does not result into a unique solution. Even to avoid these limits, an alternative approach based on stochastic simulation has been investigated.

2.3. Stochastic Simulation

[13] The simulations were performed using the sequential Gaussian simulation algorithm (sGs) by the program *sgsim* [Deutsch and Journel, 1998]. SGs operates considering an attribute (CO₂ flux from soil in this study) as the realization of a stationary multivariate Gaussian random function. The attribute values are simulated at locations defined by a grid covering the area of interest. The simulation is conditional and sequential, i.e., the variable is simulated at each unsampled location by random sampling of a Gaussian conditional cumulative distribution function defined on the basis of original data and of previously simulated data within its neighborhood.

[14] As the sGs procedure needs a multigaussian distribution, which implies first that the one-point distribution of data (i.e., histogram of data) is normal, CO₂ flux data that generally are positively skewed, have to be transformed into a normal distribution (normal scores of data). The transformation consists in substituting the original values by the corresponding quantiles of a standard normal distribution. Moreover, the normality of two-point cumulative distribution function of normal scores has to be checked for the application of the multigaussian model.

[15] The transformed data are then used in the simulation procedure. In practice, due to the multigaussian assumption, mean and variance of the Gaussian conditional

cumulative distribution function can be defined at each location as the simple kriging estimate and variance respectively [Goovaerts, 1997; Deutsch and Journel, 1998]. Simple kriging estimate and variance are computed according to the semivariogram model of normal scores. Defining the cumulative distribution function at one location means knowing the probability of any possible value to characterize that location. A random value is drawn from the conditional cumulative distribution as one "reasonable" simulated value for that location [Rautman and Istok, 1996; Castrignano *et al.*, 2002]. Once a value is simulated, it is added to the data set and can be used together with the original data to estimate the variable at the next locations of the grid. The simulation proceeds to the next grid location and loop until all nodes are simulated. Afterward, the simulated normal scores are back transformed in to values expressed in original data unit, applying the inverse of the normal score transform. This back transforming procedure needs the setting of an upper tail and of a lower tail, which represent the maximum and the minimum values allowable for the simulated values, and the definition of the extrapolation model between the original data limits and upper and lower tails. Changing the random path of grid nodes visited, by changing the starting point of simulation procedure (random-number seed), N alternative simulations can be performed and N equiprobable realizations can be drawn, each honoring the sampled data at their locations, reproducing the data univariate statistics (histogram) and the data bivariate property (experimental variogram of normal scores), within reasonable ergodic fluctuations [Deutsch and Journel, 1998].

3. Applications and Results

3.1. Treatment of the Data With the GSA Method

[16] The data histograms (Figure 2) and logarithmic probability plots (Figure 3) show two main maximums and curves with one inflection point respectively, for the areas of Solfatara di Pozzuoli, Poggio dell'Olivio, Horseshoe Lake and Vesuvio. These distributions can be interpreted as the combinations of two lognormally distributed populations of CO₂ flux. The proportion, the mean, the total CO₂ output, and the relative 90% confidence interval of each population (Table 2) were computed following the GSA method above described. The validity of the two-populations model was checked through calculation of ideal combinations of partitioned populations until a satisfactory agreement between ideal mixtures and real data was obtained. The observed bimodal distributions reflect the coexistence of (1) low CO₂ fluxes, generally connected to the biological activity in the soil (background population,

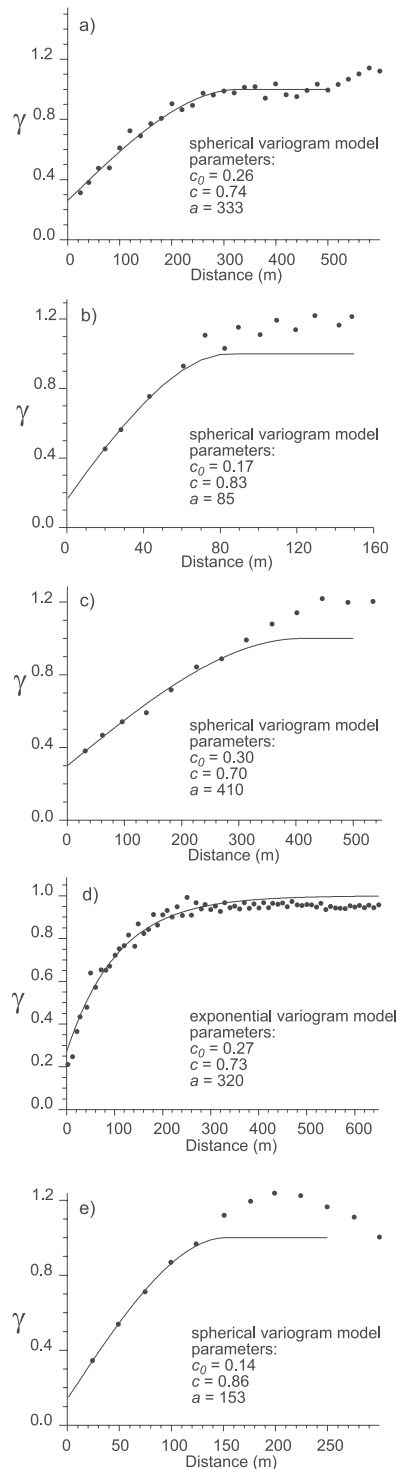


Figure 4. Omnidirectional experimental variograms (γ) of CO₂ flux normal scores: (a) Solfatara of Pozzuoli; (b) Vesuvio cone, (c) Poggio dell'Olivo, (d) Nisyros caldera, and (e) Horseshoe Lake. Lines represent the isotropic variogram models used in the simulation procedure. The parameters c_0 (nugget effect), c (sill) and a (range) refer to the variogram models (equations (1), (2) and (3)).

