

“PEAKLOCATOR 1.0, A WEB TOOL TO COMPARE EXTREME VALUE AREAS AMONG MAPS”

Domenico Granieri^{1,*}, Mattia de' Michieli Vitturi¹, Marina Bisson¹

⁽¹⁾ Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Pisa, Italy

Article history

Received June 15, 2018; accepted September 19, 2018.

Subject classification:

Numerical code; Map similarity; Spatial correspondence; Python web application; Open source code.

ABSTRACT

We present here a simple web application, PeakLocator 1.0 (hereafter referred to as PL1.0), for the analysis of gridded geo-located maps. In the present version of the code, the maps can contain up to 10 different variables with different units, not necessarily measured at the same locations, as well as the same variable recurrently measured in the time. The aim of PL1.0 is to identify regions where values lie outside the standard deviation from average values. The degree of spatial correspondence between these regions is reflected in the “fitting index” associated to the overlapping area. Here we demonstrate some possible applications of PL1.0 using published datasets, although its potential applicability extends to wide range of topics where the common demand is the comparison of two or more variables mapped over a common area or over areas partially overlapping. PL1.0 is freely accessible through a web interface and runs on any platform.

1. INTRODUCTION

The comparison of geo-spatial maps of data is useful in several research fields. In particular it is increasingly used in the environmental field and in numerous Earth science sectors. If the maps represent measurements of the same variable made at the same sampling locations then a map of simple arithmetic difference between the respective values is sufficient. The comparison becomes more complex when there are more than two variables with different measurement units and the sampling sites are different.

The representation of spatial data was born during the late 1960s [Coppock and Rhind, 1991] but only 25 years later the main principles and concepts were fully summarized in the reference work of Burrough and McDonnell [1998]. Burrough and McDonnell [1998] describe the spatial modelling of discrete and continuous data using the vector and raster representation and applying the main analysis of map algebra. The authors mention also

geostatistics as an appropriate tool to treat spatial data in the environmental field. Geostatistical techniques were developed to estimate changes in ore grade within a mine using discrete observations at specific locations, along with the associated errors in the estimates [Matheron, 1963; 1965]. Since the works of Matheron, geostatistics has evolved but has always remained faithful to the notion of *regionalized variable* that has spatial continuity from point to point, unlike random variable, but whose changes are not fully describable by a deterministic function. This approach uses discrete observations at specific locations to derive a variogram, which represents the spatial rate of change of the variable, which is then used to estimate values of the variable for the entire area of interest. The estimation procedure is called kriging, after Krige [1966], also a pioneering work in the application of geostatistical techniques to estimate changes in ore grade within a mine.

In the 1970s and 1980s, practitioners had to write their own code for geospatial analysis, but in the last 30

years powerful software has become widely and cheaply available in the public domain. At the present, such analysis can be done using Statistic or MATLAB or with GIS (Geographic Information System) platforms such as ESRI, GRASS, QGIS. These GIS platforms are able to process discrete or continuous data by using geo-coded vector or raster formats. They provide dedicated tools for making comparisons between variables by using mainly map algebra analyses on grid/matrix formats.

In order to compare different variables measured in a common area without using GIS platforms or complex software, we developed a very simple web tool in Python, called PeakLocator 1.0 (PL1.0 hereafter). It compares maps of n variables (up to 10 in the present version of the code) that have the same or different measurement units, measured in overlapping domains but not necessarily taken at the same locations. The code is then able to highlight regions where the variables are positively or negatively different from average values by some pre-determined thresholds. Alternatively, PL1.0 can query maps related to the same variable recurrently measured in the time (up to 10 times) in an effort to understand if area(s) characterized by unusual values are stable or variable in time. In addition, a quality parameter, named ‘the fitting index’, is computed to quantify the degree of spatial overlapping for the different maps. The advantages of PL1.0 are: 1) it is applicable to gridded datasets with variable sampling density and different measurement units; 2) it allows simultaneous consideration of n variables in order to identify regions of the domain where two or more variables are positively or negatively correlated; 3) it allows selection of the threshold for identifying unusual values, 1, 2 or 3 standard deviations above or below the mean for each variable; 4) it allows quantification of the spatial area over which two or more variables are correlated in relation to the whole domain; and 5) it produces gridded outputs which are readable by most contouring and mapping software and GIS tools. The only necessary requirement is that the observations must fall in a common area or in two or more areas partially overlapping, obviously with a higher significance of the analysis the larger is the common portion and the density of observations. Moreover, since the code is designed to work with grid-based data only, the sparse measurements should be infilled in a gridded pattern to cover the entire domain through a gridding procedure. This step cannot be performed directly by PL1.0, other tools can be used to interpolate the values at not sampled locations to create suitable maps to compare.

In the cases presented here, variables were infilled using the sequential Gaussian simulation approach (sGs,

Deutsch and Journel, 1998), also described in Cardellini et al., [2003]. Variables at each not sampled location were infilled by a random sampling of a Gaussian conditional cumulative distribution function defined on the basis of original data (conditioning) and previously simulated data within its neighbourhood (sequentiality). Also, while kriging provides the variance of each local estimate, differences among many sGs realizations can be used as a quantitative measure of the associated spatial uncertainty [Deutsch and Journel, 1998; Goovaerts, 2001; Cardellini et al., 2003].

The format and the description of input and output files are described in Appendix A.

2. CODE DESCRIPTION

The main function of PL1.0 is to find common areas of negative or positive extreme values between two or more maps. This goal is obtained through several steps described here and sketched in Figure 1.

Step 1. All the N maps are cropped over a common frame, given by the intersection of the domains, and interpolated on the same grid points (repositioning). For this purpose, a nearest neighbor interpolation is used, in order to avoid the creation of artificial values and to keep the areas with no values that could be present in the original maps.

Step 2. Each map is nondimensionalized and scaled in order to overcome the problem of the different units. In this phase, the average value μ_i and the standard deviation σ_i are computed over the common frame for each variable Y_i (where $i=1, \dots, n$). Then, each variable is centered with respect to its average value and divided by its standard deviation:

$$S_i = \frac{Y_i - \mu_i}{\sigma_i} \quad (1)$$

The new variable S_i will represent the number of (non-dimensional) standard deviations above or below the mean. We remark that, when μ_i and σ_i are known, the original value Y_i can be recovered from the normalized value S_i .

Step 3. The regions of extreme values $A_{i,\alpha}$, corresponding to areas where S_i exceeds (positively or negatively) a user-determined multiple of the standard deviation (“exceeding coefficient” α_i), are defined as:

$$A_{i,\alpha_i} = \left\{ \text{pixel } j \mid \frac{\alpha_i}{|\alpha_i|} (S_i - \mu_i) > |\alpha_i| \sigma_i \right\} \quad (2)$$

These regions define a masking of the original maps. It is important to note that the exceeding coefficients

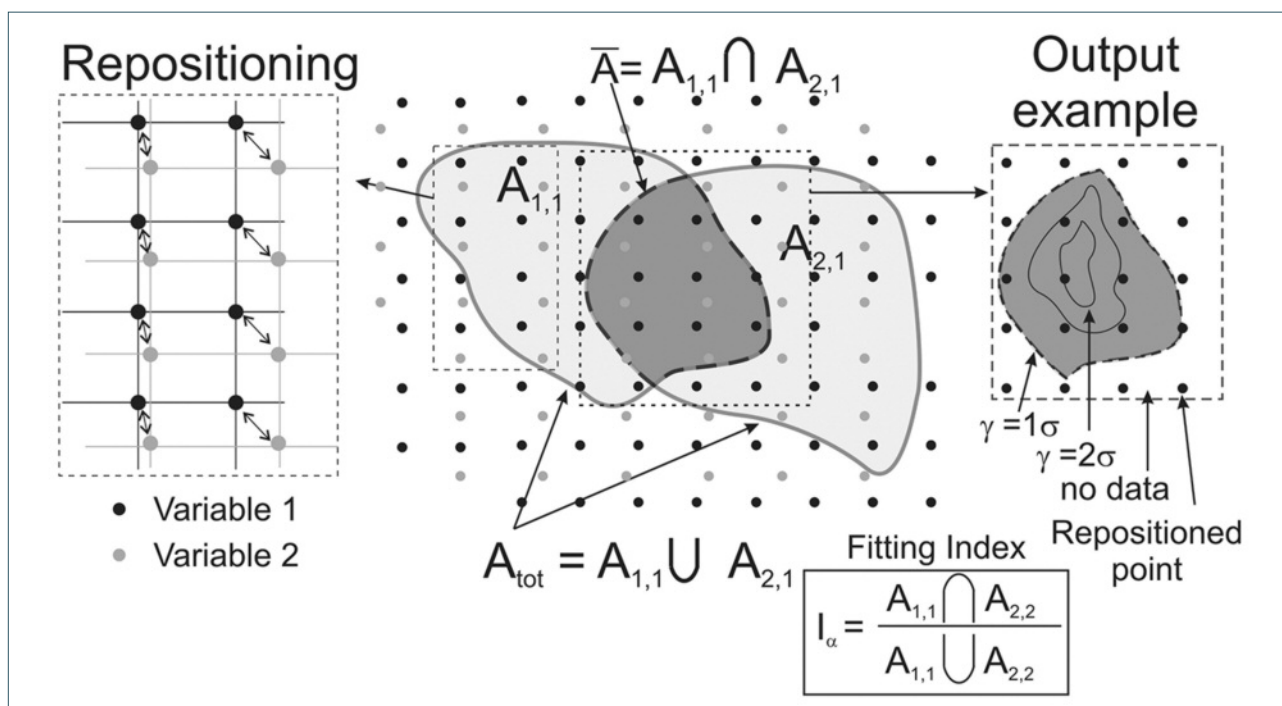


FIGURE 1. Schematic representation of the PL1.0 procedure considering the maps of the variable 1 and 2, whose areas of extreme values are $A_{1,1}$ and $A_{2,1}$, obtained setting $\alpha_1 = \alpha_2 = 1$. We remark that, in the insert representing the output, the bounded grey area represents the region where the values of both the variables exceeded the mean value by at least 1 standard deviation.

can have different values for the different maps; in this way, it is possible, for example, to search for the correlation between a higher value region of the first map with a lower value region of the second map ($\alpha_1 > \alpha_2$).

