



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

**ACCEPTED ON ANNALS OF GEOPHYSICS, 61, 2018; Doi:
10.4401/ag-7813**

**PeakLocator 1.0, a web tool to compare extreme value areas
among maps**

Domenico Granieri, Mattia de' Michieli Vitturi, Marina Bisson

Istituto Nazionale di Geofisica e Vulcanologia, Pisa, Italy

1 **PeakLocator 1.0, a web tool to compare extreme value areas among maps**

2

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

4

5 ¹ Istituto Nazionale di Geofisica e Vulcanologia, sezione di Pisa, via della Faggiola 32-56126 Pisa,
6 Italy

7

8 *Corresponding author.

9 Granieri Domenico, Istituto Nazionale di Geofisica e Vulcanologia, sezione di Pisa, via della
10 Faggiola, 32-56126 Pisa, Italy

11 E-mail address: domenico.granieri@ingv.it

12 phone +39 050 8311958

13 fax +39 050 8311942

14

15 Subject Classification: Numerical code, map similarity, spatial correspondence, Python web
16 application

17 **Abstract**

18 We present here a simple web application, PeakLocator 1.0 (hereafter referred to as PL1.0), for the
19 analysis of gridded geo-located maps. In the present version of the code, the maps can contain up to
20 10 different variables with different units, not necessarily measured at the same locations, as well as
21 the same variable recurrently measured in the time. The aim of PL1.0 is to identify regions where
22 values lie outside the standard deviation from average values. The degree of spatial correspondence
23 between these regions is reflected in the “fitting index” associated to the overlapping area. Here we
24 demonstrate some possible applications of PL1.0 using published datasets, although its potential
25 applicability extends to wide range of topics where the common demand is the comparison of two
26 or more variables mapped over a common area or over areas partially overlapping. PL1.0 is freely
27 accessible through a web interface and runs on any platform.

28 **1.0 Introduction**

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

35 The representation of spatial data was born during the late 1960s [Coppock and Rhind 1991]
36 but only 25 years later the main principles and concepts were fully summarized in the reference
37 work of Burrough and McDonnell [1998]. Burrough and McDonnell [1998] describe the spatial
38 modelling of discrete and continuous data using the vector and raster representation and applying
39 the main analysis of map algebra. The authors mention also geostatistics as an appropriate tool to
40 treat spatial data in the environmental field. Geostatistical techniques were developed to estimate
41 changes in ore grade within a mine using discrete observations at specific locations, along with the
42 associated errors in the estimates [Matheron 1963; 1965]. Since the works of Matheron,
43 geostatistics has evolved but has always remained faithful to the notion of *regionalized variable* that
44 has spatial continuity from point to point, unlike random variable, but whose changes are not fully
45 describable by a deterministic function. This approach uses discrete observations at specific
46 locations to derive a variogram, which represents the spatial rate of change of the variable, which is
47 then used to estimate values of the variable for the entire area of interest. The estimation procedure
48 is called kriging, after Krige [1966], also a pioneering work in the application of geostatistical
49 techniques to estimate changes in ore grade within a mine.

50 In the 1970s and 1980s, practitioners had to write their own code for geospatial analysis, but
51 in the last 30 years powerful software has become widely and cheaply available in the public
52 domain. At the present, such analysis can be done using Statistic or MATLAB or with GIS
53 (Geographic Information System) platforms such as ESRI, GRASS, QGIS. These GIS platforms are

54 able to process discrete or continuous data by using geo-coded vector or raster formats. They
55 provide dedicated tools for making comparisons between variables by using mainly map algebra
56 analyses on grid/matrix formats.

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

80 In the cases presented here, variables were infilled using the sequential Gaussian simulation
81 approach (sGs, Deutsch and Journel, 1998), also described in Cardellini et al. [2003]. Variables at
82 each not sampled location were infilled by a random sampling of a Gaussian conditional cumulative
83 distribution function defined on the basis of original data (conditioning) and previously simulated
84 data within its neighbourhood (sequentiality). Also, while kriging provides the variance of each
85 local estimate, differences among many sGs realizations can be used as a quantitative measure of
86 the associated spatial uncertainty [Deutsch and Journel 1998; Goovaerts 2001; Cardellini et al
87 2003].

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

89

90 **2.0 Code description**

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

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

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

$$S_i = \frac{Y_i - \mu_i}{\sigma_i}.$$

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

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

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

108 These regions define a masking of the original maps. It is important to note that the exceeding
 109 coefficients can have different values for the different maps; in this way, it is possible, for example,
 110 to search for the correlation between a higher value region of the first map ($\alpha_1 > 0$) with a lower
 111 value region of the second map ($\alpha_2 < 0$).

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

$$\bar{A} = \bigcap_i A_{i,\alpha_i}$$

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

$$I_\alpha = \frac{\bigcap_i A_{i,\alpha_i}}{\bigcup_i A_{i,\alpha_i}}$$

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

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

$$\gamma = \min_i [sign(\alpha_i) S_i]$$

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

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

125

126 **3.0 Applications**

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

129

130 **3.1. Case 1 – Finding direct correlation between higher value areas**

131 The goal of this application is the identification of common areas where the values of the
132 considered variables are at least one standard deviation above the respective averages. The maps to
133 be compared can have the same or different units and the same or different sampling density.
134 Possible practical benefits are: i) find the persistence in time of areas of extreme values or assess
135 their spatial migration; ii) select a best site to install a monitoring station; iii) find the spatial
136 correspondence among higher values of two variables or iv) more than two variables.

137 i) As a first example of the usage of PL1.0, we investigated the persistence or the spatial shift of
138 areas characterized by higher values of the soil CO₂ flux inside the crater of Solfatara (Campi
139 Flegrei, Italy). In this area, CO₂ diffuse degassing from the soil has been monitored since January
140 1998 by periodic CO₂ flux measurements over an array of fixed stations covering an area of
141 ~125.000 m² in the flat floor of the crater (Figure 2a). With time, the array has become gradually
142 more dense by increasing the number of fixed stations from 30 to 71 [Granieri et al. 2010], without
143 significant enlargement of the investigated area. For this application, we used the dataset of
144 measurements over the primary 30-point array that was repeated during 157 campaigns from
145 January 1998 to September 2007. The average value of each campaign is reported in Figure 2b.

146 Owing to the different frequency of sampling (weekly in 1998 and then almost monthly in the
147 following years), the campaigns are irregularly spaced in the time [Granieri et al. 2010].

148 We ordered the campaigns in chronological order from 1 to 157 (the numbers at the top of the graph
149 of Figure 2b are referred to some of these campaigns) and 10 of these were randomly selected as
150 training data through a procedure of generating random numbers in the appropriate interval (1-157),
151 with the condition that any duplication is avoided. The 10 maps that came up (10, 34, 46, 63, 72,
152 96, 105, 123, 143, 149, indicated by red dots in Figure 2b) were processed through PL1.0 and the
153 overall result, in terms of pixels exceeding by at least one standard deviation the mean value ($\alpha_i=1$
154 for all the maps), is shown in Figure 3. We remark again that the plotted value is the minimum
155 among the normalized values of all the maps.