Table 2), and (2) high CO₂ fluxes connected to the degassing processes of deeply derived CO₂ (hydrothermal population, Table 2). The mean CO₂ flux of background populations ranges from 1 g m⁻² d⁻¹ at Vesuvio cone, where biological activity is very low due to the absence of vegetation and of a soil rich in organic matter, to 47 g m⁻² d⁻¹ at Horseshoe Lake, where the high background value is likely due to the contribution of some diffusive degassing of deep sources. The “hydrothermal” populations are generally characterized by mean values of CO₂ fluxes (1373–3621 g m⁻² d⁻¹) 2 orders of magnitude higher than the background. However, at Vesuvio the mean CO₂ flux of the “hydrothermal” population is very low (29.5 g m⁻² d⁻¹); in this case the anomalous population is recognizable because of the very low value of the background.

[17] At Nisyros caldera, CO₂ fluxes distribute in a wide range from 0.01 to 6000 g m⁻² d⁻¹ suggesting also in this case the contribution of different sources. However, in the probability plot the log values do not fit a clear polymodal curve and the definition of the background population based on the partitioning of statistical populations is not possible. The absence of a clear polymodal distribution is most probably due to the presence of a widespread low level degassing of deeply derived CO₂ linked to the several degassing structures active in the area, i.e., phreatic craters and faults. Such features are absent in the south east sector of the area. In this zone, the values of log CO₂ flux distributed in the probability plot (not shown here) along a straight line suggesting the presence of a unique population, characterized by a mean value of 8 g m⁻² d⁻¹, which has been assumed as the biological background affecting the entire Nisyros caldera.

3.2. Applications of Sequential Gaussian Simulation

[18] Because of the nonnormality each data set has been converted in to a distribution with a mean of 0 and a unit variance, a normal scores transform, using the program *nscore* [Deutsch and Journel, 1998]. For each study area the variogram (γ) of normal scores (Figure 4) has been computed and fitted with a standardized spherical model, described by the following equations:

$$\gamma = c_0 + c \left[1.5 \left(\frac{h}{a} \right) - 0.5 \left(\frac{h}{a} \right)^3 \right], h < a \quad (1)$$

$$\gamma = 1, \quad h > a, \quad (2)$$

except for Nisyros where the best fitting results a standardized exponential model, described by the equation

$$\gamma = c_0 + \left[1 - \exp \left(-\frac{3h}{a} \right) \right], \quad (3)$$

where c_0 is the nugget effect, c is the sill, a is the range and h is the distance between samples [e.g., Isaaks and Srivastava, 1989]. The variogram models parameters used for each study area are reported in Figure 4.

[19] One hundred simulations have been performed for each data set using simulation grids of different spacing between nodes: 10 m for Solfatara of Pozzuoli and Nisyros,

Table 3. Summary Statistics of Measured CO₂ Flux (g m⁻² d⁻¹) and Simulated CO₂ Flux

Study Area	Original Data		Simulated Values (50 Realizations)			
	Mean	Standard Deviation	Mean	Standard Deviation	Maximum	Minimum
Solfatara	1207 ^a	3662.3 ^a	855–1260	2774–4873	42312–147600	0.008–2.2
Vesuvio cone	20	37.6	22.12–35.72	36.2–89.0	556–993	0.009–0.11
Poggio dell'Olivo	268 ^a	985.5 ^a	218–315	859–1248	12836–14996	0.10–0.16
Nisyros caldera	39.7	165.9	38.4–47.1	129.7–222.4	6175–9715	0.0003–0.01
Horseshoe Lake	800	1407.7	694–895	1186–1494	8670–17672	1.2–5.7

^aComputed from declustered data.

5 m for Horseshoe Lake and Poggio dell'Olivo and 2 m for the Vesuvio. Even if the values are simulated over a “point support,” they have been considered representatives of square cells centered on the grid nodes and with the side length equal to the grid spacing. The cell size was chosen, taking into account the spacing between the data and the dimension of the surveyed area, in order to obtain a good resolution in the mapping and, at the same time, to avoid a redundant number of cells limiting the CPU and memory requirements.

[20] The basic statistics, the cumulative distribution and the variogram of simulated values have been compared to those of the original data to assess the performance of simulation procedure in terms of reproduction of the features of the original data. In the case of Solfatara of Pozzuoli and Poggio dell'Olivo the statistics and the cumulative distribution of original data have been computed considering the declustering weights that have been used also in the simulation.

[21] The basic statistic (mean and standard deviation) of data samples is well reproduced by the simulated values (Table 3). This fitting is not obvious because there is no guarantee that the original sample statistics are exactly those of the broader population of all possible samples [Rautman and Istok, 1996]. A good match is found also between the cumulative distribution of the original data and simulated CO₂ fluxes from 50 randomly selected realizations, in particular the bimodal distributions of CO₂ flux values are well reproduced (Figure 3). Furthermore, the undesired smoothing effect on extreme values, produced by many interpolation algorithms (e.g., kriging), is avoided.

[22] The results obtained by the simulations are influenced by the choice of upper and lower tail and by the extrapolation mode between the original data limits and upper and lower tail. For each application the lower tail has been set to 0, 0 being the natural lowest limit of soil CO₂ fluxes. The upper tail has been estimated referring to the specific probability plots and roughly extrapolating the curves to the quantile corresponding to the maximum number of simulated values. A linear model has been used for the extrapolation of the data between the minimum measured value and the lower tail, while a hyperbolic model with ω set to 1.5 (ω = *utpar* parameter in *sgsim* code) has been used for the extrapolation of the data between the maximum measured value and the upper tail. Practice has shown that, for positively skewed distributions, the hyperbolic upper tail distribution with ω = 1.5 is a general-purpose model that yields acceptable results in a wide variety of applications [Goovaerts, 1997; Deutsch and Journal, 1998]. The consequence of upper and lower tail choice is that the maximum and minimum simulated values

are higher and lower, respectively, than the measured values, as expected because there is a very low probability that the true maximum and minimum flux values are measured in the survey (Tables 1 and 3 and Figure 2).