Then, the area \bar{A} is defined by the intersection of regions of extreme values of individual maps as:

$$\bar{A} = \bigcap_i A_{i, \alpha_i} \quad (3)$$

Finally, in order to quantify how well the extreme value regions overlap, a fitting index is defined as the ratio of the intersection of all the regions divided by the union, following Jaccard [1901]:

$$I_\alpha = \frac{\bigcap_i A_{i, \alpha_i}}{\bigcup_i A_{i, \alpha_i}} \quad (4)$$

The closer the fitting index is to 1, the larger is the similarity of the extreme value regions of the different variables, meaning that the common area \bar{A} well represents the regions of extreme values of all the maps.

In order to visualize the values of the maps in the common region, a new map is defined, with the value γ defined for each pixel in the common area:

$$\gamma = \min_i [\text{sign}(\alpha_i) S_i] \quad (5)$$

This means that if a pixel in the resultant map has

a certain value, for example 2, in that pixel all the considered variables have values exceeding (positively or negatively according to the sign of α_i) the respective mean value μ_i by at least 2 standard deviations σ_i .

PL1.0 can be run as a web tool at the website <http://peaklocator.pi.ingv.it> and the source code can be downloaded from <https://github.com/demichie/PeakLocator>.

3. APPLICATIONS

We propose and briefly discuss hereafter some applications of PL1.0, suggesting some possible practical benefits, using published datasets that were the subject of previous scientific papers.

3.1 CASE 1 | FINDING DIRECT CORRELATION BETWEEN HIGHER VALUE AREAS

The goal of this application is the identification of common areas where the values of the considered variables are at least one standard deviation above the respective averages. The maps to be compared can have the same or different units and the same or different sampling density.

Possible practical benefits are: i) find the persistence in time of areas of extreme values or assess their spatial migration; ii) select a best site to install a monitor-

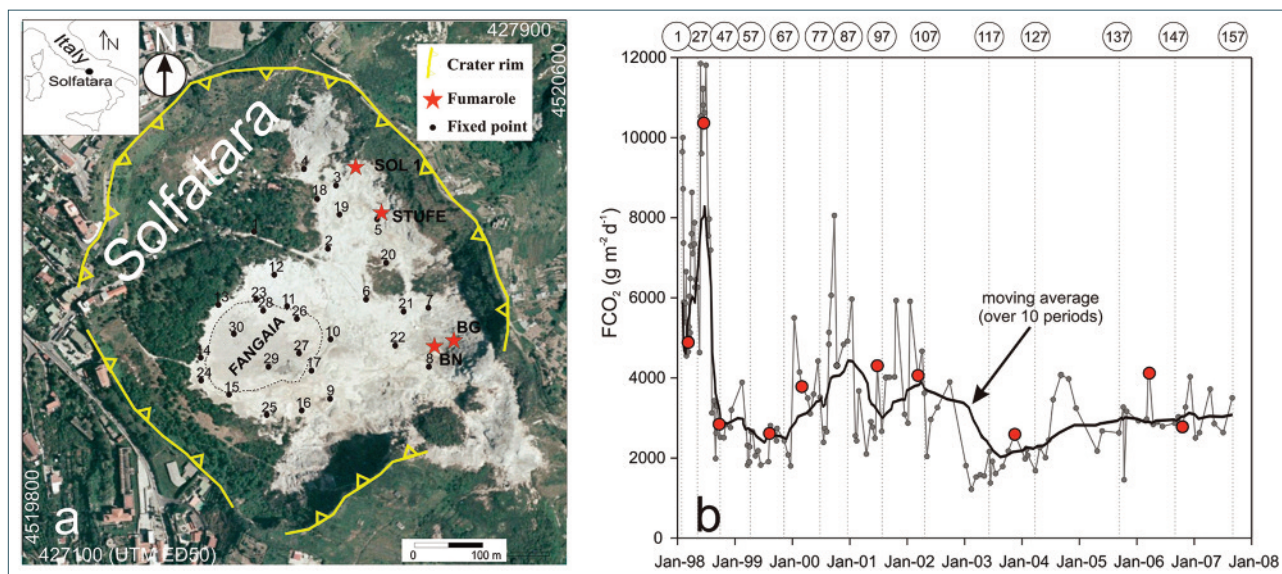


FIGURE 2. a) Array of 30 fixed points inside the crater of Solfatara to monitor the soil CO₂ emission. Fumaroles are also shown; the Fangaiia is a mud pool; b) Variation of the soil CO₂ flux in the 1998-2007 period over the 30-point array (average values of 157 campaigns). Red points refer to 10 campaigns randomly extracted to be considered in the application of PL1.0.

ing station; iii) find the spatial correspondence among higher values of two variables or iv) more than two variables.

i) As a first example of the usage of PL1.0, we investigated the persistence or the spatial shift of areas characterized by higher values of the soil CO₂ flux inside the crater of Solfatara (Campi Flegrei, Italy). In this area, CO₂ diffuse degassing from the soil has been monitored since January 1998 by periodic CO₂ flux measurements over an array of fixed stations covering an

area of ~125.000 m² in the flat floor of the crater (Figure 2a). With time, the array has become gradually more dense by increasing the number of fixed stations from 30 to 71 [Granieri et al., 2010], without significant enlargement of the investigated area. For this application, we used the dataset of measurements over the primary 30-point array that was repeated during 157 campaigns from January 1998 to September 2007. The average value of each campaign is reported in Figure 2b. Owing to the different frequency of sampling (weekly in 1998 and then almost monthly in the following years), the campaigns are irregularly spaced in the time [Granieri et al., 2010].

We ordered the campaigns in chronological order from 1 to 157 (the numbers at the top of the graph of Figure 2b are referred to some of these campaigns) and 10 of these were randomly selected as training data through a procedure of generating random numbers in the appropriate interval (1-157), with the condition that any duplication is avoided. The 10 maps that came up (10, 34, 46, 63, 72, 96, 105, 123, 143, 149, indicated by red dots in Figure 2b) were processed through PL1.0 and the overall result, in terms of pixels exceeding by at least one standard deviation the mean value ($\alpha_1=1$ for all the maps), is shown in Figure 3. We remark again that the plotted value is the minimum among the normalized values of all the maps.

Although the campaigns, randomly selected, are characterized by different average values (cfr. Figure 2b), one main common area of high values is the fumarole area of SOL1 and STUFE (zone 1 in Figure 3), located at the intersection of the NW-SE and NE-SW main fault

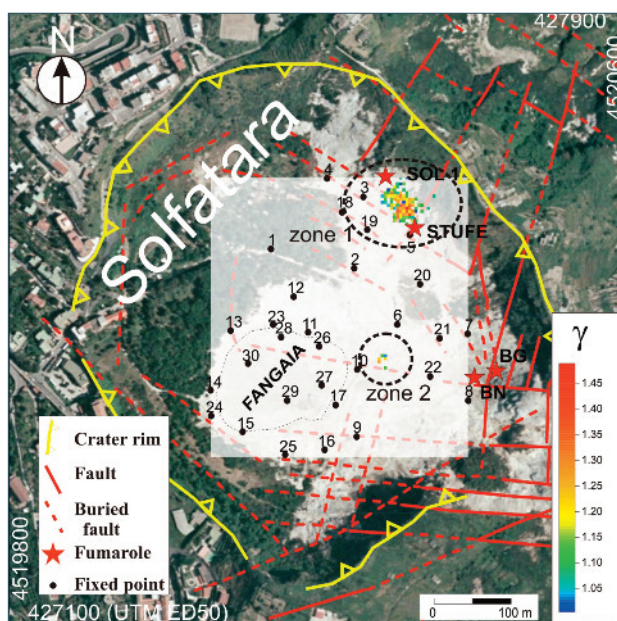


FIGURE 3. Areas of highest values resulting from 10 randomly selected campaigns after the application of PL1.0. The main tectonic structures come from Isaia et al., [2015].

