



Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: [www.elsevier.com/locate/jappgeo](http://www.elsevier.com/locate/jappgeo)

## A geophysical $k$ -means cluster analysis of the Solfatara–Pisciarelli volcano-geothermal system, Campi Flegrei (Naples, Italy)

M.G. Di Giuseppe <sup>a,\*</sup>, A. Troiano <sup>a</sup>, D. Patella <sup>a,b</sup>, M. Piochi <sup>a</sup>, S. Carlino <sup>a</sup>

<sup>a</sup> Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Napoli Osservatorio Vesuviano, Naples, Italy

<sup>b</sup> Department of Physics, University Federico II, Naples, Italy

### ARTICLE INFO

#### Article history:

Received 21 January 2016

Received in revised form 17 March 2017

Accepted 2 June 2017

Available online xxxx

#### Keywords:

$k$ -means

Multivariate clustering

Geophysical parameters

Volcano-geothermal system

Solfatara–Pisciarelli

Campi Flegrei

### ABSTRACT

We present an application of the  $k$ -means cluster analysis technique to a trivariate geophysical parameter set distributed over a 3 km thick and 1.2 km long section through the geothermally active Solfatara–Pisciarelli area of the Campi Flegrei caldera (Italy). The data set consists of resistivity, P-wave velocity and density parameters, extracted from previously modeled univariate sections through the selected profile and mapped in a joint parameter 3D space. Six compact and separated clusters have been identified. A subsequent linear regression statistics has allowed two distinct branches of clusters to be recognized, each represented in the joint parameter space by the equation of a plane. Each equation and its partial derivatives define the peculiar combinations among the used parameters. Once mapped back to the space domain, the final image has provided a grouping of the structures in two superimposed blocks with an irregular interface located at 2 km mean depth. Based on drilling data, subsurface stratigraphy and recent seismic activity, the interface between the two blocks has been interpreted in terms of a substantial change in the rocks rheology inside the Solfatara–Pisciarelli volcano-geothermal system, namely the brittle-to-ductile transition.

© 2017 Published by Elsevier B.V.

### 1. Introduction

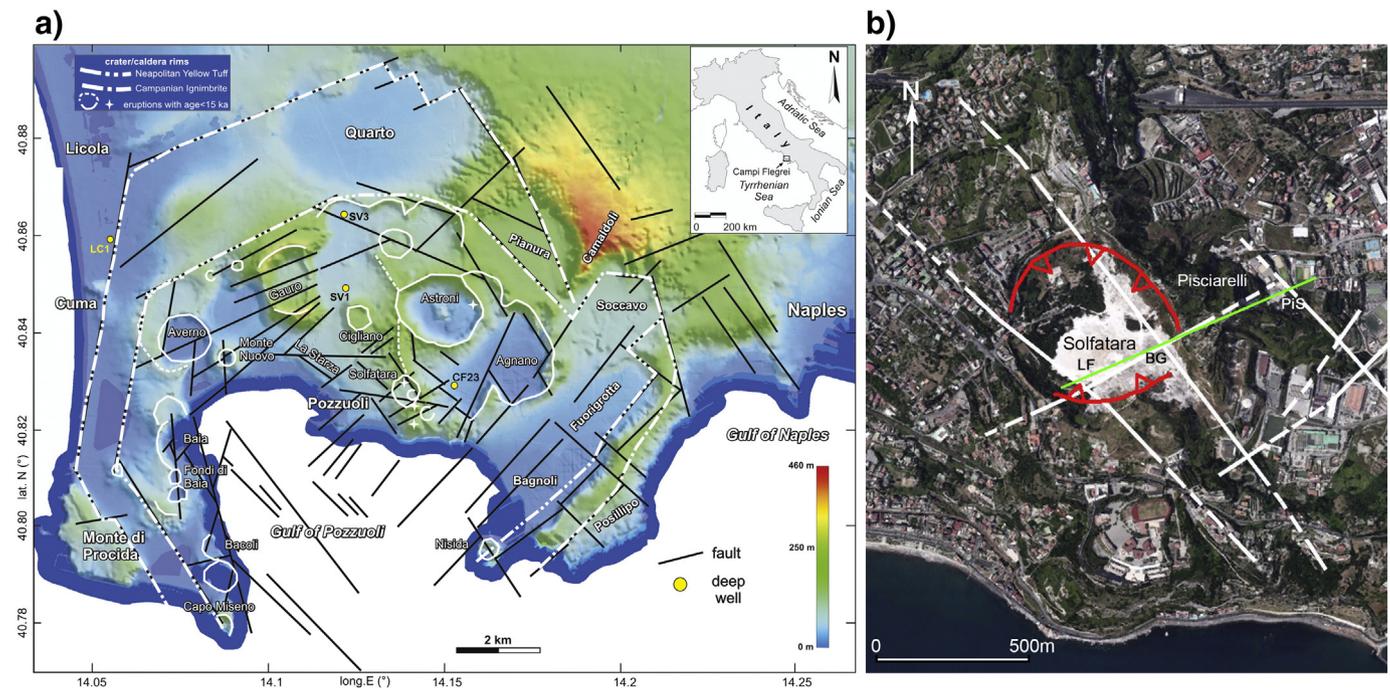
Geophysical methods are worldwide applied to investigate buried structure, furnishing structural models generally originated by single physical parameter (univariate approach). However, in principle, only effective integration from the different geophysical methods (multivariate approach) may provide an unambiguous, self-constrained geological model, suitable for the multiple rocks physical properties (Bosch, 1999; Bosch et al., 2002). The main question is how to correctly handle multiple datasets in the absence of well-defined a-priori relationships between them. Actually, the very high variability range of chemical and physical conditions within the Earth (Lees and VanDecar, 1991) represents the primary limit to find theoretical, or even empirical but universal, relationships between the distinct physical parameters. In this framework, cluster analysis is considered a promising auxiliary tool towards the integration of distinct geophysical methods (Dietrich et al., 1998; Dietrich and Tronicke, 2009; Altdorff and Dietrich, 2010; Song et al., 2010; Di Giuseppe et al., 2014). It can provide an unambiguous linking between well-delineated spatial areas and a compact range of values of the geophysical parameters useful for geological