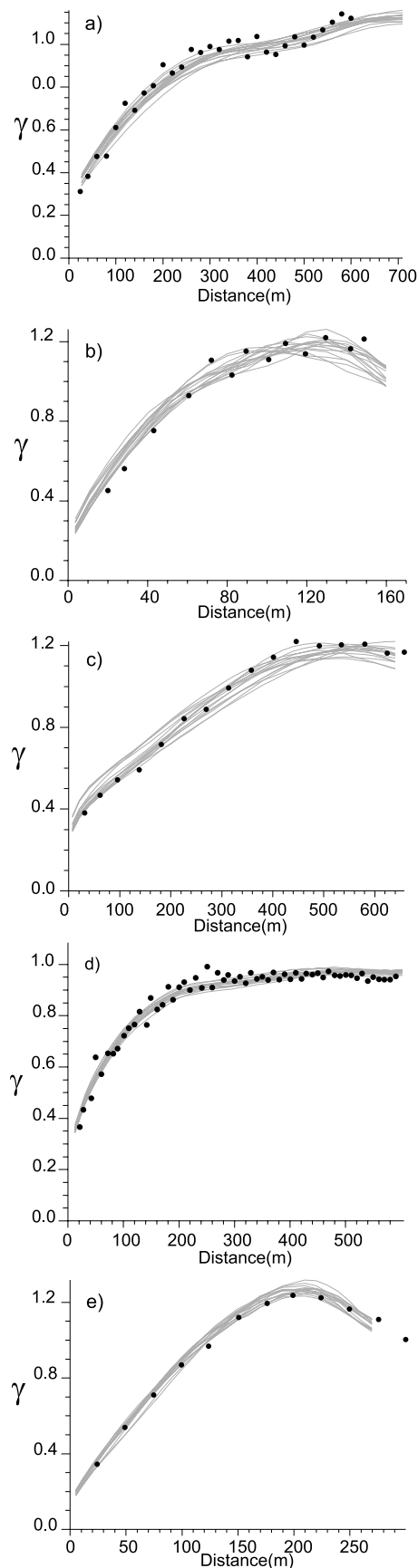
[23] Figure 5 compares normal scores omnidirectional variograms of original data with those computed for 15 randomly selected realizations of each study area. There is a good agreement in the spatial continuity between the sample and the realizations variograms. The variograms of different realizations show some fluctuations (Figure 5); these fluctuations are common and are referred to “ergodic fluctuations” [Goovaerts, 1997, 1999a; Deutsch and Journal, 1998]. Ergodic fluctuations are caused by the discrepancy between the realization statistics and the corresponding model parameter [Deutsch and Journal, 1998], and their magnitude is related to the relative dimension of the simulated domain with respect to the range of the model variogram [Goovaerts, 1999a; Rautman and Istok, 1996], i.e., tends to decrease as the ratio between the size of the modeled domain and the range of the variogram increases. Deutsch and Journal [1998] suggest that ergodic fluctuations generally are minor when the size of the modeled domain is in excess of 10 times the range of correlation.

3.2.1. Probabilistic Summary of a Set of Simulations: Mapping of the Diffuse Degassing Structures

[24] The different level of reproduction of sample statistics may lead one to choose the realization fitting best the imposed statistics for mapping the CO₂ flux. This choice would be correct only if we consider the original sample statistics exact and unquestionable, but this is not true, hence all the realizations should be considered equiprobable [Deutsch and Journal, 1998].

[25] One of the possible post processing of the sets of realizations to define the extent of the diffuse degassing structures (DDS) consists of drawing maps of the probability that the CO₂ flux at any location is above a cutoff value. The probability is computed at any location from the proportion of all simulated values above the cutoff at that location.

[26] Figure 6 gives the maps of the probability of exceeding a CO₂ flux value, which was assumed to be a reasonable upper limit for a background CO₂ flux. Different cutoff values have been chosen for each area from inspection of the probability plots (Figure 3) to discriminate values belonging to the background populations from those of the anomalous populations. More in detail, cutoff values of 2, 50, 20 and 100 g m⁻² d⁻¹ have been selected for Vesuvio cone, Solfatara of Pozzuoli, Poggio dell'Olivo and Horseshoe Lake, respectively. Only 10% of background samples and a number ranging from 90% to 98% of the anomalous populations fall above these cutoff values. A



cutoff value of $20 \text{ g m}^{-2} \text{ d}^{-1}$, i.e., 2 times higher than the mean estimated for the background CO₂ flux, has been chosen for Nisyros caldera where the absence of a clear bimodal distribution did not permit a more sophisticated analysis (see section 3.1).

[27] The maps of Figure 6 represent the probability that each location belongs to the DDS. In general the DDS are controlled by tectonic and volcanic structures: At Solfatara of Pozzuoli, the DDS is connected to the presence of NW-SE regional trending faults and to ENE-WSW fractures generated by pressure variations within the underlying hydrothermal system [Chiodini *et al.*, 2001], at Nisyros the DDS match the NE-SW and NNE-SSW regional faults and the ancient hydrothermal craters, and at Poggio dell'Olivio the DDS develops along a NW-SE fault [Chiodini *et al.*, 1999]. The straight boundaries of the Horseshoe Lake DDS suggest that in this case the degassing process is controlled by tectonics, while the small Vesuvio anomaly is coincident with an ancient crater rim of the volcano.

3.2.2. Quantification of Total CO₂ Output

[28] The quantitative estimation of the total amount of CO₂ released from a DDS is a primary objective of our study and in particular of researches devoted to surveillance of active volcanoes. In this framework, it is important to quantify the uncertainty of the total CO₂ output quantification, for a correct interpretation of the temporal variations. The total CO₂ output is computed for each realization by summing the products of simulated value of each grid cell by the cell surface. The mean and the standard deviation of the 100 simulated values of total CO₂ output, computed for the 100 realizations, are assumed to be the characteristic values of the CO₂ release and of its uncertainty for each area (Table 4). Total CO₂ releases vary from 0.95 t d^{-1} , estimated for Vesuvio DDS, to 1500 t d^{-1} computed for Solfatara DDS, and the uncertainties are always lower than $\pm 12\%$ of the total CO₂ output.