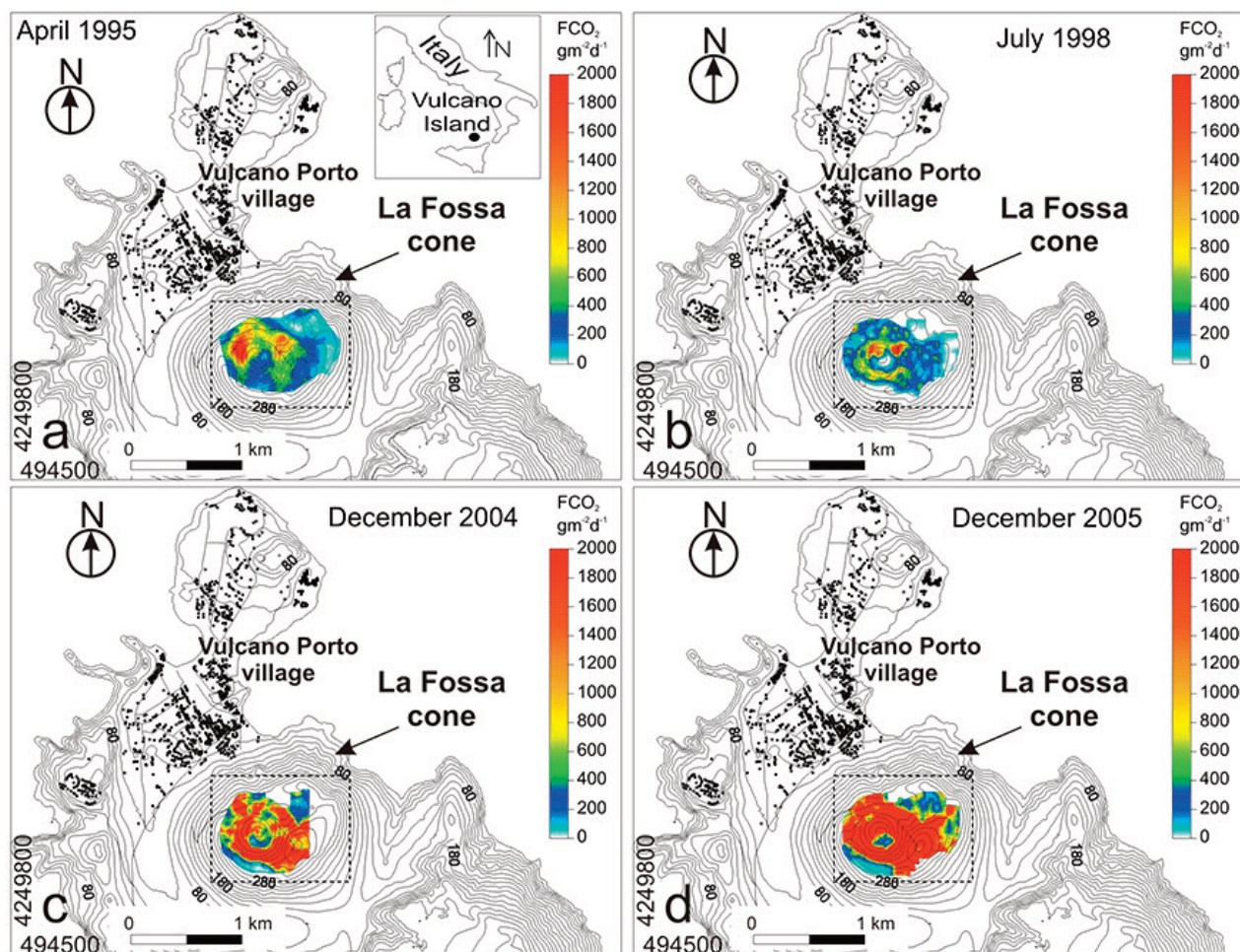


FIGURE 4. Maps of 4 soil CO_2 flux campaigns at the crater of La Fossa (Vulcano island), modified from Granieri et al., [2006]. Dashed rectangle encompasses the area zoomed in Figure 5.

systems of the Solfatara crater [Isaia et al., 2015], and a less extended area of high values is in the center of the crater (zone 2), near to the Fangaia mud pool. We observe that for this test the fitting index is quite low ($I_\alpha \sim 0.028$), expressing the fact that the resulting common area (1653 m^2) represents only a small portion (approximately 3%) of the area where all maps exceeded the respective average values by one standard deviation. Although of limited size, the persistence through time of these areas displaying significant CO_2 emission, could suggest a better connection between the surface and the feeding hydrothermal system of Solfatara than surrounding zones.

ii) We used another test bed to demonstrate PL1.0 as a tool for selecting optimal sites for automated continuous sampling stations, starting from a limited number of explorative field surveys. In the case presented here, we investigated the CO_2 flux from the soil at La Fossa crater of Vulcano (Aeolian Archipelago, Italy). CO_2 soil flux is monitored here by periodic field surveys over a discrete number of sample locations and continuously

through automated stations [Inguaggiato et al., 2018]. A similar combined approach is applied in different volcanoes worldwide [Granieri et al., 2003], e.g., at Campi Flegrei, Vesuvio, Etna, Stromboli, Vulcano (Italy), Masaya (Nicaragua), Poa's (Costa Rica), La Palma (Canary Islands, Spain), San Salvador, San Miguel, San Vincente and Santa Ana (El Salvador), Usu (Japan), Mammoth Mountain (California, USA). Some of these experiences have fallen short due to non-ideal choices of automated measurement sites. In fact, if the purpose of the observation is to monitor deeper volcanic processes and changes in the levels of volcanic activity in real time, the device must be placed in a sector of the crater where the degassed CO_2 is derived from the hydrothermal/magmatic system, as mainly supported by measured fluxes orders of magnitude higher than the values of the local biological CO_2 background (typically around $20\text{-}50 \text{ gm}^{-2}\text{d}^{-1}$ for volcanic soils, Chiodini et al. 2008, Granieri et al. 2003). Furthermore, the station needs to be in a location not too exposed to volcanic gases, generally acidic and with detrimental impacts on

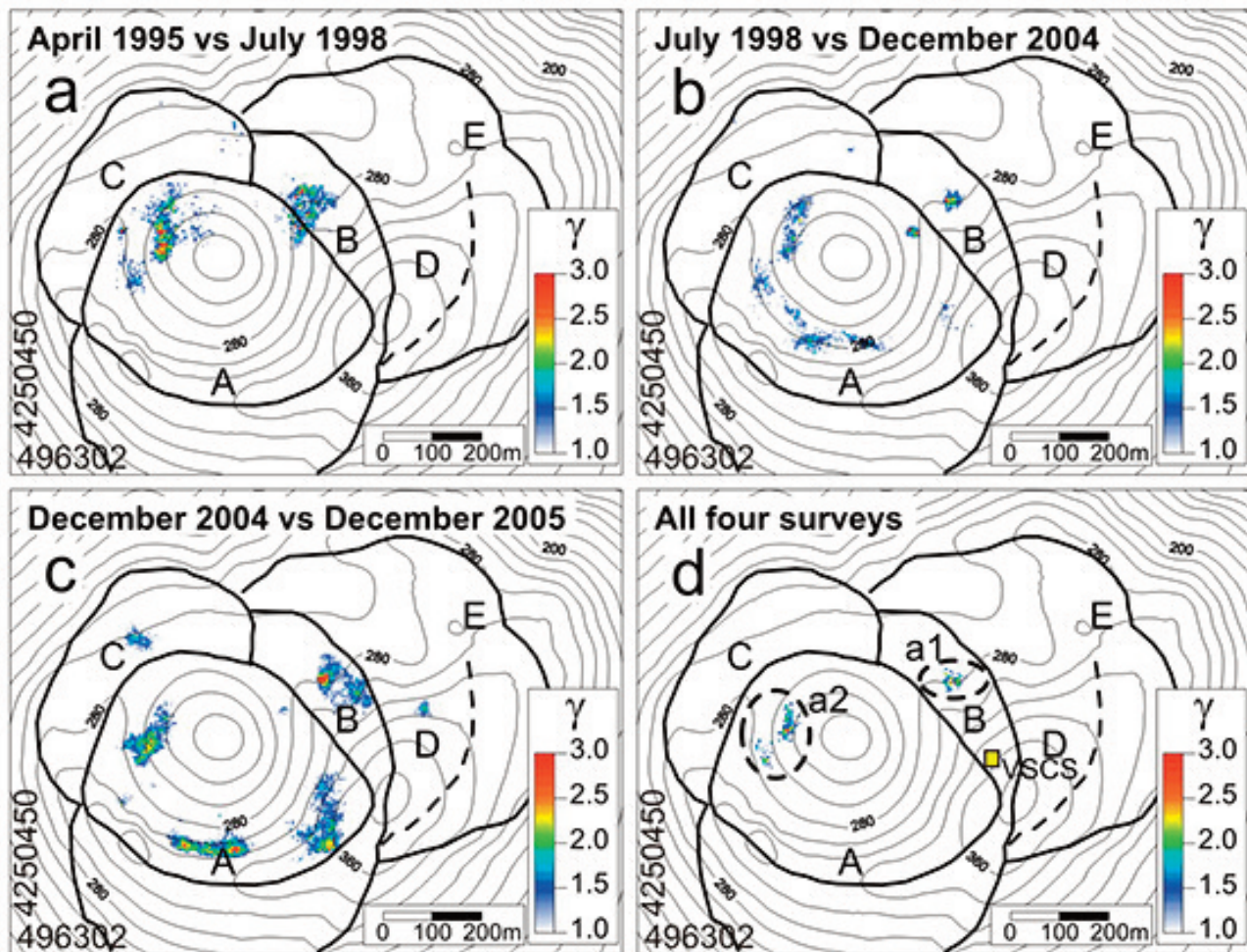


FIGURE 5. Common areas of higher soil CO_2 flux considering pairs of campaigns (a, b, c) and all 4 campaigns (d). The rims of the present-day crater (A) and of paleo-craters that were active in different times are indicated (B-Pietre Cotte; C-Palizzi and Commenda; D-Unidentified; E-Punte Nere, from Frazzetta and La Volpe [1991]). In d) common areas are indicated by dashed ellipses (a1 and a2) and the position of the automated station (VSCS) operative since 2007 by a yellow box.

measuring sensors and electronics, and easily accessible to operators for maintenance.