interpretation. Among the various statistical approaches, the  $k$ -means cluster analysis technique ( $k$ CAT) application to geophysical data results already tested (Audebert et al., 2014; Ward et al., 2014). In particular, it has already been fruitfully used in the Campanian District (Italy) on a joint high-resolution controlled-source audiomagnetotelluric (CSAMT) and seismic profiles across the Irpinia active fault (Di Giuseppe et al., 2014).

In this paper, the post-inversion scheme based on the  $k$ CAT technique is applied to the Solfatara–Pisciarelli area, within the Campi Flegrei volcanic caldera (Naples, Italy; Fig. 1), which represents the most active zone in terms of hydrothermal manifestations and local seismicity in the caldera (e.g., Rosi and Sbrana, 1987; D'Auria et al., 2011; Piochi et al., 2014). The multiple dataset in the S–P area consists of a CSAMT–MT resistivity model (Troiano et al., 2014), a P-wave velocity model (Battaglia et al., 2008) and a density contrast model (Capuano et al., 2013). The obtained image of the volcano-geothermal area shows interesting relations with lithological, petrophysical and structural information from drillhole and field work studies (Rosi and Sbrana, 1987; De Vivo et al., 1989; Piochi et al., 2014; Petrosino et al., 2012). The application of the  $k$ CAT method to the S–P area aims at showing that the clustering approach can be used as an effective method to characterize complex volcano-geothermal structures and to retrieve local relationships between distinct physical parameters.

\* Corresponding author.

E-mail address: [mariagiulia.digiuseppe@ingv.it](mailto:mariagiulia.digiuseppe@ingv.it) (M.G. Di Giuseppe).



**Fig. 1.** a) Aerial view of the Campi Flegrei composite caldera (Naples, Italy). The main faults and the location of the deep boreholes are reported, together with the Neapolitan Yellow Tuff and the Campanian Ignimbrite rims and the main caldera craters. b) Aerial view of the Solfatara crater and adjacent Pisciarelli area. The green line indicates the common geophysical survey line. The white lines indicate the major faults. Red lines indicate the Solfatara crater rims. BG and LF indicate Bocca Grande and La Fangaia (mud pool) main fumarole fields, respectively. PIS indicates the main Pisciarelli vent. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 2. Volcanological outline