[29] The variation of the total CO₂ output computation is visually appreciable in Figure 7 where a map of the simulated "expected" values at any cell (E-type estimates, obtained through a pointwise linear average of all the realizations [Deutsch and Journel, 1998]) is shown for each area together with two simulated maps relative to the realization giving the highest and the lowest value of total CO₂ output.

[30] The total CO₂ outputs, estimated by sGs method, do not differ significantly from those estimated by GSA, however there are main differences among uncertainties estimated by the two methods. For a comparison of the uncertainties, the 5th and 95th percentiles relative to the 100 simulated values of total CO₂ output, are reported in Table 4. These two percentile values are the limits of the central 90% of the total CO₂ output simulated values and permit an easy comparison with the central 90% confidence

Figure 5. (opposite) Variograms of normal scores of CO₂ flux: (a) Solfatara of Pozzuoli; (b) Vesuvio cone, (c) Poggio dell'Olivio, (d) Nisyros caldera, and (e) Horseshoe Lake. The variograms (γ) of normal scores of original values (dots) are compared to the variograms of normal scores of simulated values (shaded lines) of 15 realization randomly selected for each area.

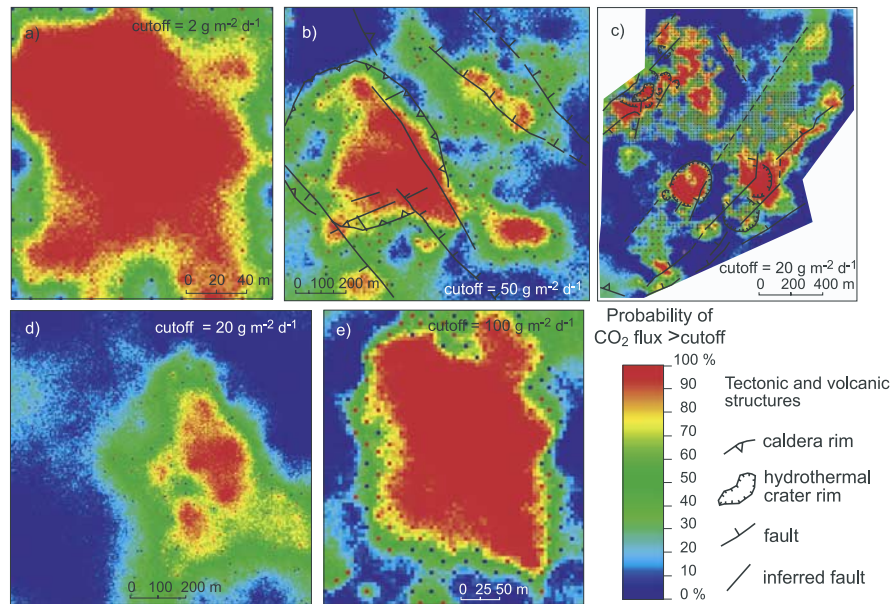


Figure 6. Probability maps of CO₂ flux: (a) Vesuvio cone; (b) Solfatara of Pozzuoli, (c) Nisyros caldera, (d) Poggio dell'Olivio, and (e) Horseshoe Lake. Color scale shows the probability of CO₂ flux exceeding specific threshold selected to discriminate background from anomalous CO₂ degassing (see the text).

interval estimated by GSA. In any case, the uncertainties estimated by GSA method are much higher than those estimated by sGs method, being the 5th–95th percentile range, computed by sGs, much lower than the 90% confidence interval estimated through the GSA approach.

3.2.3. Total CO₂ Output Uncertainty and Sampling Design

[31] There are many factors influencing the magnitude of the error of an estimate, e.g., the number of samples, their spatial arrangement, the nature of phenomena under study, etc. [Isaaks and Srivastava, 1989]. In this section we approach this problem and try to derive empirical criteria for the definition of a reasonable sampling design to obtain an estimation of the total CO₂ output with an acceptable uncertainty. On the basis of the entire sample data, the computed total CO₂ output oscillates from $\pm 5\%$ to $\pm 12\%$ in the different areas (Table 4).

[32] In order to investigate the relation between the uncertainty and number of conditioning data, several subsets with different number of samples were randomly created from the original data, without giving consideration to whether or not the sample statistics of the subsets match those of the complete data set. For each subset, 100 simulations have been performed and the mean CO₂ total output and the standard deviation have been computed, as described in the previous section.

[33] In all cases, the mean total CO₂ output seems to be relatively insensitive to the number of conditioning data until that number exceeds a threshold which, in each area, is lower than 50% of the available samples (Figure 8). Below this limit, variable for each area, the computed total CO₂ output starts to oscillate intensively. This finding (1) confirms that our experimental data sets can be considered exhaustive and (2) suggests that in each area there exists a minimum number of samples below which the conditioning

data are inadequate to describe the study field and to provide a reliable quantitative estimation of the total CO₂ output.

[34] Figure 9a shows that the standard deviation of the mean total CO₂ output is linked with a hyperbolic function to the number of measurements. However, the correlation between number of measurements and uncertainty is specific to each survey and can not be used as a general law to design sampling of new areas.

[35] A more general relation can be derived considering the number of samples falling in the area contained by a circle with radius equal to the range of the CO₂ flux variogram (circle range area (CRA)) instead of the total samples number (Figure 9b). In this case the same hyperbolic function can adequately fit the data of the five different surveys, suggesting that a suitable sampling density can be derived from the range of correlation of the parameter under study. Figure 9b, for example, suggests that uncertainty lower than 10% are obtained with a sampling density that guarantees at least 90 samples in the CRA.