Since the last eruption of 1888-1890, the crater of Vulcano showed an intensive and persistent degassing activity, with periods of enhanced degassing involving both the fumaroles and the soil [Granieri et al., 2006, 2014, Paonita et al., 2013]. As reported in Granieri et al. [2014], a CO_2 output of $216 (\pm 83)$ tons per day (td^{-1}) represents the “background” soil CO_2 emission from the crater in the present quiescent stage of the volcano whereas higher CO_2 emissions were recorded in December 2004 and December 2005, with CO_2 emission rates of 700 td^{-1} and 1600 td^{-1} , respectively [Granieri et al., 2014]. These values were estimated on the basis of fifteen field surveys performed during the period from 1995 to 2010, each of them based on a different number of samples, taken at different locations approximately over the same crater area of $\sim 1 \text{ km}^2$ through the “accumulation chamber” portable device [Granieri et al., 2014].

We started from 4 maps of soil CO_2 flux from April 1995, July 1998, December 2004, and December 2005 at La Fossa crater of Vulcano and reported in Granieri et al., [2014]. They are indicative of the mean CO_2 output (266 td^{-1} in April 1995), low CO_2 output (162 td^{-1} in July 1998), high CO_2 output (700 td^{-1} in December 2004), and very high CO_2 output (1600 td^{-1} in December 2005). These maps, derived through the sGs approach [Cardellini et al., 2003], are shown in Figure 4.

First, PL1.0 was applied to pairs of maps, following a temporal order in their coupling, and then the code was applied to all four maps combined.

Results from the comparison of pairs of maps (Figure 5a,b,c) highlighted that regions exceeding the mean values by 1 standard deviation are always contained within the present crater rim (crater A) or in a restricted sector outside to the crater rim, along the NE edge (encompassed by the crater A and the paleo-crater B). The combined map of all four surveys confirmed the existence of two common highest-emission areas (a1 and a2

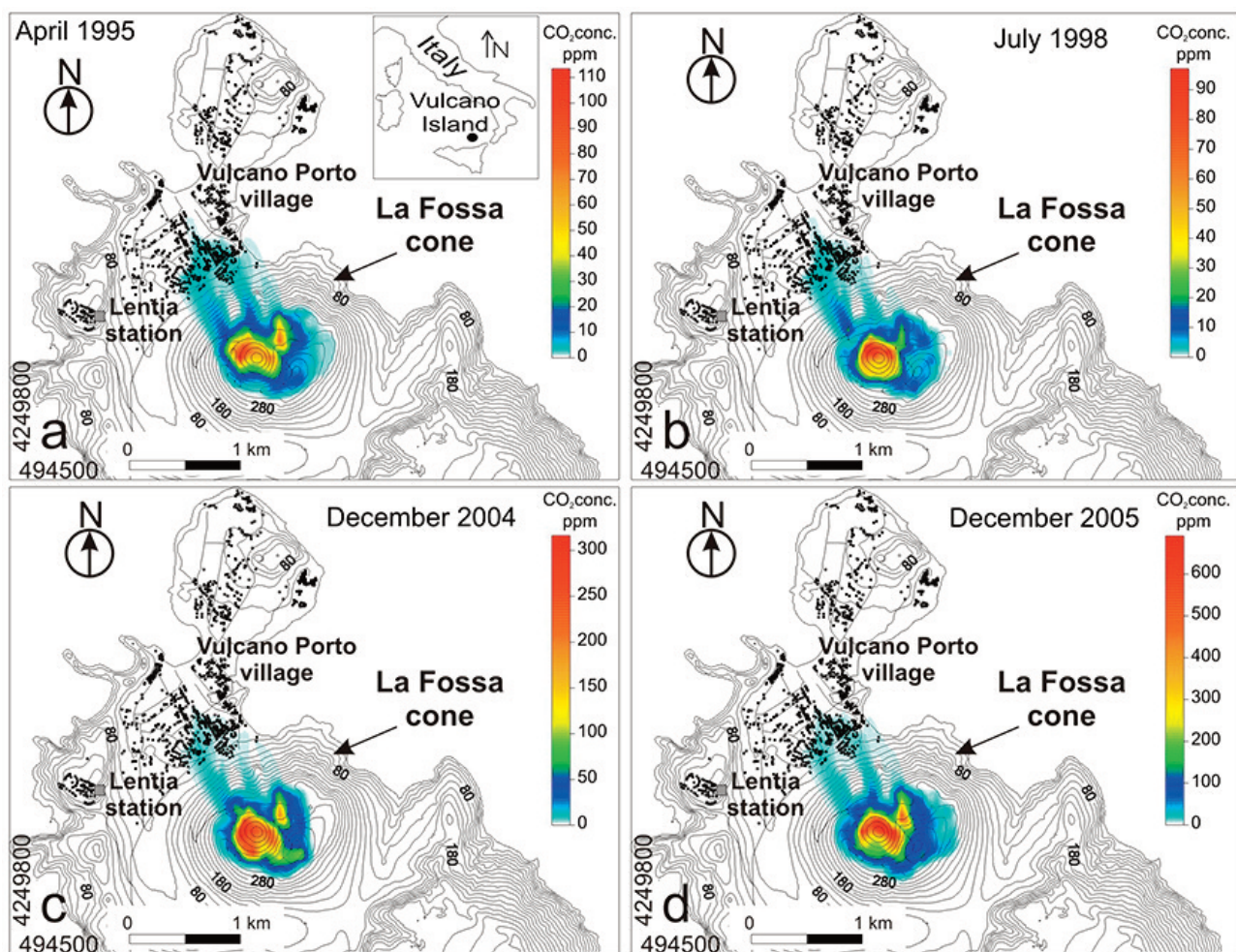


FIGURE 6. Maps of the air CO_2 concentration at the crater of La Fossa in different times under a SSE wind (see text). The site of the meteorological station (Lentia station) is shown.

in Figure 5d, for a total surface of 3458 m^2), representing cumulatively about 2% of the total area ($I\alpha \sim 0.02$), where all maps exceed the respective mean values by at least 1 standard deviation. The low value of the fitting index likely results from the appearance of new higher values in the southern sector of the crater after the first period of observation. Although a1 and a2 areas are both formally suitable (Figure 5) for the placement of an automatic station, other considerations concerning the safe distance to preserve sensors and electronics of the station and a more comfortable position for maintenance might encourage the operator to select the area a1 as the most appropriate. In reality, an automated station to monitor the soil diffuse CO_2 degassing at the summit of La Fossa cone (VSCS station, Figure 5) is located in a site near to a1 area since September 2007 [Inguaggiato et al., 2018].

iii) Starting from different levels of soil CO_2 diffuse degassing recently measured at the crater of Vulcano (from 162 td^{-1} in July 1998 to 1600 td^{-1} in December 2005), we performed a further test of PL1.0 in an attempt

to identify common regions of the domain where air CO_2 concentrations, whose levels are mainly due to the volcanic source [Granieri et al., 2014], exceed the respective mean values by 1 standard deviation. In order to model CO_2 plume dispersion, we applied the DISGAS code [Costa et al., 2005, 2016; Granieri et al., 2013, 2014, 2017]. Despite being rare at Vulcano, for this test we imposed a wind from SSE since it draws a dangerous situation for the village of Vulcano Porto lying downwind with respect to this wind direction. The average speed of 1.87 m/s was extracted from the dataset of a local meteorological station (Lentia station), covering the 8-year time period from May 2008 to February 2016 (for detail see Vita et al. 2012 and Granieri et al. 2017). Resulting maps of the plume dispersion for four periods are reported in Figure 6.

Results show that the common area of modeled extreme values of ambient CO_2 concentration is encompassed by the rim of the present crater with the exception of a lobe in the external NE sector (Figure 7). The inhabited area of Vulcano Porto is substantially unaf-