156 Although the campaigns, randomly selected, are characterized by different average values (cfr.
157 Figure 2b), one main common area of high values is the fumarole area of SOL1 and STUFE (zone 1
158 in Figure 3), located at the intersection of the NW-SE and NE-SW main fault systems of the
159 Solfatara crater [Isaia et al. 2015], and a less extended area of high values is in the center of the
160 crater (zone 2), near to the Fangaia mud pool. We observe that for this test the fitting index is quite
161 low ($I_\alpha \sim 0.028$), expressing the fact that the resulting common area (1653 m²) represents only a
162 small portion (approximately 3%) of the area where all maps exceeded the respective average
163 values by one standard deviation. Although of limited size, the persistence through time of these
164 areas displaying significant CO₂ emission, could suggest a better connection between the surface
165 and the feeding hydrothermal system of Solfatara than surrounding zones.

166 ii) We used another test bed to demonstrate PL1.0 as a tool for selecting optimal sites for automated
167 continuous sampling stations, starting from a limited number of explorative field surveys. In the
168 case presented here, we investigated the CO₂ flux from the soil at La Fossa crater of Vulcano
169 (Aeolian Archipelago, Italy). CO₂ soil flux is monitored here by periodic field surveys over a
170 discrete number of sample locations and continuously through automated stations [Inguaggiato et

171 al. 2018]. A similar combined approach is applied in different volcanoes worldwide [Granieri et al.
172 2003], e.g., at Campi Flegrei, Vesuvio, Etna, Stromboli, Vulcano (Italy), Masaya (Nicaragua), Poa's
173 (Costa Rica), La Palma (Canary Islands, Spain), San Salvador, San Miguel, San Vincente and Santa
174 Ana (El Salvador), Usu (Japan), Mammoth Mountain (California, USA). Some of these experiences
175 have fallen short due to non-ideal choices of automated measurement sites. In fact, if the purpose of
176 the observation is to monitor deeper volcanic processes and changes in the levels of volcanic
177 activity in real time, the device must be placed in a sector of the crater where the degassed CO₂ is
178 derived from the hydrothermal/magmatic system, as mainly supported by measured fluxes orders of
179 magnitude higher than the values of the local biological CO₂ background (typically around 20-50
180 gm⁻²d⁻¹ for volcanic soils, Chiodini et al. 2008, Granieri et al. 2003). Furthermore, the station needs
181 to be in a location not too exposed to volcanic gases, generally acidic and with detrimental impacts
182 on measuring sensors and electronics, and easily accessible to operators for maintenance.

183 Since the last eruption of 1888-1890, the crater of Vulcano showed an intensive and persistent
184 degassing activity, with periods of enhanced degassing involving both the fumaroles and the soil [
185 Granieri et al. 2006, 2014, Paonita et al. 2013]. As reported in Granieri et al. [2014], a CO₂ output
186 of 216 (± 83) tons per day (td⁻¹) represents the “background” soil CO₂ emission from the crater in
187 the present quiescent stage of the volcano whereas higher CO₂ emissions were recorded in
188 December 2004 and December 2005, with CO₂ emission rates of 700 td⁻¹ and 1600 td⁻¹, respectively
189 [Granieri et al. 2014]. These values were estimated on the basis of fifteen field surveys performed
190 during the period from 1995 to 2010, each of them based on a different number of samples, taken at
191 different locations approximately over the same crater area of ~1 km² through the “accumulation
192 chamber” portable device [Granieri et al. 2014].

193 We started from 4 maps of soil CO₂ flux from April 1995, July 1998, December 2004, and
194 December 2005 at La Fossa crater of Vulcano and reported in Granieri et al. [2014]. They are
195 indicative of the mean CO₂ output (266 td⁻¹ in April 1995), low CO₂ output (162 td⁻¹ in July 1998),
196 high CO₂ output (700 td⁻¹ in December 2004), and very high CO₂ output (1600 td⁻¹ in December

197 2005). These maps, derived through the sGs approach [Cardellini et al. 2003], are shown in Figure
198 4.

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

201 Results from the comparison of pairs of maps (Figure 5a,b,c) highlighted that regions exceeding the
202 mean values by 1 standard deviation are always contained within the present crater rim (crater A) or
203 in a restricted sector outside to the crater rim, along the NE edge (encompassed by the crater A and
204 the paleo-crater B). The combined map of all four surveys confirmed the existence of two common
205 highest-emission areas (a1 and a2 in Figure 5d, for a total surface of 3458 m²), representing
206 cumulatively about 2% of the total area ($I_{\alpha} \sim 0.02$), where all maps exceed the respective mean
207 values by at least 1 standard deviation. The low value of the fitting index likely results from the
208 appearance of new higher values in the southern sector of the crater after the first period of
209 observation. Although a1 and a2 areas are both formally suitable for the placement of an automatic
210 station, other considerations concerning the safe distance to preserve sensors and electronics of the
211 station and a more comfortable position for maintenance might encourage the operator to select the
212 area a1 as the most appropriate. In reality, an automated station to monitor the soil diffuse CO₂
213 degassing at the summit of La Fossa cone (VSCS station, Figure 5) is located in a site near to a1
214 area since September 2007 [Inguaggiato et al. 2018].

215 iii) Starting from different levels of soil CO₂ diffuse degassing recently measured at the crater of
216 Vulcano (from 162 td⁻¹ in July 1998 to 1600 td⁻¹ in December 2005), we performed a further test of
217 PL1.0 in an attempt to identify common regions of the domain where air CO₂ concentrations, whose
218 levels are mainly due to the volcanic source [Granieri et al. 2014], exceed the respective mean
219 values by 1 standard deviation. In order to model CO₂ plume dispersion, we applied the DISGAS
220 code [Costa et al. 2005, 2016, Granieri et al. 2013, 2014, 2017]. Despite being rare at Vulcano, for
221 this test we imposed a wind from SSE since it draws a dangerous situation for the village of
222 Vulcano Porto lying downwind with respect to this wind direction. The average speed of 1.87 m/s

223 was extracted from the dataset of a local meteorological station (Lentia station), covering the 8-year
224 time period from May 2008 to February 2016 (for detail see Vita et al. 2012 and Granieri et al.
225 2017). Resulting maps of the plume dispersion for four periods are reported in Figure 6.

226 Results show that the common area of modeled extreme values of ambient CO₂ concentration is
227 encompassed by the rim of the present crater with the exception of a lobe in the external NE sector
228 (Figure 7). The inhabited area of Vulcano Porto is substantially unaffected by the crater-derived
229 CO₂ plume, confirming the finding of Granieri et al. [2014]. The large value of the fitting index (I_{α}
230 ~ 0.60) obtained comparing the four maps, confirmed that the region of the higher CO₂
231 concentration, essentially concentrated in the present crater rim (A in Figure 5), persists in time,
232 despite different degassing levels measured at the volcano.