[36] The empirical relation derived from Figure 9b can be a useful tool for the design of a soil CO₂ flux survey. For example, a suitable sampling design, to obtain a reliable estimation of the total CO₂ output, can be defined

Table 4. Total CO₂ Output and Uncertainties Derived From 100 sGs Realizations

Study Area	Mean, $t\ d^{-1}$	Standard Deviation, $t\ d^{-1}$	5th–95th Percentiles, $t\ d^{-1}$
Solfatara of Pozzuoli	1513.00	136.00	1266.4–1729.7
Vesuvio cone	0.95	0.11	0.77–1.14
Poggio dell'Olivio	233.50	27.90	191.9–274.6
Nisyros caldera	84.00	3.82	79.80–89.47
Horseshoe Lake	104.30	5.00	96.20–112.99

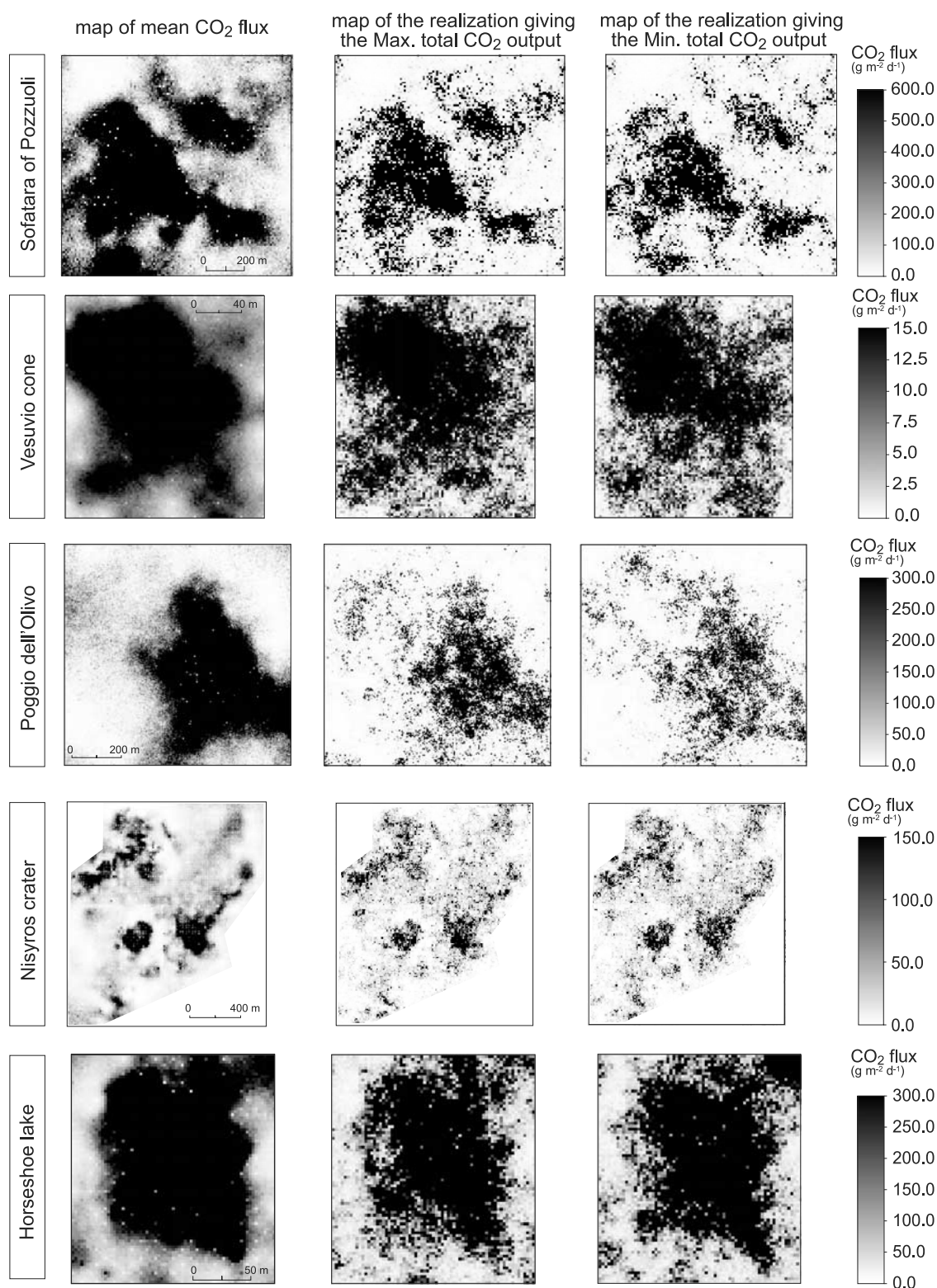


Figure 7. CO₂ flux maps. For each area are reported, from the left to the right, the map of mean CO₂ flux obtained from a pointwise linear averaging of all 100 the realizations (E-type estimates), the map relative to the realization giving the maximum estimate of the total CO₂ output, and the map relative to the realization giving the minimum estimate of the total CO₂ output are shown.

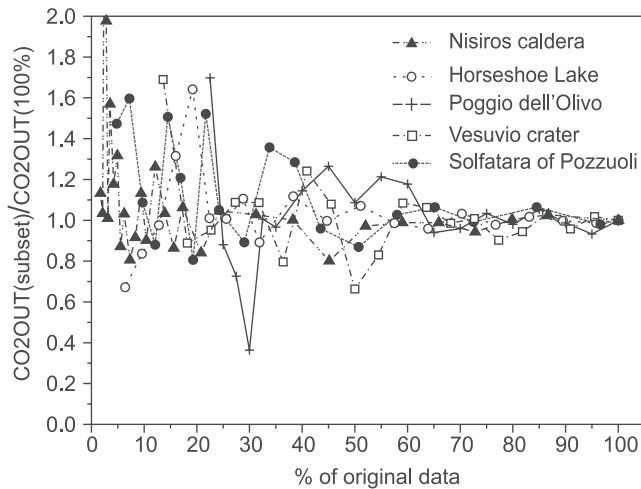


Figure 8. Variation of the total CO₂ output estimate as a function of the number of conditioning data reported as percentage of the total sample number. The total CO₂ outputs computed from each subset of data are standardized to the total CO₂ output computed from the entire data set of the corresponding area (CO₂OUT(subset)/CO₂OUT(100%)).

after a quick, preliminary survey aimed to roughly define the range of the correlation of soil CO₂ fluxes. In the absence of any information on soil CO₂ flux values, an approximate estimate of the CRA and consequently of the suitable sampling density can be done considering that (1) the CRA extension is close to the spatial extent of degassing structures (Figure 10) and (2) in many cases that extent can be evaluated “a priori” on the basis of macroscopic field evidences (e.g., bare soils, focused emission of gas, dead vegetation, etc.).