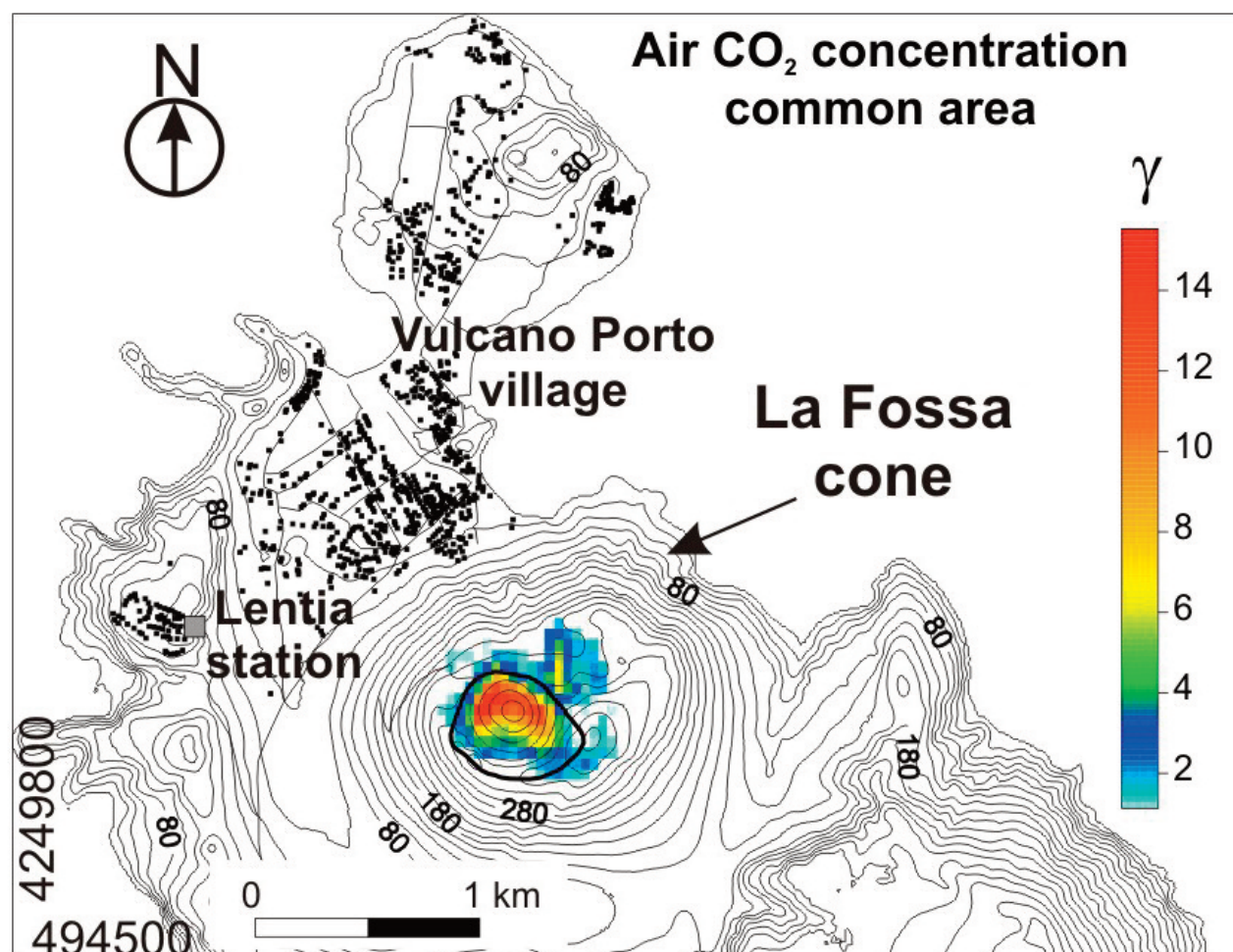


FIGURE 7. Common area of higher air CO₂ concentration considering the four maps of Figure 6. The thicker black curve encompasses the present-day crater (A in Figure 5).

ected by the crater-derived CO₂ plume, confirming the finding of Granieri et al. [2014]. The large value of the fitting index ($I_{\alpha} \sim 0.60$) obtained comparing the four maps, confirmed that the region of the higher CO₂ concentration, essentially concentrated in the present crater rim (A in Figure 5), persists in time, despite different degassing levels measured at the volcano.

iv) In this application of PL1.0 we were interested in the spatial correlation between three variables, having the same measurement units and different statistics (mean, standard deviation, skewness, kurtosis, etc.), which were simultaneously measured at the same locations. In this specific case, we used concentrations (in $\mu\text{g/L}$) of arsenic (As), uranium (U) and vanadium (V) in 328 water samples, collected in the Vicano-Cimino volcanic district (central Italy). Data were taken from Cinti et al. [2015] and they are provided as a test case in the Supplementary material. In some areas of central Italy, the concentration of heavy metals in drinkable waters can pose a hazard for people, in particular the high level of As, often above the threshold imposed by a specific

European directive [EC Directive, 1998]. Although not regulated by the countries of the European Union, high concentrations of U and V can be toxic, and for this reason they were measured in the area along with the As concentration. The main factor controlling the concentration of As, U and V in waters resulted to be a combination of the lithology where waters move together with the thermal and redox conditions of the water-rock interaction [Cinti et al., 2015]. Maps derived from the data of Cinti et al. [2015] are shown in Figure 8,a,b,c.

By correlating the entire dataset, it is not possible to find any significant correlation among the pairs of the three elements ($r=+0.12, -0.03, 0.00$ for U-V, As-V, As-U pairs, respectively, as reported in the first column of Table 1). The comparison consisted simply of calculating the correlation coefficient between the concentration of variable 1 and 2, 1 and 3, and 2 and 3, with no regard for the spatial position of the samples. The main disadvantage of this procedure is the inability to discriminate if the overall absence of correspondence in the whole area may mask a higher degree of similarity in sub-regions of the

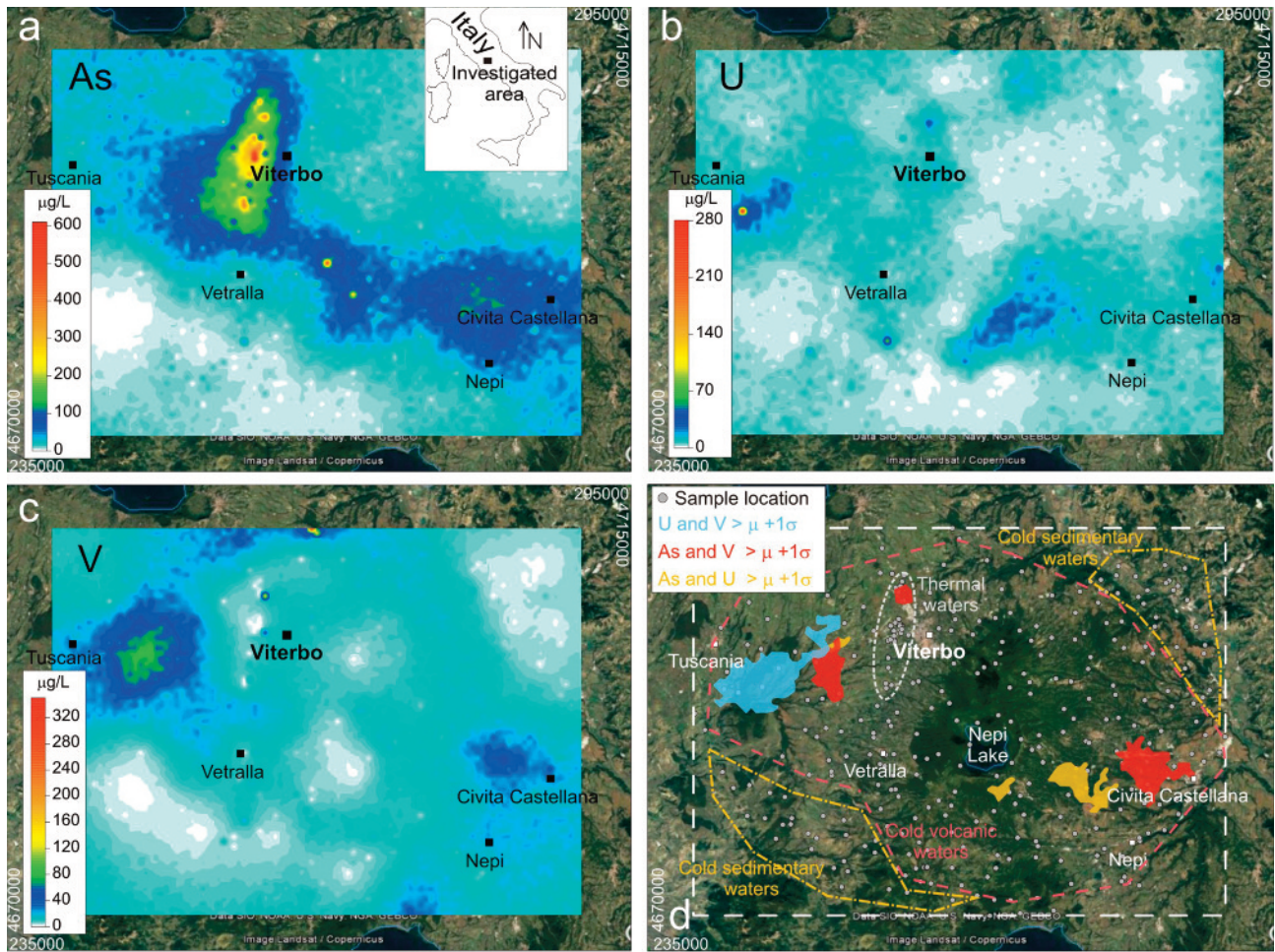


FIGURE 8. Common area of higher air CO_2 concentration considering the four maps of Figure 6. The thicker black curve encompasses the present-day crater (A in Figure 5).

Couple	r (all samples)	r cold-volcanic waters	r cold-sedimentary waters	r thermal waters	r bubbling pools
U-V	+0.12	+0.17	+0.63	+0.28	-0.06
As-V	-0.03	+0.06	+0.83	-0.06	+0.25
As-U	0.00	+0.13	+0.68	-0.16	+0.18

TABLE 1. Correlation coefficient (r) between the water concentration of uranium (U), vanadium (V), and arsenic (As), considering all samples (first column) or clustering the samples on the basis of four water types.

domain. In order to explore this possibility, we clustered the samples into four categories on the basis of the water type, as identified by the authors, somehow reflecting the spatial distribution of the samples (Figure 8d). Results highlighted a strong positive correlation between U and V in cold-sedimentary ($r=+0.63$) and thermal waters ($r=+0.28$), a minor correlation in cold-volcanic waters ($r=+0.17$), and a low (and negative) correlation in bubbling pools ($r=-0.06$). Similarly, a strong correlation comes out between As and V and between As and U in cold-sedimentary waters ($r=+0.83$ and $+0.68$, respec-

tively), moderate or low correlation in cold-volcanic waters ($r=+0.06$ and $+0.13$, respectively), and low (and negative) correlation in thermal waters ($r=-0.06$ and -0.16 , respectively). Table 1 summarizes these relationships.