233 iv) In this application of PL1.0 we were interested in the spatial correlation between three variables,
234 having the same measurement units and different statistics (mean, standard deviation, skewness,
235 kurtosis, etc.), which were simultaneously measured at the same locations. In this specific case, we
236 used concentrations (in $\mu\text{g/L}$) of arsenic (As), uranium (U) and vanadium (V) in 328 water samples,
237 collected in the Vicano-Cimino volcanic district (central Italy). Data were taken from Cinti et al.
238 [2015] and they are provided as a test case in the Supplementary material. In some areas of central
239 Italy, the concentration of heavy metals in drinkable waters can pose a hazard for people, in
240 particular the high level of As, often above the threshold imposed by a specific European directive
241 [EC Directive 1998]. Although not regulated by the countries of the European Union, high
242 concentrations of U and V can be toxic, and for this reason they were measured in the area along
243 with the As concentration. The main factor controlling the concentration of As, U and V in waters
244 resulted to be a combination of the lithology where waters move together with the thermal and
245 redox conditions of the water-rock interaction [Cinti et al. 2015]. Maps derived from the data of
246 Cinti et al. [2015] are shown in Figure 8,a,b,c.

247 By correlating the entire dataset, it is not possible to find any significant correlation among the pairs
248 of the three elements ($r=+0.12$, -0.03 , 0.00 for U-V, As-V, As-U pairs, respectively, as reported in
249 the first column of Table 1). The comparison consisted simply of calculating the correlation
250 coefficient between the concentration of variable 1 and 2, 1 and 3, and 2 and 3, with no regard for
251 the spatial position of the samples. The main disadvantage of this procedure is the inability to
252 discriminate if the overall absence of correspondence in the whole area may mask a higher degree
253 of similarity in sub-regions of the domain. In order to explore this possibility, we clustered the
254 samples into four categories on the basis of the water type, as identified by the authors, somehow
255 reflecting the spatial distribution of the samples. Results highlighted a strong positive correlation
256 between U and V in cold-sedimentary ($r=+0.63$) and thermal waters ($r=+0.28$), a minor correlation
257 in cold-volcanic waters ($r=+0.17$), and a low (and negative) correlation in bubbling pools ($r=-0.06$).
258 Similarly, a strong correlation comes out between As and V and between As and U in cold-
259 sedimentary waters ($r=+0.83$ and $+0.68$, respectively), moderate or low correlation in cold-volcanic
260 waters ($r=+0.06$ and $+0.13$, respectively), and low (and negative) correlation in thermal waters ($r=-$
261 0.06 and -0.16 , respectively). Table 1 summarizes these relationships.

262 PL1.0 was applied for the As-U, As-V, and U-V pairs, starting from the whole datasets of 328 water
263 samples. For this application, the units and the sampling density of the three variables are the same,
264 but, as previously stated, this is not required for the applicability of the code. Results of the
265 procedure allowed easier identification of common areas of extreme values (Figure 8d), with a
266 fitting index of about 0.15 for the couple U-V, and lower values for As-V and As-U pairs ($I_{\alpha} \sim 0.09$
267 and ~ 0.08 , respectively), confirming the correspondence between higher value areas and water
268 types, but with more spatial information than a simple statistical correlation. In fact, common areas
269 are all found in the spatial domain of the cold-volcanic waters (Figure 8d), where the correlation of
270 the three considered pairs is moderate (second column of Table 1), rather than in the cold-
271 sedimentary waters where the correlation is high (third column of Table 1), suggesting that the high

272 correlation in the cold-sedimentary samples is likely due to the numerically largest group of low-
273 medium concentration values whilst that highest values (“anomalies”, *sensu lato*) are well
274 correlated in the domain of the cold-volcanic waters. As a general conclusion, this means that if the
275 statistical value of a correlation may not be in doubt, the “physical” meaning ascribed to that value
276 is open to interpretation for the peculiar nature of the geochemical data. That is the conclusion
277 reached through the application of PL1.0, without the necessity of knowing a-priori a further
278 parameter, i.e., the water type, on which the clustering and the subsequent correlation “for groups”
279 was based.

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

282 This tool, available in the ArcToolBox extension, presents a calculator-like interface and provides
283 as output a Boolean geo-raster that identifies the areas where a request is satisfied (or not) by
284 combining mathematical and logical operators. The results of the comparison are shown in Figure 9.

285 The anomalous regions obtained with PL1.0 are enclosed by the curves resulting from the GIS tool,
286 showing very good spatial agreement. The results of this comparison, as other ones not shown,
287 confirmed that the code has performed well and it is satisfactorily robust. Then, when a comparison
288 among gridded maps is required, it can provide a reliable alternative to the more established and
289 complicated approach of GIS tools, for which a wealth of knowledge and experience is necessary.

290 In addition, it adds further information about the user’s request. For instance, in the aforementioned
291 case study the GIS tool pulls out the curve enclosing the common areas in which values of the
292 couples are higher than their means + 1 standard deviation (as required) whilst the PL1.0
293 application provides a pixel tessellation with different levels of the exceedence of the mean from +1
294 standard deviation up to +4.5 standard deviations.

295
296 **3.2. Case 2 – Finding inverse correlation between investigated areas**

297 The goal of this application is the identification of common areas where the positive values of one
298 variable (calculated as average value plus 1,2 or 3 standard deviations) correlates with the negative
299 values of one another (calculated as average value minus 1, 2 or 3 standard deviations). Possible
300 practical benefits are: i) identify common areas of the domain where lithological, structural or
301 physico-chemical features of the medium favor the escape of a gas, compound or metal in the near-
302 surface environment and inhibit the output of another.

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

311

312 **3.3. Case 3 – Finding direct correlation between areas with very extreme values (2σ or 3σ** 313 **above/below the mean)**

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

318 This case concerns the comparison between the tails of the value distribution in order to bring out
319 the location of sub-region(s) of the domain characterized by very extreme values of the measured
320 variables. For example, when measurements are *normally* distributed, a quite common peculiarity
321 for geochemical data, no more than 5% of all measurements are included within the interval higher

322 than +2 standard deviations or lower than -2 standard deviations on both sides of the mean, and
323 slightly over 0.1% of all measurements fall within the interval higher than +3 standard deviations or
324 lower than -3 standard deviations.

325 In this application, we consider the dataset of As, U, and V concentration in water samples for
326 values exceeding the respective mean by at least 2 standard deviations. Results of PL1.0 permitted
327 highlighting of a larger zone of high values for the couple U-V (with a fitting index of about 0.1) in
328 proximity to the Tuscania village (Figure 11), where actually the contamination of groundwater
329 posed serious problems for human health [Cinti et al. 2015] and two smaller areas (with negligible
330 values of the fitting index) of simultaneous As-V and As-U highest values near to Viterbo town and
331 Civita Castellana village, respectively (Figure 11).