4. Conclusions

[37] The application of the stochastic simulation methodology, sequential Gaussian simulation, to the diffuse CO₂ from soils can constitute a valid approach to the definition and characterization of diffuse degassing structures, DDS, affecting active volcanoes and geothermal areas. The approach, based on the production of several equiprobable images of the CO₂ flux, permits both mapping of the DDS and quantification of the total CO₂ output from surveyed areas. Moreover, the production of a set of equiprobable realizations of the same phenomena allows us to quantify the uncertainty of our representation of the reality.

[38] The method has been applied to four volcanic areas (Solfatara di Pozzuoli, Nisyros caldera, Horseshoe Lake, and Vesuvio) and to one nonvolcanic area (Poggio Olivo), all of which are characterized by active CO₂ degassing processes. For each data set, 100 simulations have been produced. The simulated values at unsampled locations reproduce well the univariate and bivariate statistics of the original samples, suggesting that the sGs method produces more realistic representations than the traditional estimation technique, kriging, which reproduces neither the univariate (histogram) nor the bivariate (variogram) statistics of the soil diffuse CO₂ fluxes.

[39] On the basis of sGs results, different types of maps can be chosen to highlight the DDS. In this work we show the maps relative to the maximum and minimum CO₂ output realizations and the map of the mean values (E-type). However, a more appropriate visualization of the DDS is obtained through probability maps. Soil CO₂ fluxes are typically characterized by polymodal density distribution, generally consisting in the combination of low CO₂ fluxes, connected to the biological activity in the soil (background population), and high CO₂ flux populations generated by degassing processes of deeply derived CO₂. On the basis of suitable threshold values, chosen from probability plots of samples data, the probability that values exceed this threshold can be mapped for each area. Because the threshold values reasonably divide background from deeply derived CO₂ fluxes, these maps represent the probability of each “location” to belong to the DDS. The soil CO₂ flux maps of the studied areas both define the extension DDS and emphasize a strict control of tectonic-volcanic structures on the degassing process.

[40] The total amount of CO₂ released through diffuse degassing has been estimated from each simulated realization attributing to the surface of each grid cell the

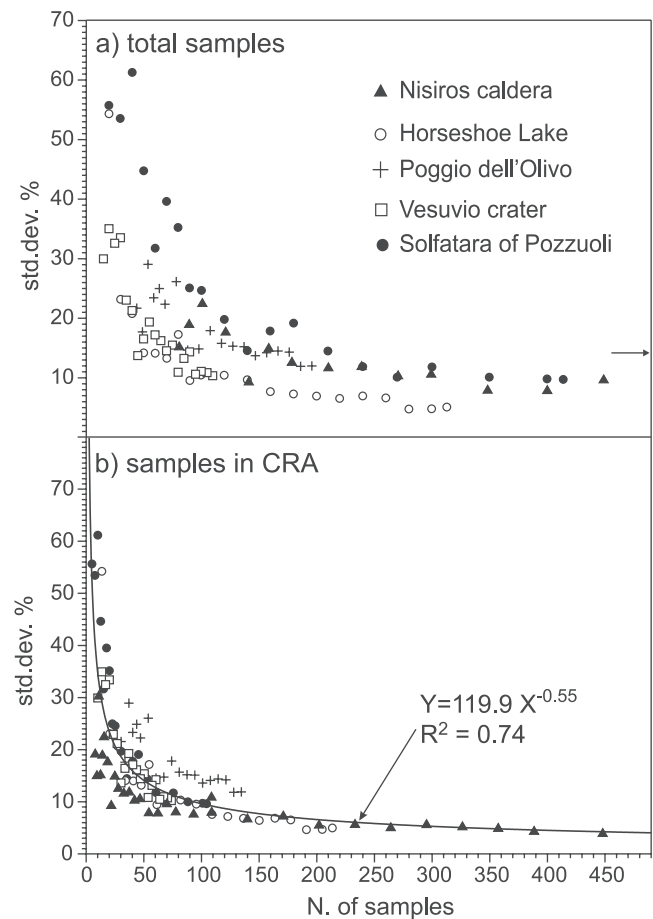


Figure 9. (a) Uncertainty of total CO₂ output versus number of conditioning data. (b) Uncertainty of total CO₂ versus number of conditioning data falling in a circle with radius equal to the range of the CO₂ flux variogram (CRA) and the best fitting function.