PL1.0 was applied for the As-U, As-V, and U-V pairs, starting from the whole datasets of 328 water samples. For this application, the units and the sampling density of the three variables are the same, but, as previously stated, this is not required for the applicability of the code. Results of the procedure allowed easier identification of common areas of extreme values (Figure 8d),

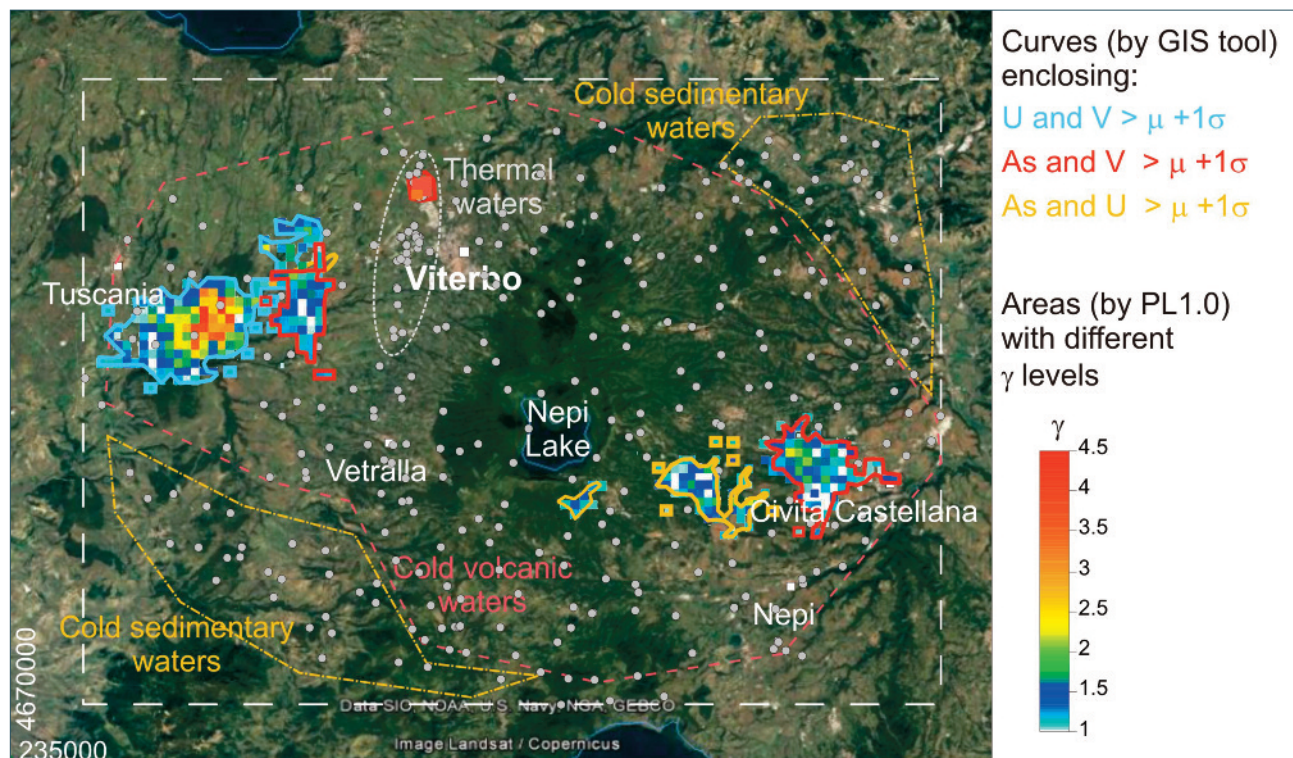


FIGURE 9. Comparison between the GIS tool Raster Calculation (colored curves) and PL1.0 outputs (tessellated areas) for the case study iv (see text).

with a fitting index of about 0.15 for the couple U-V, and lower values for As-V and As-U pairs ($I_{\alpha} \sim 0.09$ and ~ 0.08 , respectively), confirming the correspondence between higher value areas and water types, but with more spatial information than a simple statistical correlation. In fact, common areas are all found in the spatial domain of the cold-volcanic waters (Figure 8d), where the correlation of the three considered pairs is moderate (second column of Table 1), rather than in the cold-sedimentary waters where the correlation is high (third column of Table 1), suggesting that the high correlation in the cold-sedimentary samples is likely due to the numerically largest group of low-medium concentration values whilst that highest values (“anomalies”, *sensu lato*) are well correlated in the domain of the cold-volcanic waters. As a general conclusion, this means that if the statistical value of a correlation may not be in doubt, the “physical” meaning ascribed to that value is open to interpretation for the peculiar nature of the geochemical data. That is the conclusion reached through the application of PL1.0, without the necessity of knowing a-priori a further parameter, i.e., the water type, on which the clustering and the subsequent correlation “for groups” was based.

The performance of PL1.0 was tested by comparing the results of this last case study against the prediction from the “Raster Calculation” tool available in ARCGIS 10.5 software (ESRI platform).

This tool, available in the ArcToolBox extension, presents a calculator-like interface and provides as output a Boolean geo-raster that identifies the areas where a request is satisfied (or not) by combining mathematical and logical operators. The results of the comparison are shown in Figure 9. The anomalous regions obtained with PL1.0 are enclosed by the curves resulting from the GIS tool, showing very good spatial agreement. The results of this comparison, as other ones not shown, confirmed that the code has performed well and it is satisfactorily robust. Then, when a comparison among gridded maps is required, it can provide a reliable alternative to the more established and complicated approach of GIS tools, for which a wealth of knowledge and experience is necessary. In addition, it adds further information about the user’s request. For instance, in the aforementioned case-study the GIS tool pulls out the curve enclosing the common areas in which values of the couples are higher than their means + 1 standard deviation (as required) whilst the PL1.0 application provides a pixel tessellation with different levels of the exceedence of the mean from +1 standard deviation up to +4.5 standard deviations.

3.2 CASE 2 | FINDING INVERSE CORRELATION BETWEEN INVESTIGATED AREAS

The goal of this application is the identification of common areas where the positive values of one vari-

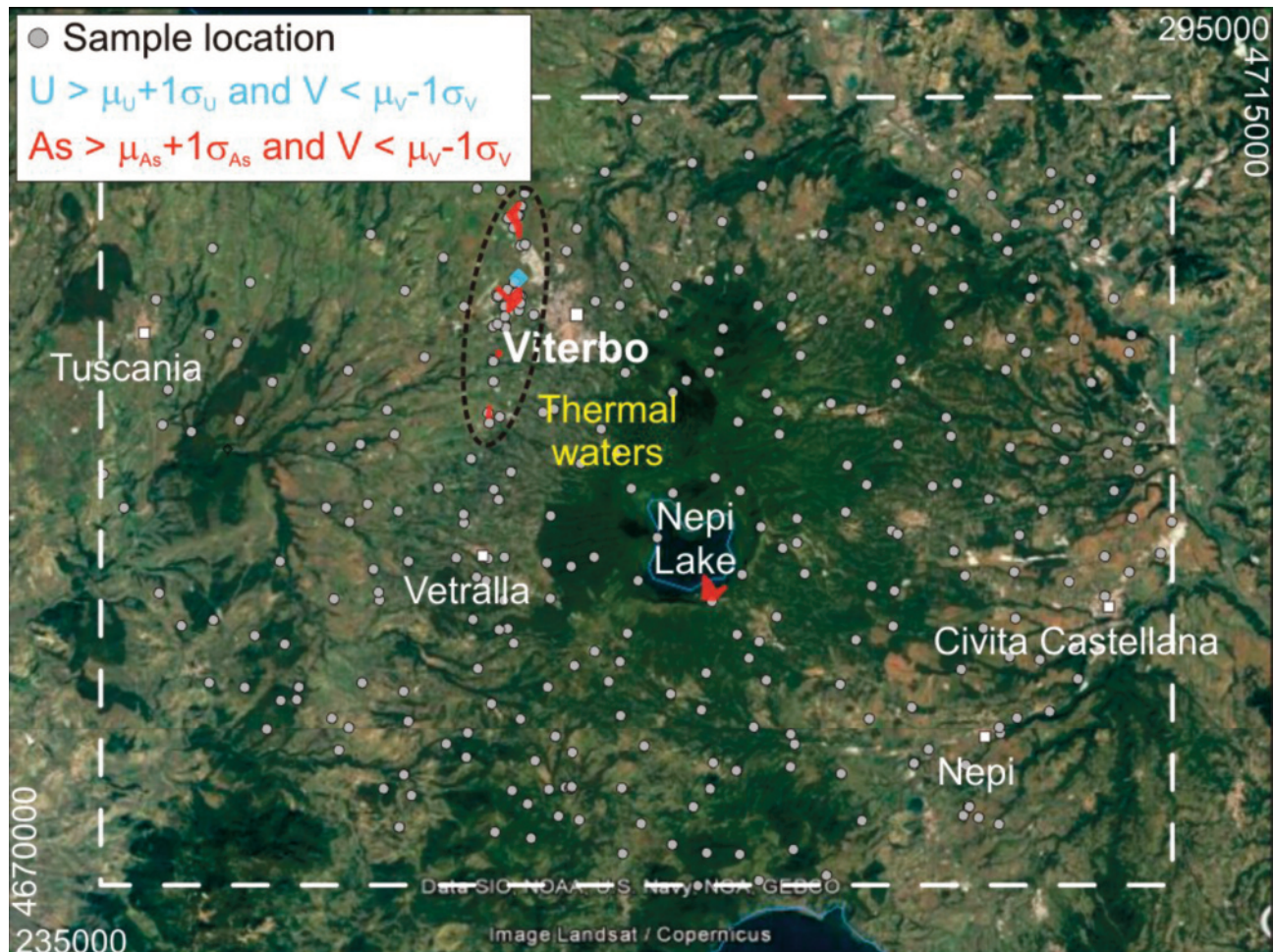


FIGURE 10. Resulting areas where water concentrations of U and As one standard deviation above their respective means coincide with values of the V concentration one standard deviation below its mean.

able (calculated as average value plus 1, 2 or 3 standard deviations) correlate with the negative values of one another (calculated as average value minus 1, 2 or 3 standard deviations). Possible practical benefits are: i) identify common areas of the domain where lithological, structural or physico-chemical features of the medium favor the escape of a gas, compound or metal in the near-surface environment and inhibit the output of another.