### 2.1. The Campi Flegrei caldera

The Campi Flegrei (CF) is a complex caldera with a ca. 2 km average diameter (Fig. 1a), formed during two large volume eruptions, namely the 39 ka BP Campanian Ignimbrite and the 15 ka BP Neapolitan Yellow Tuff (Vitale and Isaia, 2014, and reference therein). The caldera, site of volcanism up to 1538 CE (Rosi and Sbrana, 1987), has always been characterized by intense deformation producing fault systems, a marine terrace and significant changes of the ground level (Parascandola, 1947; Dvorak and Mastrolorenzo, 1991; Rosi and Sbrana, 1987). The most recent episodes of large ground deformation dated at 1970–72 and 1982–84, originated a cumulative maximum uplift of over 3.5 m, accompanied by intense seismicity. In particular, during the 1982–84 crisis > 15,000 earthquakes, mainly occurring in swarms, were recorded, with magnitude up to 4.2 and hypocentres located within the first 4 km of depth mostly around the Solfatara area (De Natale et al., 2006; D'Auria et al., 2011).

Geophysical surveys and drilling explorations have so far allowed a quite good knowledge of the subsurface structure of the caldera (see Rosi and Sbrana, 1987, and reviews in Troiano et al., 2014, and Piochi et al., 2014). The CF deep structure resulting from active and passive seismic, electromagnetic and potential field data is given in Troiano et al. (2014). In synthesis, a coherent source model suggests a quenched magma reservoir (likely a mush) localized below 4–5 km of depth, offshore in the caldera centre, overlying an extended magma sill found at 7–8 km of depth (De Natale et al., 2006; Battaglia et al., 2008). Based on data from deep drillings (Fig. 1a), the structure at depths < 3 km is filled by mostly pyroclastic and volcanoclastic deposits; in the caldera centre the infilling sequence is ca. < 1750 m and sediments were recovered at ca. > 2250 m of depth (Rosi and Sbrana, 1987; Piochi et al., 2014). Furthermore, the buried structure is characterized by a thermo-metamorphic layer at ca. 2500 m of depth, where temperatures reach up to ~350 °C, and three major hydrothermal facies, according to

temperatures and fluid salinity, i.e. propylitic, phyllic and argillic (Rosi and Sbrana, 1987; De Vivo et al., 1989; Piochi et al., 2014; Mormone et al., 2015). Clay minerals in the argillic and phyllic facies are illite, glauconite, chlorite and montmorillonite, pyrite is the main sulphide at depth ca. > 1700 m.

### 2.2. The Solfatara-Pisciarelli hydrothermal area

The Solfatara-Pisciarelli (S-P) area (Fig. 1b) is the most active zone within the CF caldera. It is well-defined by solfataric mineralizations resulting from the intense hydrothermal activity (Rosi and Sbrana, 1987; Piochi et al., 2015).

The Solfatara volcano is a tuff cone formed ca. 4100 years ago, which generated a low-magnitude explosive eruption that ejected tephra over a small area (< 1 km<sup>2</sup>) (Di Vito et al., 1999). Two main fault systems dissect the cone, determining the angular shape of the crater: a NE-SW one and a SE-NW fault system, both associated with the main fumarolic vents. The last one trends offshore where it produced the seafloor deformation (Milia et al., 2000). Hydrothermal activity has long been considered a direct indicator of the present volcanic dynamics. Currently, it exhibits the most impressive degassing manifestations with temperatures up to 150°–160 °C inside the crater. A direct relationship has always been observed between the increase of hydrothermal activity in the Solfatara crater and the ground uplift in the CF caldera (ov.ingv.it). A numerical modelling of the unrests (Troiano et al., 2011, and references therein) highlighted a strong correlation between chemical composition of the S-P fumaroles, seismicity and ground movements, likely resulting from periodic injections of hot CO<sub>2</sub>-rich fluids at the base of a relatively shallow hydrothermal system.

The Pisciarelli area is located outside the Solfatara crater. It extends from the north-eastern slopes of the Solfatara volcano to the western margin of the nearby Agnano plain (Fig. 1a). Near the Solfatara, the Pisciarelli area is characterized by a fumarole field (Fig. 1b), which has manifested an increase of activity marked by a nearly linear trend of the peak temperatures, from about 97 °C up to around 112 °C, from

2003 up to date (Piochi et al., 2015). The increase of activity has also been marked by the opening of new vigorous vents and degassing pools, and by local seismic activity (D'Auria et al., 2011; Piochi et al., 2015).