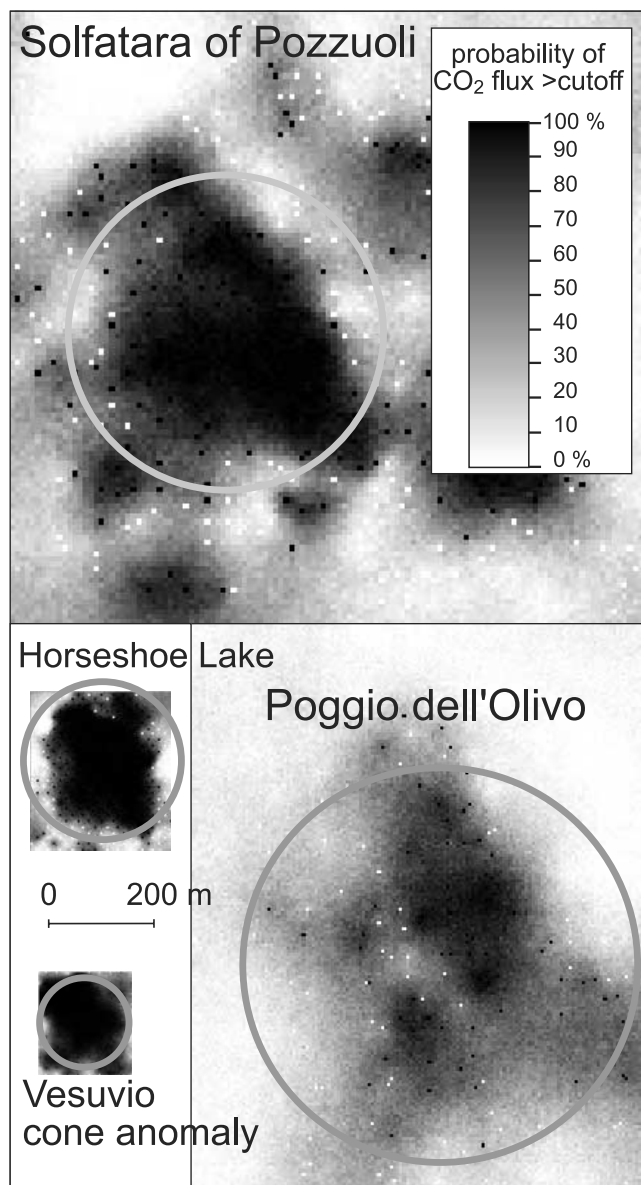


Figure 10. The dimension of the circles with radius equal to the range of the CO₂ flux variograms (CRA) are compared to the probability maps of CO₂. Grey scale shows the probability of CO₂ flux exceeding specific threshold (the same used in Figure 6) and highlights diffuse degassing structures (DDS). The extension of the CRAs approximately matches the dimension of the DDSs.

corresponding simulated flux value, creating a set of equiprobable values for the total CO₂ output. The mean total CO₂ output of the 100 realizations is assumed as the “characteristic” value for each study area. The uncertainty associated to this estimation is taken as the standard deviation of all the possible values obtained from the 100 realizations. Solfatara of Pozzuoli releases through diffuse degassing $1513 \pm 136 \text{ t d}^{-1}$ of CO₂, Poggio dell’Olivio releases $233 \pm 28 \text{ t d}^{-1}$, Nisyros caldera releases $84 \pm 3.82 \text{ t d}^{-1}$, Horseshoe Lake releases $104 \pm 5 \text{ t d}^{-1}$ and the small degassing area on the Vesuvio cone flank releases $0.95 \pm 0.11 \text{ t d}^{-1}$. The characteristic values of simulated

total CO₂ output from the different areas do not differ significantly from the values previously estimated with the GSA method, but the most interesting aspect of sGs method is the assessment of the uncertainty.

[41] There are many factors influencing the magnitude of the uncertainty of an estimate, most of them are not preventable. However, to obtain a reliable estimation of the total CO₂ output through sGs we can at least adopt an appropriate sampling design for the CO₂ flux survey. The results of this study suggest that a suitable sampling density is related to the range of correlation of the parameter under study. On the basis of the results on the five different areas, an empirical relation between the expected uncertainty and the number of the samples which fall in the CRA (an area corresponding to the circle of radius equal to the range of correlation of the CO₂ fluxes defined by the variogram) has been derived. This relation may represent a useful tool for planning future CO₂ flux surveys. This paper focuses the attention of CO₂ flux surveyors on the necessity of choosing tools that allows some assessment of estimation uncertainty, in order to efficiently use soil diffuse degassing measurements in the framework of volcanic surveillance.

[42] **Acknowledgments.** This work was financially supported by EU (GEOWARN Project IST 1999-12310), by GNV-INGV (Volcanology National Group of Italy) and by MIUR (GEOCO2 Project).

References

- Allard, P., et al., Eruptive and diffuse emissions of CO₂ from Mount Etna, *Nature*, 351(6325), 387–391, 1991.
- Bergfeld, D., F. Goff, and C. J. Janik, Elevated carbon dioxide flux at the Dixie Valley geothermal field, Nevada: Relations between surface phenomena and the geothermal reservoir, *Chem. Geol.*, 177(1–2), 43–66, 2001.
- Brantley, S. L., and K. W. Koepnick, Measured carbon-dioxide emissions from Oldoinyo-Lengai and the skewed distribution of passive volcanic fluxes, *Geology*, 23(10), 933–936, 1995.
- Brombach, T., J. C. Hunziker, G. Chiodini, C. Cardellini, and L. Marini, Soil diffuse degassing and thermal energy fluxes from the southern Lakki plain, Nisyros (Greece), *Geophys. Res. Lett.*, 28(1), 69–72, 2001.
- Castrignano, A., N. Lopez, M. Prudeniano, and P. Steduto, Estimation or stochastic simulation in soil science?, paper presented at 7th International Meeting on Soils With Mediterranean Type of Climate: Preserving the Mediterranean Soils in the Third Millenium, Soc. Ital. Della Sci. Della Suolo, Valenzano, Bari, Italy, 2002.
- Chiodini, G., F. Frondini, and B. Raco, Diffuse emission of CO₂ from the Fossa crater, Vulcano Island (Italy), *Bull. Volcanol.*, 58, 41–50, 1996.
- Chiodini, G., R. Cioni, M. Guidi, B. Raco, and L. Marini, Soil CO₂ flux measurements in volcanic and geothermal areas, *Appl. Geochem.*, 13(5), 543–552, 1998.
- Chiodini, G., F. Frondini, D. M. Kerrick, J. Rogie, F. Parello, L. Peruzzi, and A. R. Zanzari, Quantification of deep CO₂ fluxes from central Italy: Examples of carbon balance for regional aquifers and of soil diffuse degassing, *Chem. Geol.*, 159(1–4), 205–222, 1999.
- Chiodini, G., F. Frondini, C. Cardellini, D. Granieri, L. Marini, and G. Ventura, CO₂ degassing and energy release at Solfatara volcano, Campi Flegrei, Italy, *J. Geophys. Res.*, 106(B8), 16,213–16,221, 2001.
- David, M., *Geostatistical Ore Reserve Estimation*, 364 pp., Elsevier Sci., New York, 1977.
- Deutsch, C. V., and A. G. Journel, *GSLIB: Geostatistical Software Library and Users Guide*, 369 pp., Oxford Univ. Press, New York, 1998.
- Etiopie, G., Subsoil CO₂ and CH₄ and their advective transfer from faulted grassland to the atmosphere, *J. Geophys. Res.*, 104(D14), 16,889–16,894, 1999.
- Etiopie, G., P. Beneduce, M. Calcara, P. Favali, F. Frugoni, M. Schiattarella, and G. Smriglio, Structural pattern and CO₂-CH₄ degassing of Ustica Island, southern Tyrrhenian basin, *J. Volcanol. Geotherm. Res.*, 88(4), 291–304, 1999.
- Evans, W. C., M. L. Sorey, B. M. Kennedy, D. A. Stonestrom, J. D. Rogie, and D. L. Shuster, High CO₂ emissions through porous media: Transport mechanisms and implications for flux measurement and fractionation, *Chem. Geol.*, 177(1–2), 15–29, 2001.