We present this case by inspecting the same dataset of As, U, and V concentration in 328 water samples used in the previous application for values exceeding the mean by (+/-) 1 standard deviation. Results of this PL1.0 application highlighted the absence of negatively correlated areas for the couple As-U (Figure 10), and the existence of two small “inversely correlated” areas for the pairs U-V and As-V (with α of 0.02). Interestingly, these resulting areas, with the exception of a sub-area located near Nepi Lake, fall in the domain of the thermal waters. This correspondence is likely linked to the anoxic conditions of the thermal waters, differently influencing the solubility of As, U, and V [Cinti et al., 2015], but the full explanation of the process is beyond the scope of the present paper.

3.3 CASE 3 | FINDING DIRECT CORRELATION BETWEEN AREAS WITH VERY EXTREME VALUES (2σ OR 3σ ABOVE/BELOW THE MEAN)

The last application of PL1.0 is aimed to identify “very anomalous” regions of the domain where the values of two or more variables exceed (or are below) the mean by 2 or 3 standard deviations. This could be useful, for example, in the field of the mining exploration when simultaneous high values of grade, thickness and density of ore could suggest an ideal site for the mining.

This case concerns the comparison between the tails of the value distribution in order to bring out the location of sub-region(s) of the domain characterized by very extreme values of the measured variables. For example, when measurements are *normally* distributed, a quite common peculiarity for geochemical data, no more than 5% of all measurements are included within the interval higher than +2 standard deviations or lower than -2 standard deviations on both sides of the mean, and slightly over 0.1% of all measurements fall within the interval higher than +3 standard deviations or lower

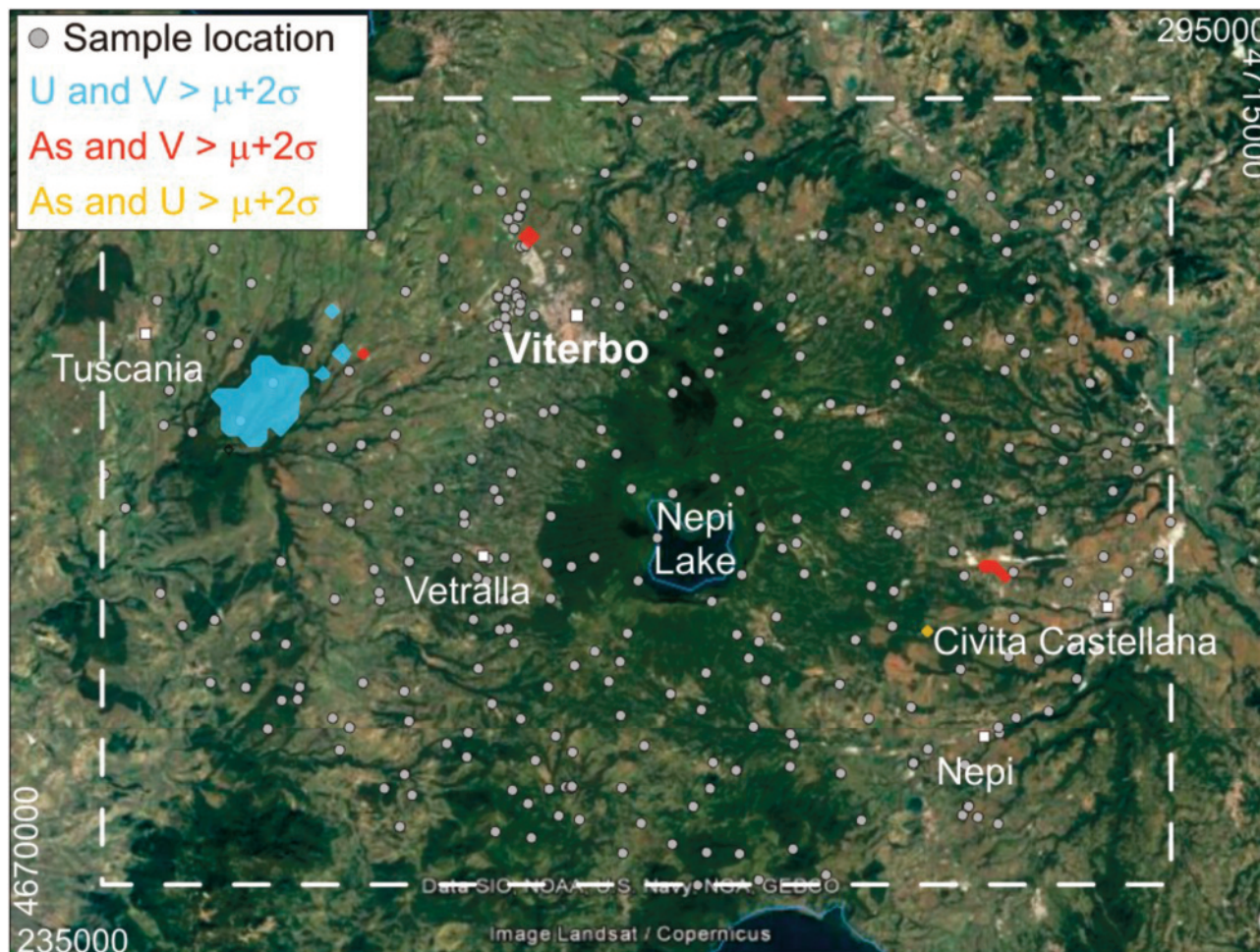


FIGURE 11. Common areas with very extreme values for pairs of elements where their concentrations in water are two standard deviations above their means.

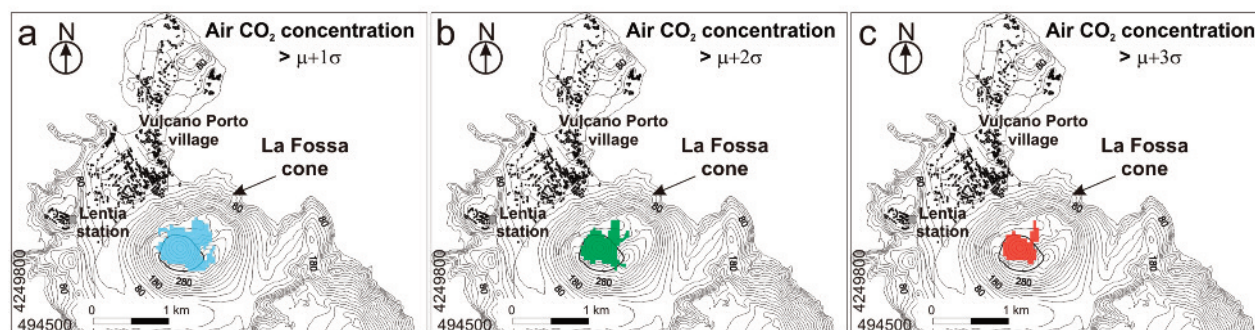


FIGURE 12. Common area of air CO₂ concentration considering the four maps of Figure 6 with values exceeding the mean by a) one, b) two, and c) three standard deviations. The map in a) is the same of Figure 7.

than -3 standard deviations.

In this application, we consider the dataset of As, U, and V concentration in water samples for values exceeding the respective mean by at least 2 standard deviations. Results of PL1.0 permitted highlighting of a larger zone of high values for the couple U-V (with a fitting index of about 0.1) in proximity to the Tuscania

village (Figure 11), where actually the contamination of groundwater posed serious problems for human health [Cinti et al., 2015] and two smaller areas (with negligible values of the fitting index) of simultaneous As-V and As-U highest values near to Viterbo town and Civita Castellana village, respectively (Figure 11).

We applied again PL1.0 to the air CO₂ concentra-

tion maps (temporally spaced) at La Fossa crater of Vulcano (see Section 3.1), but here considering values exceeding the mean by 2 and 3 standard deviations. Code results highlighted the persistence over time of a common area inside the present-day crater rim (cfr. Figure 6), where, as expected, the common area of very extreme values progressively reduces its extension when larger thresholds (from $+1\sigma$ to $+3\sigma$) are considered (Figure 12a,b,c). For this application, even for the largest threshold, a good similarity of common areas for the four CO₂ concentration maps (cfr. Figure 6) is obtained, as suggested by the fitting index of 0.43 (Figure 12c).

4. CONCLUSIONS

In this study, we describe the new PL1.0 web application for analyzing gridded maps. We obtained some robust results using data from there literature, which demonstrate the reliability of PL1.0 in finding regions of common extreme values among two or more maps of a single or different variables measured over the same area or over partially overlapping areas. A major advantage of this procedure is the ability to (i) prove the spatial correspondence of “anomalous” regions for up to ten variables measured simultaneously; and (ii) highlight the persistence or the migration in time of an “anomalous” region for a single variable, measured up to ten times. Similarly, the procedure allows identification of regions of the domain where the largest values of a variable (1, 2 or 3 standard deviations above the mean) are spatially correlated with the lowest values (1, 2 or 3 standard deviations below the mean) of another variable. PL1.0 is able to accomplish these tasks for data with different measurement units and sampling densities. The quantitative correspondence among resulting areas is indicated by a fitting index which is the measurement of how well they overlap.

Although the spatial correspondence of mapped variables is a frequent demand of geologists and scientists in the environmental field only a few attempts have been made by these communities to do more than make comparisons “by eye”. The purpose of this study is to provide a user-friendly and suitable web tool for approaching this topic without using GIS platforms or complex software for which a solid base of knowledge is necessary.

As the next step, our objective is to implement the proposed procedure in cloud platforms for the analysis of geospatial datasets (i.e. Google Earth Engine, GEE) in order to reach a wider audience.