332 We applied again PL1.0 to the air CO₂ concentration maps (temporally spaced) at La Fossa crater of
333 Vulcano (see Section 3.1), but here considering values exceeding the mean by 2 and 3 standard
334 deviations. Code results highlighted the persistence over time of a common area inside the present-
335 day crater rim (cfr. Figure 6), where, as expected, the common area of very extreme values
336 progressively reduces its extension when larger thresholds (from +1 σ to +3 σ) are considered
337 (Figure 12a,b,c). For this application, even for the largest threshold, a good similarity of common
338 areas for the four CO₂ concentration maps (cfr. Figure 6) is obtained, as suggested by the fitting
339 index of 0.43 (Figure 12c).

340

341 **4. Conclusions**

342 In this study, we describe the new PL1.0 web application for analyzing gridded maps. We obtained
343 some robust results using data from there literature, which demonstrate the reliability of PL1.0 in
344 finding regions of common extreme values among two or more maps of a single or different
345 variables measured over the same area or over partially overlapping areas. A major advantage of
346 this procedure is the ability to (i) prove the spatial correspondence of “anomalous” regions for up to
347 ten variables measured simultaneously; and (ii) highlight the persistence or the migration in time of

348 an “anomalous” region for a single variable, measured up to ten times. Similarly, the procedure
349 allows identification of regions of the domain where the largest values of a variable (1, 2 or 3
350 standard deviations above the mean) are spatially correlated with the lowest values (1, 2 or 3
351 standard deviations below the mean) of another variable. PL1.0 is able to accomplish these tasks for
352 data with different measurement units and sampling densities. The quantitative correspondence
353 among resulting areas is indicated by a fitting index which is the measurement of how well they
354 overlap.

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

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

362

363 **Acknowledgements**

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

367

368 **References**

369 Burrough P.A. and McDonnell R.A. (1998). Principles of Geographic Information System, Oxford,
370 Oxford University Press, 330pp.

371

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

375

376 Chiodini, G., Caliro S., Cardellini C., Avino R., Granieri D., Schmidt A. (2008). Carbon isotopic
377 composition of soil CO₂ efflux, a powerful method to discriminate different sources feeding soil
378 CO₂ degassing in volcanic-hydrothermal areas. *Earth Planet. Sci. Lett.*, 274, 372-379,
379 doi:10.1016/j.epsl.2008.07.051.

380

381 Cinti, D., Poncia, P. P., Brusca, L., Tassi F., Quattrocchi F., Vaselli O. (2015). Spatial distribution
382 of arsenic, uranium and vanadium in the volcanic-sedimentary aquifers of the Vicano-Cimino
383 Volcanic District (Central Italy). *Journal of Geochemical Exploration* 152, 123-133.

384

385 Coppock, J. T., and Rhind, D. W. (1991). The history of GIS. *Geographical information systems:*
386 *Principles and applications*, 1(1), 21-43.

387

388 Costa, A., Macedonio, G., Chiodini, G. (2005). Numerical model of gas dispersion emitted from
389 volcanic sources. *Ann. Geophys.* 48, 805–815.

390

391 Costa, A., Macedonio, G. (2016). DISGAS: A model for passive DISpersion of GAS. *Rapporti*
392 *tecnic* INGV, n. 332, Istituto Nazionale Di Geofisica e Vulcanologia, Italy.
393 <http://datasim.ov.ingv.it/models/disgas.html>.

394

395 Deutsch C.V., Journel A.G. (1998). *GSLIB, geostatistical software library and users guide*. Oxford
396 Univ Press, New York.

397

398 EC Directive (1998). Council Directive 98/83/EC of 3 November 1998 on the Quality of Water
399 Intended for Human Consumption. European Commission, Brussels.

400

401 Frazzetta, G., La Volpe, L. (1991). Volcanic history and maximum expected eruption at La Fossa di
402 Vulcano (Aeolian Island, Italy). *Acta Vulcanol.* 1, 107–113.

403

404 Goovaerts P. (2001). Geostatistical modeling of uncertainty in soil science. *Geoderma*, 103, 3–26.

405

406 Granieri, D., Chiodini, G., Marzocchi, W., Avino, R. (2003). Continuous monitoring of CO₂ soil
407 diffuse degassing at Phlegraean Fields (Italy): influence of environmental and volcanic parameters.
408 *Earth Planet. Sci. Lett.*, 212, 167-179. doi:10.1016/S0012-821X(03)00232-2

409

410 Granieri, D., Carapezza, M.L., Chiodini, G., Avino, R., Caliro, S., Ranaldi, M., Ricci, T., Tarchini,
411 L. (2006). Correlated increase in CO₂ fumarolic content and diffuse emission from La Fossa crater
412 (Vulcano, Italy): Evidence of volcanic unrest or increasing gas release from a stationary deep
413 magma body? *Geophys. Res. Lett.* 33, L13316. doi:10.1029/2006GL026460.

414

415 Granieri, D., Avino, R., Chiodini G. (2010). Carbon dioxide diffuse emission from the soil: ten
416 years of observations at Vesuvio and Campi Flegrei (Pozzuoli), and linkages with volcanic activity.
417 *Bulletin of Volcanology* 72, 103–118.

418

419 Granieri, D., Costa, A., Macedonio, G., Bisson, M., Chiodini, G. (2013). Carbon dioxide in the
420 urban area of Naples: Contribution and effects of the volcanic source. *J. Volcanol. Geotherm. Res.*
421 260, 52–61. doi:10.1016/j.jvolgeores.2013.05.003
422

423 Granieri, D., Carapezza, M.L., Barberi, F., Ranaldi, M., Ricci, T., Tarchini, L. (2014). Atmospheric
424 dispersion of natural carbon dioxide emissions on Vulcano Island, Italy. *J. Geophys. Res. Solid*
425 *Earth* 119. doi: <http://dx.doi.org/10.1002/2013JB010688>.
426

427 Granieri, D., Vita, F., Inguaggiato, S. (2017). Volcanogenic SO₂, a natural pollutant: measurements,
428 modeling and hazard assessment at Vulcano Island (Aeolian Archipelago, Italy). *Environmental*
429 *Pollution* 231, 219-228.
430

431 Inguaggiato, S., Diliberto I.S., Federico C., Paonita A., Vita F. (2018). Review of the evolution of
432 geochemical monitoring, networks and methodologies applied to the volcanoes of the Aeolian Arc
433 (Italy). *Earth-Science Reviews*, 176, 241–276
434

435 Isaia, R., Vitale, S., Di Giuseppe, M.G., Iannuzzi, E., Tramparulo, F.D.A., Troiano, A. (2015).
436 Stratigraphy, structure, and volcano-tectonic evolution of Solfatara maar-diatreme (Campi Flegrei,
437 Italy). *GSA Bulletin* 9/10, 1485–1504. doi: 10.1130/B31183.1
438

439 Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des
440 Jura". *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, 547–579.
441

442 Krige, D.G., 1966. Two-dimensional weighted moving average trend surfaces for ore evaluation.
443 *J. South. Afr. Inst. Min. Metall.* 66, 13–38.
444

- 445 Matheron, G., 1963. Principles of geostatistics. *Econ. Geol.* 58, 1246–1266.
- 446
- 447 Matheron, G., 1965. *Les variables régionalisées et leur estimation*. Masson, Paris.
- 448
- 449 Paonita, A., Federico, C., Bonfanti, P., Capasso, G., Inguaggiato, S., Italiano, F., Madonia, P.,
450 Pecoraino, G., Sortino, F. (2013). The episodic and abrupt geochemical changes at La Fossa
451 fumaroles (Vulcano Island, Italy) and related constraints on the dynamics, structure, and
452 compositions of the magmatic system. *Geochim. Cosmochim. Acta* 120, 158–178.
453 doi:10.1016/j.gca.2013.06.015.
- 454
- 455 Vita, F., Inguaggiato, S., Bobrowski, N., Calderone, L., Galle, B., Parello, F. (2012). Continuous
456 SO₂ flux measurements for Vulcano Island, Italy. *Ann. Geophys.* 55, 2. doi: 10.4401/ag-5759.