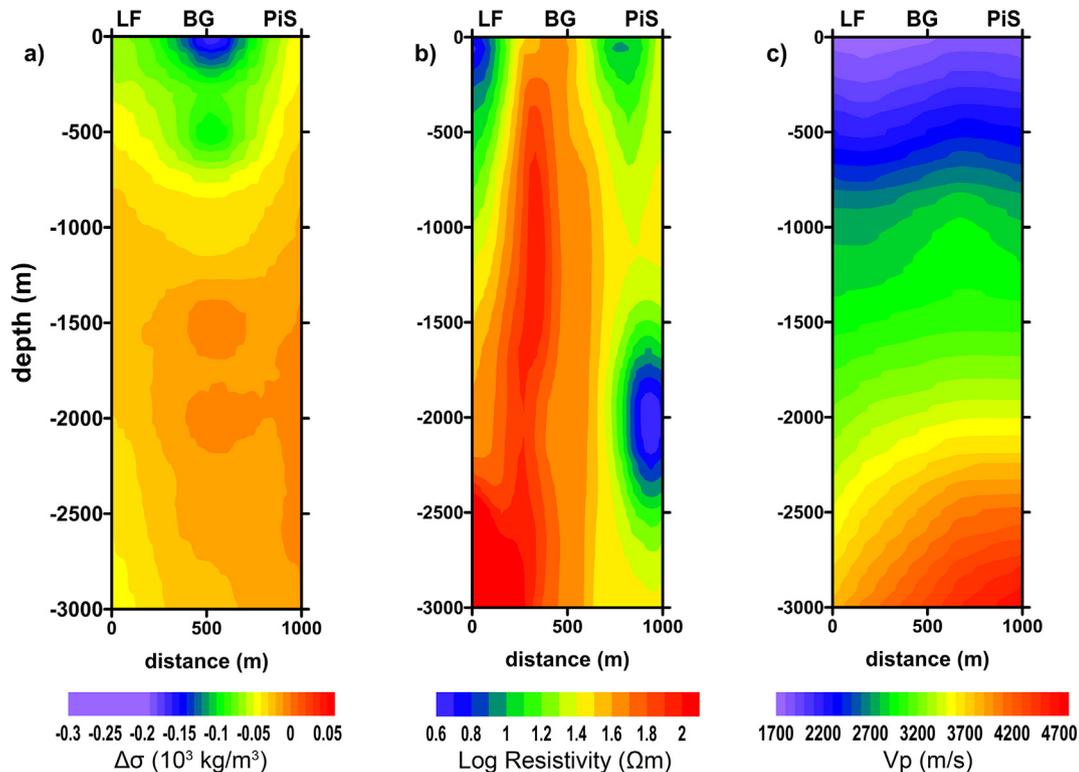
### 3. Geophysical outline of the Solfatara-Pisciarelli hydrothermal area

Geophysical applications have allowed a detailed imaging of the subsurface structure of the Solfatara crater. A high degree of heterogeneities characterizes the first few hundred meters of depth. Starting from the pioneering gravity survey by Oliveri del Castillo et al. (1968), the crater has become the site of several geophysical surveys. Bruno et al. (2007), integrating near-surface TDEM, ERT, and seismic profiles with hydrogeological investigations, characterizing the subsoil beneath the crater, down to about a few ten meters depth. Petrosino et al. (2012) obtained an image of the shallow crustal structure of Solfatara by correlating the local seismic noise with the results of a volcanological and morphostructural survey. Letort et al. (2012) inferred the main subsurface features of the Solfatara crater by correlating active and passive seismic data with CO<sub>2</sub> flux and soil temperature maps. Troiano et al. (2014) obtained a CSAMT-MT resistivity model along a 1.2 km long section across the Solfatara-Pisciarelli area. The modeled section across the profile has essentially been divided in three zones. In such a way a water-saturated, high-pressurized geothermal reservoir under the Solfatara crater, a vertical plume like structure associated with a steam/gas-saturated column, and a hydrothermally mineralized, clay-rich body localized beneath the Pisciarelli area have been identified. Moreover, a few very recent electrical tomography surveys have detailed the near-surface of the Solfatara crater. (Byrdina et al., 2014; Di Giuseppe et al., 2015; Isaia et al., 2015).