Acknowledgements. We wish to thank Amanda B. Clarke of the Arizona State University for her critical reading and helpful suggestions. We thank the associate Editor and two anonymous reviewers for their comments that greatly improved the manuscript.

APPENDIX

Appendix A. Input and output file format

The input files have to be in the ASCII grid format (.grd). The “.grd” file is described below after modification from the tutorial of the Surfer Golden® software (vers. 8.0).

ASCII grid files (.grd) contain five header lines that provide information about the size and limits of the grid, followed by a list of Z values. The fields within ASCII grid files must be space delimited.

The listing of Z values follows the header information in the file. The Z values are stored in row-major order starting with the minimum Y coordinate. The first Z value in the grid file corresponds to the lower left corner of the map. This can also be thought of as the southwest corner of the map, or, more specifically, the grid node of minimum X and minimum Y. The second Z value is the next adjacent grid node in the same row (the same Y coordinate but the next higher X coordinate). When the maximum X value is reached in the row, the list of Z values continues with the next higher row, until all the rows of Z values have been included. The general format of an ASCII grid file is:

id (the identification string DSAA that identifies the file as an ASCII grid file);

nx ny (where nx is the integer number of grid lines along the X axis (columns) and ny is the integer number of grid lines along the Y axis (rows));

xlo xhi (where xlo is the minimum X value of the grid and xhi is the maximum X value of the grid);

ylo yhi (where ylo is the minimum Y value of the grid and yhi is the maximum Y value of the grid);

zlo zhi (where zlo is the minimum Z value of the grid and zhi is the maximum Z value of the grid).

Then:

grid row 1

grid row 2

grid row 3

...

These are the rows of Z values of the grid, organized in row order. Each row has a constant Y coordinate. Grid row 1 corresponds to ylo and the last grid row corre-

sponds to yhi. Within each row, the Z values are arranged from xlo to xhi.

The following example grid file is ten rows high by ten columns wide. The first five lines of the file contain header information. X ranges from 0 to 9, Y ranges from 0 to 7, and Z ranges from 25 to 97.19. The first Z value shown corresponds to the lower left corner of the map and the following values correspond to the increasing X positions along the bottom row of the grid file. This file has a total of 100 Z values.

```
DSAA
10 10
0.0 9.0
0.0 7.0
25.00 97.19
91.03 77.21 60.55 46.67 52.73 64.05 41.19 54.99 44.30
25.00
96.04 81.10 62.38 48.74 57.50 63.27 48.67 60.81 51.78
33.63
92.10 85.05 65.09 53.01 64.44 65.64 52.53 66.54 59.29
41.33
94.04 85.63 65.56 55.32 73.18 70.88 55.35 76.27 67.20
45.78
97.19 82.00 64.21 61.97 82.99 80.34 58.55 86.28 75.02
48.75
91.36 78.73 64.05 65.60 82.58 81.37 61.16 89.09 81.36
54.87
86.31 77.58 67.71 68.50 73.37 74.84 65.35 95.55 85.92
55.76
80.88 75.56 74.35 72.47 66.93 75.49 86.39 92.10 84.41
55.00
74.77 66.02 70.29 75.16 60.56 65.56 85.07 89.81 74.53
51.69
70.00 54.19 62.27 74.51 55.95 55.42 71.21 74.63 63.14
44.99
```

Output files

Output files are in ASCII grid format (.grd) such as the input files. They have to be saved (and eventually renamed) as “namefile.grd” and can be read by several commercial plotting software (e.g., Surfer Golden ©) or by GIS tools.

REFERENCES

- Burrough P.A. and McDonnell R.A. (1998). *Principles of Geographic Information System*, Oxford, Oxford University Press, 330pp.
- Cardellini, C., Chiodini, G., Frondini, F. (2003). 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.
- Chiodini, G., Caliro S., Cardellini C., Avino R., Granieri D., Schmidt A. (2008). Carbon isotopic composition of soil CO₂ efflux, a powerful method to discriminate different sources feeding soil CO₂ degassing in volcanic-hydrothermal areas. *Earth Planet. Sci. Lett.*, 274, 372-379, doi:10.1016/j.epsl.2008.07.051.
- Cinti, D., Poncia, P. P., Brusca, L., Tassi F., Quattrocchi F., Vaselli O. (2015). Spatial distribution of arsenic, uranium and vanadium in the volcanic-sedimentary aquifers of the Vicano-Cimino Volcanic District (Central Italy). *Journal of Geochemical Exploration* 152, 123-133.
- Coppock, J. T., and Rhind, D. W. (1991). The history of GIS. *Geographical information systems: Principles and applications*, 1(1), 21-43.
- Costa, A., Macedonio, G., Chiodini, G. (2005). Numerical model of gas dispersion emitted from volcanic sources. *Ann. Geophys.* 48, 805-815.
- Costa, A., Macedonio, G. (2016). DISGAS: A model for passive DISpersion of GAS. *Rapporti tecnici INGV*, n. 332, Istituto Nazionale Di Geofisica e Vulcanologia, Italy. <http://datasim.ov.ingv.it/models/disgas.html>.
- Deutsch C.V., Journel A.G. (1998). *GSLIB, geostatistical software library and users guide*. Oxford Univ Press, New York.
- EC Directive (1998). Council Directive 98/83/EC of 3 November 1998 on the Quality of Water Intended for Human Consumption. European Commission, Brussels.
- Frazzetta, G., La Volpe, L. (1991). Volcanic history and maximum expected eruption at La Fossa di Vulcano (Aeolian Island, Italy). *Acta Vulcanol.* 1, 107-113.
- Goovaerts P. (2001). Geostatistical modeling of uncertainty in soil science. *Geoderma*, 103, 3-26.
- Granieri, D., Chiodini, G., Marzocchi, W., Avino, R. (2003). Continuous monitoring of CO₂ soil diffuse degassing at Phlegraean Fields (Italy): influence of environmental and volcanic parameters. *Earth Planet. Sci. Lett.*, 212, 167-179. doi:10.1016/S0012-821X(03)00232-2
- Granieri, D., Carapezza, M.L., Chiodini, G., Avino, R., Caliro, S., Ranaldi, M., Ricci, T., Tarchini, L. (2006). Correlated increase in CO₂ fumarolic content and diffuse emission from La Fossa crater (Vulcano, Italy): Evidence of volcanic unrest or increasing gas release from a stationary deep magma body? *Geophys. Res. Lett.* 33, L13316.

- doi:10.1029/2006GL026460.
- Granieri, D., Avino, R., Chiodini G. (2010). Carbon dioxide diffuse emission from the soil: ten years of observations at Vesuvio and Campi Flegrei (Pozzuoli), and linkages with volcanic activity. *Bulletin of Volcanology* 72, 103-118.
- Granieri, D., Costa, A., Macedonio, G., Bisson, M., Chiodini, G. (2013). Carbon dioxide in the urban area of Naples: Contribution and effects of the volcanic source. *J. Volcanol. Geotherm. Res.* 260, 52-61. doi:10.1016/j.jvolgeores.2013.05.003
- Granieri, D., Carapezza, M.L., Barberi, F., Ranaldi, M., Ricci, T., Tarchini, L. (2014). Atmospheric dispersion of natural carbon dioxide emissions on Vulcano Island, Italy. *J. Geophys. Res. Solid Earth* 119. doi: <http://dx.doi.org/10.1002/2013JB010688>.
- Granieri, D., Vita, F., Inguaggiato, S. (2017). Volcanogenic SO₂, a natural pollutant: measurements, modeling and hazard assessment at Vulcano Island (Aeolian Archipelago, Italy). *Environmental Pollution* 231, 219-228.
- Inguaggiato, S., Diliberto I.S., Federico C., Paonita A., Vita F. (2018). Review of the evolution of geochemical monitoring, networks and methodologies applied to the volcanoes of the Aeolian Arc (Italy). *Earth-Science Reviews*, 176, 241-276
- Isaia, R., Vitale, S., Di Giuseppe, M.G., Iannuzzi, E., Tramparulo, F.D.A., Troiano, A. (2015). Stratigraphy, structure, and volcano-tectonic evolution of Solfatara maar-diatreme (Campi Flegrei, Italy). *GSA Bulletin* 9/10, 1485-1504. doi: 10.1130/B31183.1
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, 547-579.
- Krige, D.G., 1966. Two-dimensional weighted moving average trend surfaces for ore evaluation. *J. South. Afr. Inst. Min. Metall.* 66, 13-38.
- Matheron, G., 1963. Principles of geostatistics. *Econ. Geol.* 58, 1246-1266.
- Matheron, G., 1965. Les variables régionalisées et leur estimation. Masson, Paris.
- Paonita, A., Federico, C., Bonfanti, P., Capasso, G., Inguaggiato, S., Italiano, F., Madonia, P., Pecoraino, G., Sortino, F. (2013). The episodic and abrupt geochemical changes at La Fossa fumaroles (Vulcano Island, Italy) and related constraints on the dynamics, structure, and compositions of the magmatic system. *Geochim. Cosmochim. Acta* 120, 158-178. doi:10.1016/j.gca.2013.06.015.
- Vita, F., Inguaggiato, S., Bobrowski, N., Calderone, L., Galle, B., Parello, F. (2012). Continuous SO₂ flux measurements for Vulcano Island, Italy. *Ann. Geophys.* 55, 2. doi: 10.4401/ag-5759.

*CORRESPONDING AUTHOR: Domenico GRANIERI,

Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa

Pisa, Italy

email: domenico.granieri@ingv.it

© 2018 the Istituto Nazionale di Geofisica e Vulcanologia.

All rights reserved.