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

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

459

460

461 **Figure captions**

462
463 Figure 1. Schematic representation of the PL1.0 procedure considering the maps of the variable 1
464 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,
465 in the insert representing the output, the bounded grey area represents the region where the values
466 of both the variables exceed the mean value by at least 1 standard deviation.

467 Figure 2. a) Array of 30 fixed points inside the crater of Solfatara to monitor the soil CO₂ emission.
468 Fumaroles are also shown; the Fangaia is a mud pool; b) Variation of the soil CO₂ flux in the 1998-
469 2007 period over the 30-point array (average values of 157 campaigns). Red points refer to 10
470 campaigns randomly extracted to be considered in the application of PL1.0.

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

473 Figure 4. Maps of 4 soil CO₂ flux campaigns at the crater of La Fossa (Vulcano island), modified
474 from Granieri et al. [2006]. Dashed rectangle encompasses the area zoomed in Figure 5.

475 Figure 5. Common areas of higher soil CO₂ flux considering pairs of campaigns (a, b, c) and all 4
476 campaigns (d). The rims of the present-day crater (A) and of paleo-craters that were active in
477 different times are indicated (B-Pietre Cotte; C-Palizzi and Commenda; D-Unidentified; E-Punte
478 Nere, from Frazzetta and La Volpe [1991]). In d) common areas are indicated by dashed ellipses (a1
479 and a2) and the position of the automated station (VSCS) operative since 2007 by a yellow box.

480
481 Figure 6. Maps of the air CO₂ concentration at the crater of La Fossa in different times under a SSE
482 wind (see text). The site of the meteorological station (Lentia station) is shown.

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

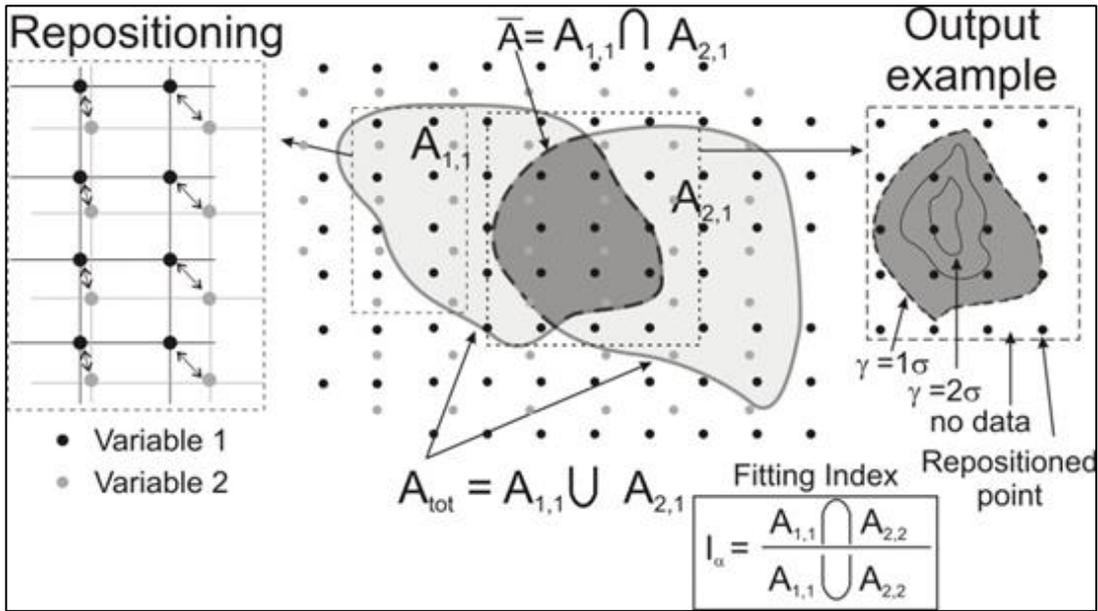
485
486 Figure 8. Maps of the water content of a) arsenic (As), b) uranium (U), and c) vanadium (V). Data
487 from Cinti et al. [2015]; d) common areas of extreme values for pairs of elements are shown with
488 different colors. Dashed lines encompass the different domains of the water type (see Cinti et al.
489 2015).

490
491 Figure 9. Comparison between the GIS tool Raster Calculation (colored curves) and PL1.0 outputs
492 (tessellated areas) for the case study iv (see text).

493
494 Figure 10. Resulting areas where water concentrations of U and As one standard deviation above
495 their respective means coincide with values of the V concentration one standard deviation below its
496 mean.

497
498 Figure 11. Common areas with very extreme values for pairs of elements where their concentrations
499 in water are two standard deviations above their means.

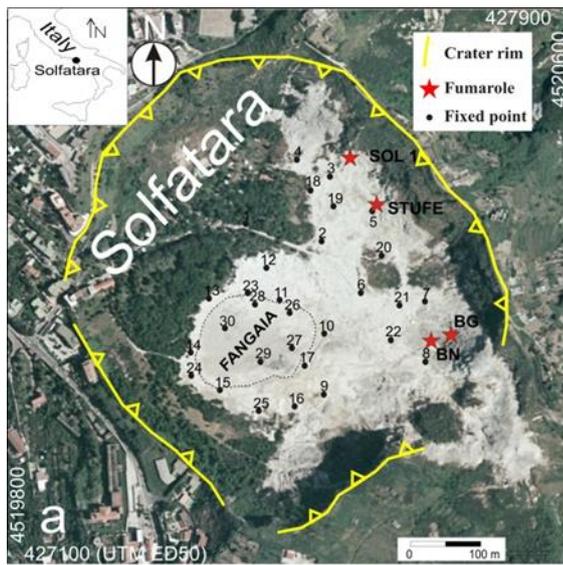
500
501 Figure 12. Common area of air CO₂ concentration considering the four maps of Figure 6 with
502 values exceeding the mean by a) one, b) two, and c) three standard deviations. The map in a) is the
503 same of Figure 7.