#### 3.1. Cluster analysis

The datasets utilized in the *k*CAT application to the S-P geothermal area are the CSAMT-MT resistivity ( $\rho$ ) model (Troiano et al., 2014), the density contrast ( $\Delta\sigma$ ) model by Capuano et al. (2013) and the P-wave velocity ( $V_p$ ) model by Battaglia et al. (2008). Both the density contrast and P-wave velocity models have been extracted from large scale 3D models of the whole CF caldera, whereas the  $\log\rho$  values belong to a more resolved 2D section. The three univariate sections across the green line in Fig. 1b are separately depicted in Fig. 2.

Before illustrating the *k*CAT analysis, a few lines of comment are needed about the nature of the comparison among different geophysical distributions with different resolution on a common grid. In principle, the integration of datasets from different experiments, made at different times and by different groups, is a problematic approach. Nevertheless, owning a multiple dataset in a complex and highly risky volcanic area, is an uncommon, favorable opportunity, deserving more insights. In the present case, the crucial aspect is that the  $\log\rho$ ,  $V_p$  and  $\Delta\sigma$  sections of Fig. 2 result from interpolations onto independent sets of grid points. This result particularly true considering the intrinsic difference in resolution exist between the three considered models. As stated before, the  $V_p$  and  $\Delta\sigma$  sections have been extracted from large-scale 3D models of the whole CF caldera, whereas the  $\log\rho$  values belong to a more resolved 2D section. Furthermore, while a uniformly-spaced grid is commonly used in seismic and gravity inversion, in magnetotellurics a non-uniform model grid, which progressively coarsens with both depth and distance from the measurement sites, is required (e.g. Vozoff, 1987). Therefore, to make the sections as much comparable as possible, we decided to use as common basis for the *k*CAT analysis, the wide-mesh grid of the deeper portion of the CSAMT-MT section. Following Bedrosian et al.'s (2007) general



**Fig. 2.** a) Density contrast ( $\Delta\sigma$ ) model section extracted from the 3D gravity model of the CF caldera (Capuano et al., 2013). b) Resistivity model section obtained from the 3D inversion of the real and imaginary parts of the impedance tensor (Troiano et al., 2014). A common logarithmic scale is used for the resistivity. c) P-wave velocity ( $V_p$ ) model section extracted from the 3D tomographic inversion based on travel time recordings from passive and active seismic experiments over the whole the CF caldera (Battaglia et al., 2008). The three geophysical sections refer to the common line depicted in green in Fig. 1b.

discussion about this problem, in order to test whether the  $V_p$  and  $\Delta\sigma$  estimated 2D distribution and the original 3D distribution were statistically similar (Muñoz et al., 2010), the Q-Q plot method (Isaaks and Srivastava, 1990) was adopted. Fig. 3 shows the result of this statistical check, where the Q-Q plots for the initial 3D dataset (plot a) and the estimated 2D profile (plot b) are drawn. We observe a marked similarity of the two Q-Q plots, which ensures that the interpolation procedure has retained the original statistical distribution, without introducing artifacts.

Coming now to the  $k$ CAT analysis, at first we have integrated the univariate resistivity, velocity and gravity models, by projecting the set of variables  $\mathbf{x}_i = (\log\rho, V_p, \Delta\sigma)_i$  ( $i = 1, 2, \dots, N$ , being  $N$  the total number of mesh points) onto a common joint parameter space (JPS). Technically, after extracting the independent data at the nodes of a regular mesh covering the sections across the common profile, each triplet of values belonging to the same point in the model sections has been projected onto a 3D space graphically represented as in Fig. 4a, i.e. as a triangular system with mutually orthogonal axes  $\Delta\sigma$ ,  $\log\rho$  and  $V_p$ . The JPS scatterplot in Fig. 4a thus provides a first insight into the parameter correlation.

At this point, a new set of variables,  $\mathbf{z}_i$  ( $i = 1, 2, \dots, N$ ), was derived by sphering the data, i.e. by transforming the data into variables with zero mean value (using the euclidean distance) and a covariance matrix given by the identity matrix (Klemelä, 2009). The dataset was

transformed at this stage, because if the standard deviations of the parameters are different, one might dominate in the clustering procedure.