- Farrar, C. D., M. L. Sorey, W. C. Evans, J. F. Howle, B. D. Kerr, B. M. Kennedy, C. Y. King, and J. R. Southon, Forest-killing diffuse CO₂ emission at Mammoth Mountain as a sign of magmatic unrest, *Nature*, 376(6542), 675–678, 1995.
- Gerlach, T. M., M. P. Doukas, K. A. McGee, and R. Kessler, Soil efflux and total emission rates of magmatic CO₂ at the Horseshoe Lake tree kill, Mammoth Mountain, California, 1995–1999, *Chem. Geol.*, 177(1–2), 101–116, 2001.
- Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford Univ. Press, New York, 1997.
- Goovaerts, P., Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties, *Stochastic Environ. Res. Risk Assess.*, 13(3), 161–182, 1999a.
- Goovaerts, P., Geostatistics in soil science: State-of-the-art and perspectives, *Geoderma*, 89(1–2), 1–45, 1999b.
- Goovaerts, P., Geostatistical modeling of uncertainty in soil science, *Geoderma*, 103(1–2), 3–26, 2001.
- Hernandez, P. A., N. M. Perez, J. M. Salazar, S. Nakai, K. Notsu, and H. Wakita, Diffuse emission of carbon dioxide, methane, and helium-3 from Teide volcano, Tenerife, Canary Islands, *Geophys. Res. Lett.*, 25(17), 3311–3314, 1998.
- Isaaks, E. D., and R. M. Srivastava, *An Introduction to Applied Geostatistics*, 561 pp., Oxford Univ. Press, New York, 1989.
- Istok, J. D., and C. A. Rautman, Probabilistic assessment of ground-water contamination. 2. Results of case study, *Ground Water*, 34(6), 1050–1064, 1996.
- Kerrick, D. M., M. A. McKibben, T. M. Seward, and K. Caldeira, Convective hydrothermal CO₂ emission from high heat-flow regions, *Chem. Geol.*, 121(1–4), 285–293, 1995.
- Lapen, D. R., G. C. Topp, H. N. Hayhoe, E. G. Gregorich, and W. E. Curnoe, Stochastic simulation of soil strength/compaction and assessment of corn yield risk using threshold probability patterns, *Geoderma*, 104(3–4), 325–343, 2001.
- Lewicki, J. L., and S. L. Brantley, CO₂ degassing along the San Andreas fault, Parkfield, California, *Geophys. Res. Lett.*, 27(1), 5–8, 2000.
- Lin, Y. P., and T. K. Chang, Geostatistical simulation and estimation of the spatial variability of soil zinc, *J. Environ. Sci. Health, Part A*, 35(3), 327–347, 2000.
- Lin, Y. P., T. K. Chang, and T. P. Teng, Characterization of soil lead by comparing sequential Gaussian simulation, simulated annealing simulation and kriging methods, *Environ. Geol.*, 41(1–2), 189–199, 2001.
- Marty, B., and I. N. Tolstikhin, CO₂ fluxes from mid-ocean ridges, arcs and plumes, *Chem. Geol.*, 145(3–4), 233–248, 1998.
- Rautman, C. A., and J. D. Istok, Probabilistic assessment of ground-water contamination. 1. Geostatistical framework, *Ground Water*, 34(5), 899–909, 1996.
- Rogie, J. D., D. M. Kerrick, M. L. Sorey, G. Chiodini, and D. L. Galloway, Dynamics of carbon dioxide emission at Mammoth Mountain, California, *Earth Planet. Sci. Lett.*, 188(3–4), 535–541, 2001.
- Salazar, J. M. L., P. A. Hernandez, N. M. Perez, G. Melian, J. Alvarez, F. Segura, and K. Notsu, Diffuse emission of carbon dioxide from Cerro Negro volcano, Nicaragua, Central America, *Geophys. Res. Lett.*, 28(22), 4275–4278, 2001.
- Seward, T. M., and D. M. Kerrick, Hydrothermal CO₂ emission from the Taupo Volcanic Zone, New Zealand, *Earth Planet. Sci. Lett.*, 139(1–2), 105–113, 1996.
- Sinclair, A. J., Selection of threshold values in geochemical data using probability graphs, *J. Geochem. Explor.*, 3, 129–149, 1974.
- Welles, J. M., T. H. Demetriades-Shah, and D. K. McDermitt, Considerations for measuring ground CO₂ effluxes with chambers, *Chem. Geol.*, 177(1–2), 3–13, 2001.
- Williams, S. N., S. J. Schaefer, M. L. Calvache, and D. Lopez, Global carbon-dioxide emission to the atmosphere by volcanoes, *Geochim. Cosmochim. Acta*, 56(4), 1765–1770, 1992.

C. Cardellini and F. Frondini, Dipartimento di Scienze della Terra, Università di Perugia, Piazza dell'Università, I-06100 Perugia, Italy. (geochem@unipg.it; frondini@unipg.it)

G. Chiodini, Osservatorio Vesuviano, INGV, 328 Via Diocleziano, I-80124 Napoli, Italy. (chiod@ov.ingv.it)