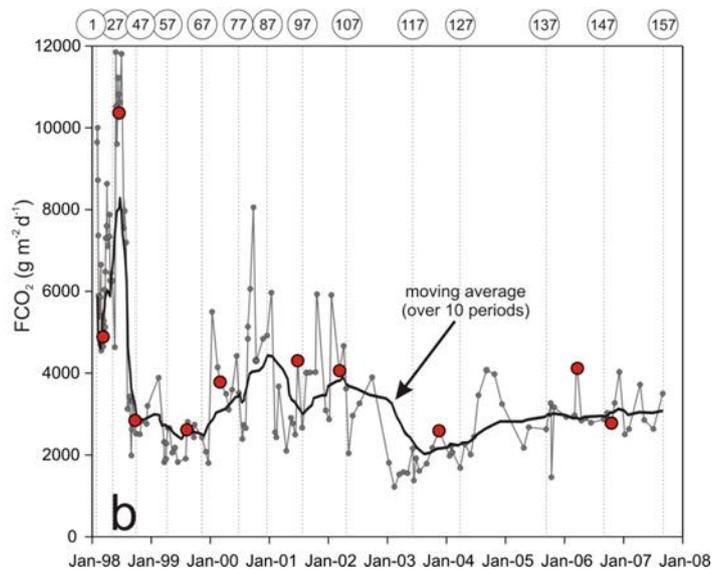
513

514 Figure 1

515

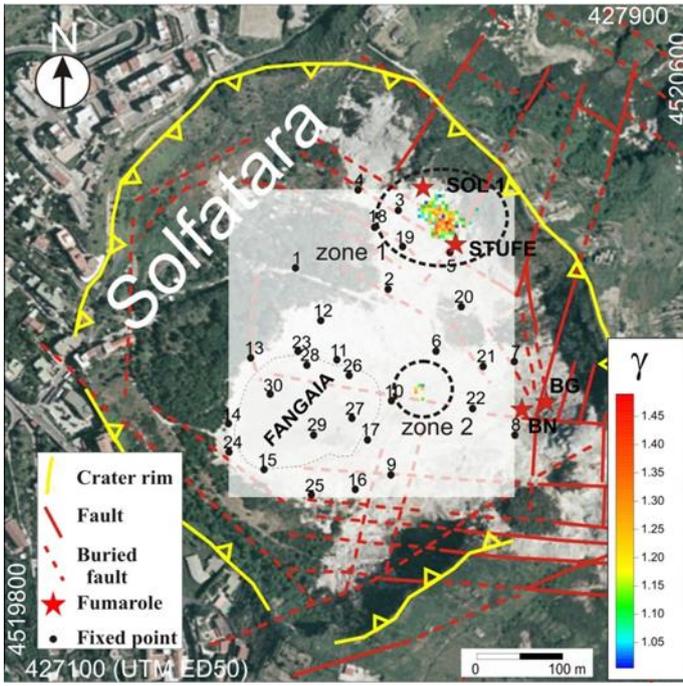


516



517 Figure 2

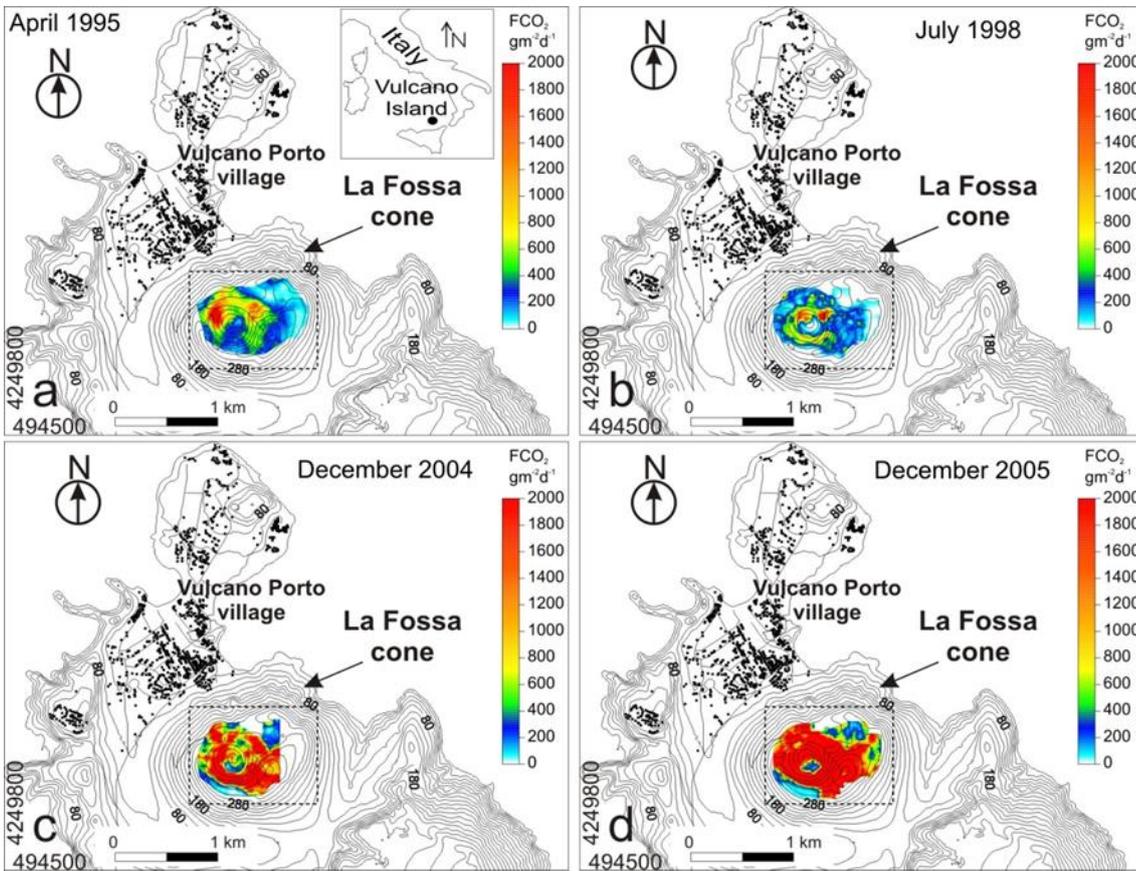
518



519

520 Figure 3

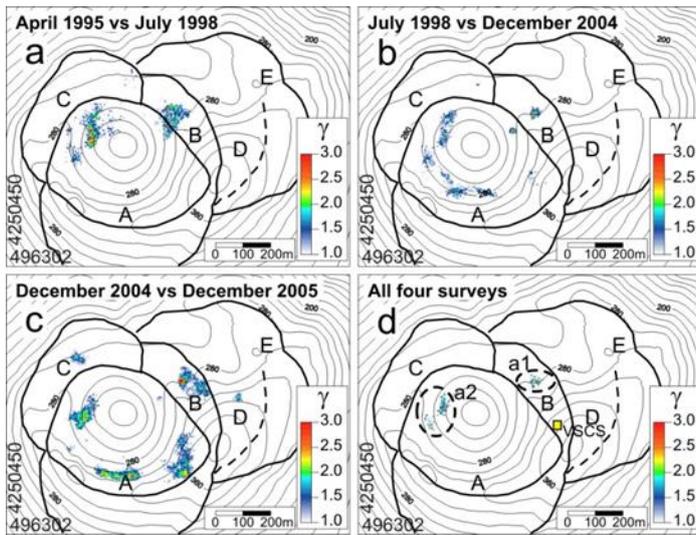
521



522

523 Figure 4

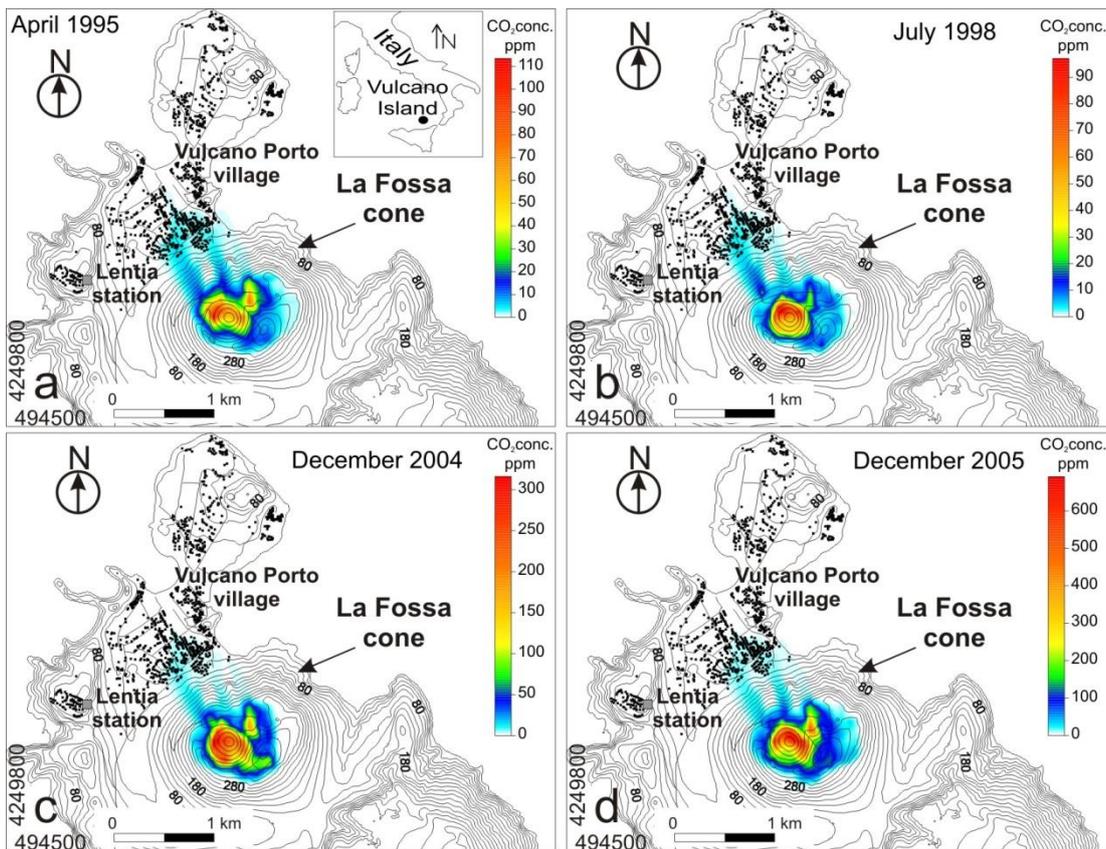
524



525

526 Figure 5

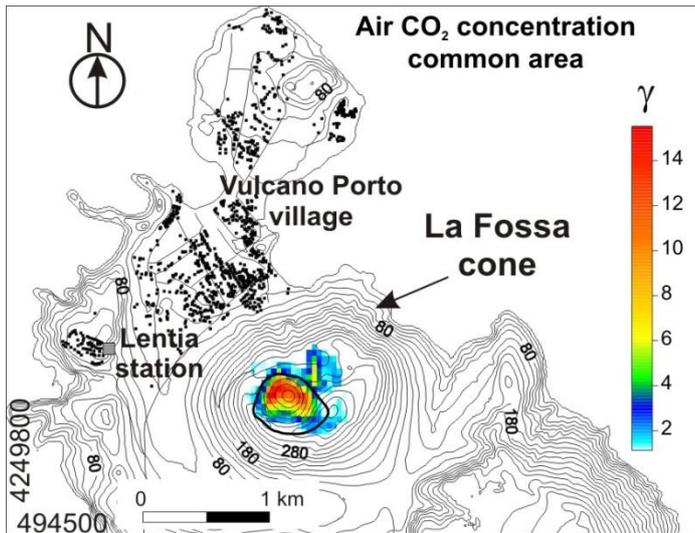
527



528

529 Figure 6

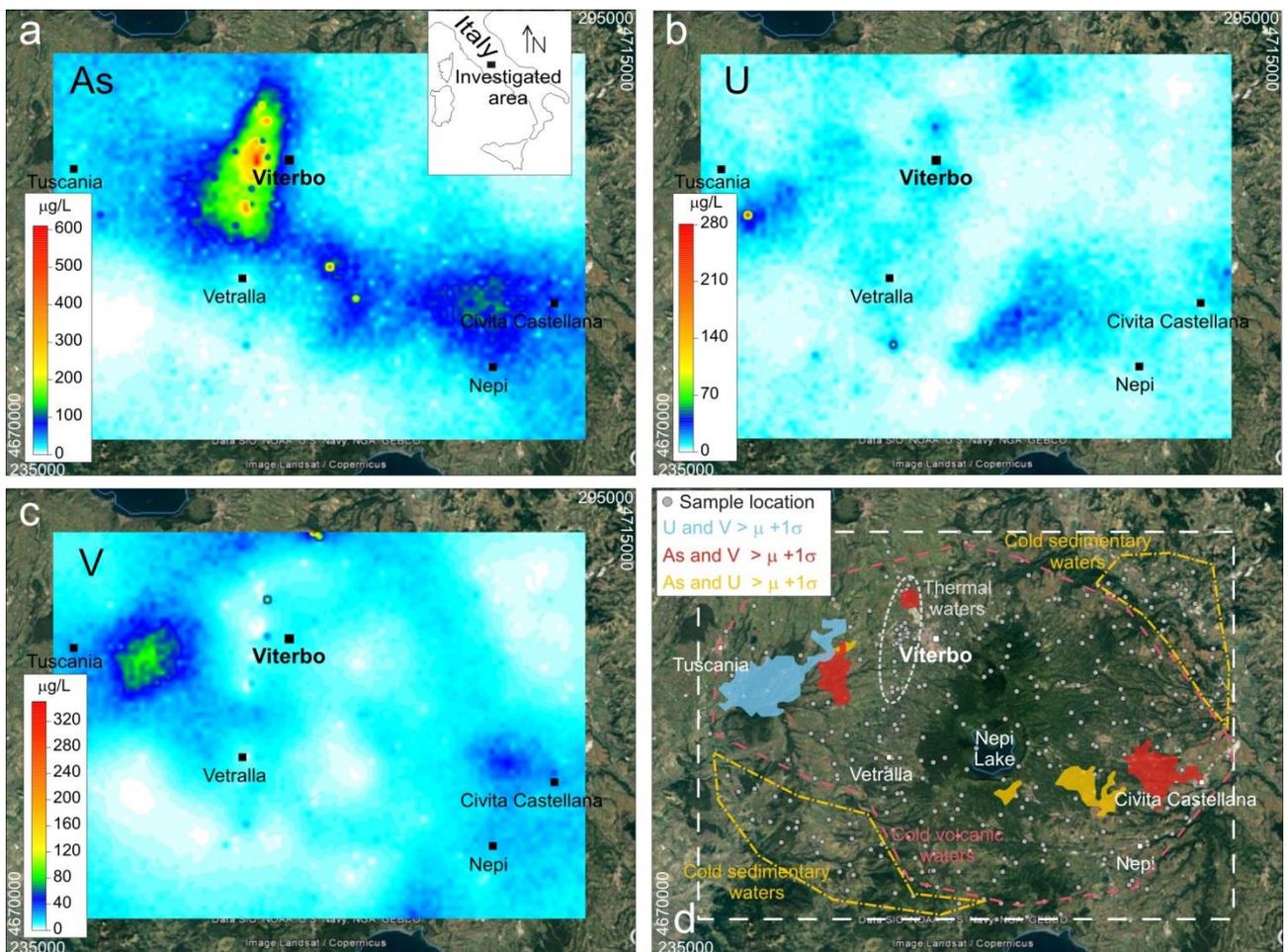
530



531

532 Figure 7

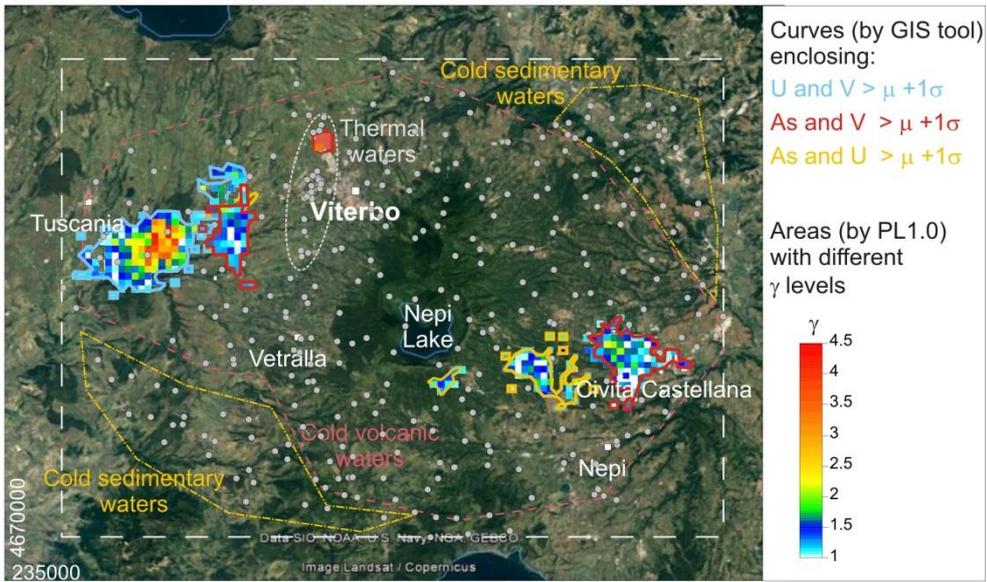
533



534

535 Figure 8

536



537

538 Figure 9

539



540

541 Figure 10

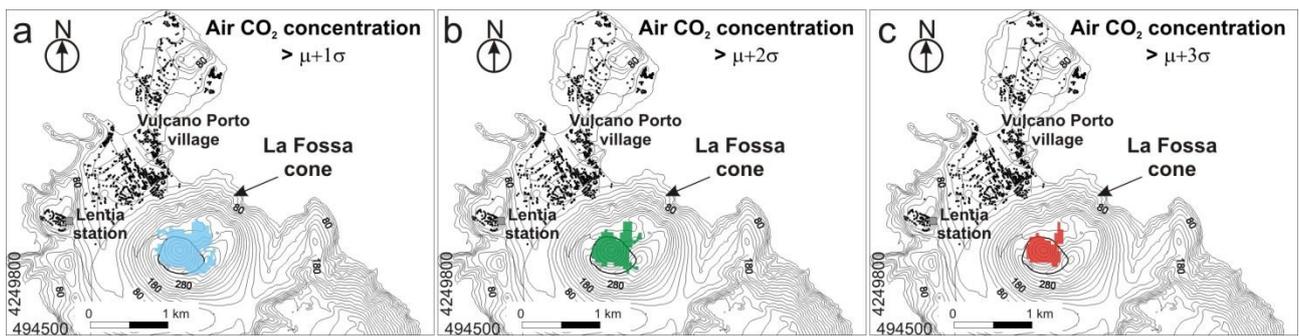
542



543

544 Figure 11

545



546

547 Figure 12

548 **Appendix**

549

550 **Appendix A. Input and output file format**

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

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

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

562 The general format of an ASCII grid file is:

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

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

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

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

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

569

570 Then:

571 grid row 1

572 grid row 2

573 grid row 3

574 ...

575 These are the rows of Z values of the grid, organized in row order. Each row has a constant Y
576 coordinate. Grid row 1 corresponds to ylo and the last grid row corresponds to yhi. Within each
577 row, the Z values are arranged from xlo to xhi.

578 The following example grid file is ten rows high by ten columns wide. The first five lines of the file
579 contain header information. X ranges from 0 to 9, Y ranges from 0 to 7, and Z ranges from 25 to
580 97.19. The first Z value shown corresponds to the lower left corner of the map and the following
581 values correspond to the increasing X positions along the bottom row of the grid file. This file has a
582 total of 100 Z values.

```
583 DSAA  
584 10 10  
585 0.0 9.0  
586 0.0 7.0  
587 25.00 97.19  
588 91.03 77.21 60.55 46.67 52.73 64.05 41.19 54.99 44.30 25.00  
589 96.04 81.10 62.38 48.74 57.50 63.27 48.67 60.81 51.78 33.63  
590 92.10 85.05 65.09 53.01 64.44 65.64 52.53 66.54 59.29 41.33  
591 94.04 85.63 65.56 55.32 73.18 70.88 55.35 76.27 67.20 45.78  
592 97.19 82.00 64.21 61.97 82.99 80.34 58.55 86.28 75.02 48.75  
593 91.36 78.73 64.05 65.60 82.58 81.37 61.16 89.09 81.36 54.87  
594 86.31 77.58 67.71 68.50 73.37 74.84 65.35 95.55 85.92 55.76  
595 80.88 75.56 74.35 72.47 66.93 75.49 86.39 92.10 84.41 55.00  
596 74.77 66.02 70.29 75.16 60.56 65.56 85.07 89.81 74.53 51.69  
597 70.00 54.19 62.27 74.51 55.95 55.42 71.21 74.63 63.14 44.99  
598
```

599 Output files

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