The adopted transformation is expressed by the following relationship:

$$\mathbf{z}_i = \Lambda^{-1/2} \mathbf{Q}^T (\mathbf{x}_i - \bar{\mathbf{x}}), \quad (i = 1, 2, \dots, N), \quad (1)$$

where the columns of  $\mathbf{Q}$  are the eigenvectors that are obtained from the sample covariance matrix,  $\Lambda$  is a diagonal matrix of the corresponding eigenvalues, and  $\bar{\mathbf{x}}$  is the average set of variables.

This homogenizing procedure provides a set of normalized values for each of the mesh grid elements, and consequently for each spatial location. In this way, a new JPS is defined, in which the statistical correlation in the dataset can be fully investigated. To this purpose, as previously anticipated, the cluster analysis technique has been chosen.

Clustering is a process of organizing a set of data into groups via an optimization criterion, in such a way that the observations within any particular group are more similar to each other than they are to the observations belonging to a different cluster (Jain, 2010). A wide range of partitioning methods is available in the literature that are distinct according to the nature of the objective function and the algorithm used to come up with the final clustering (for further details, see Halkidi et al., 2001).

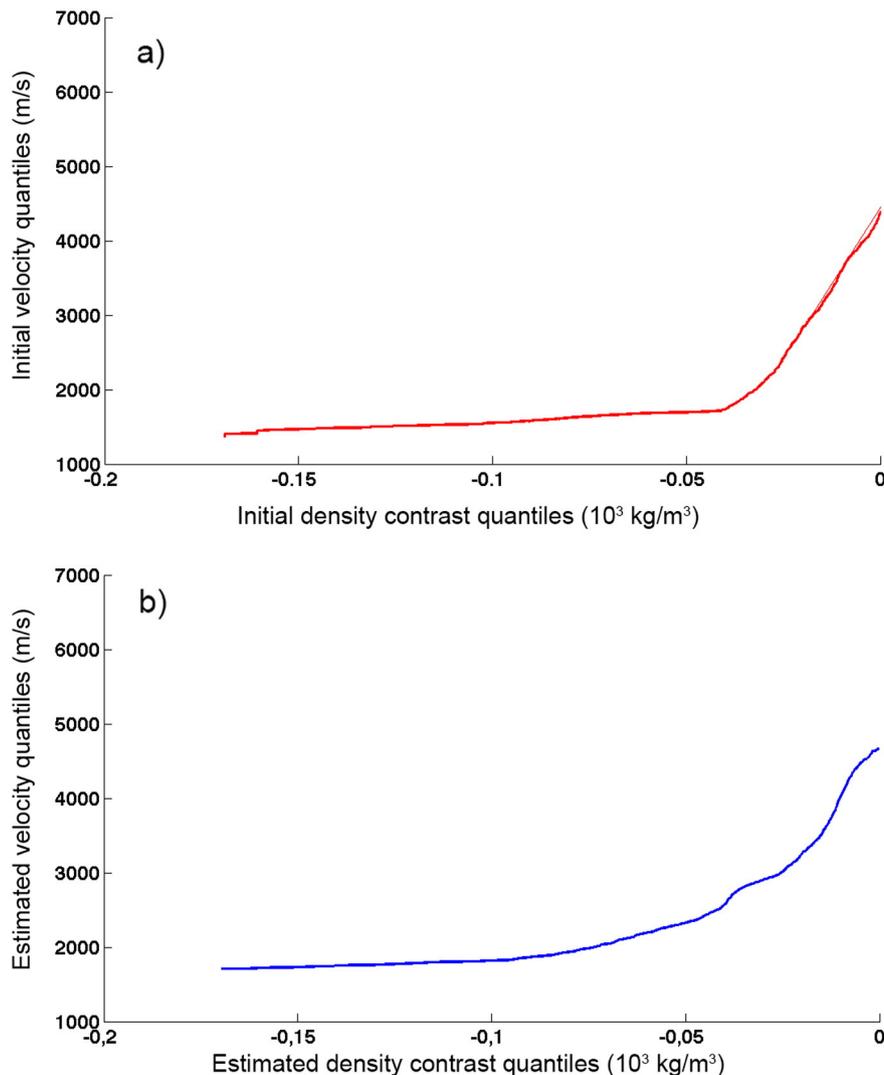
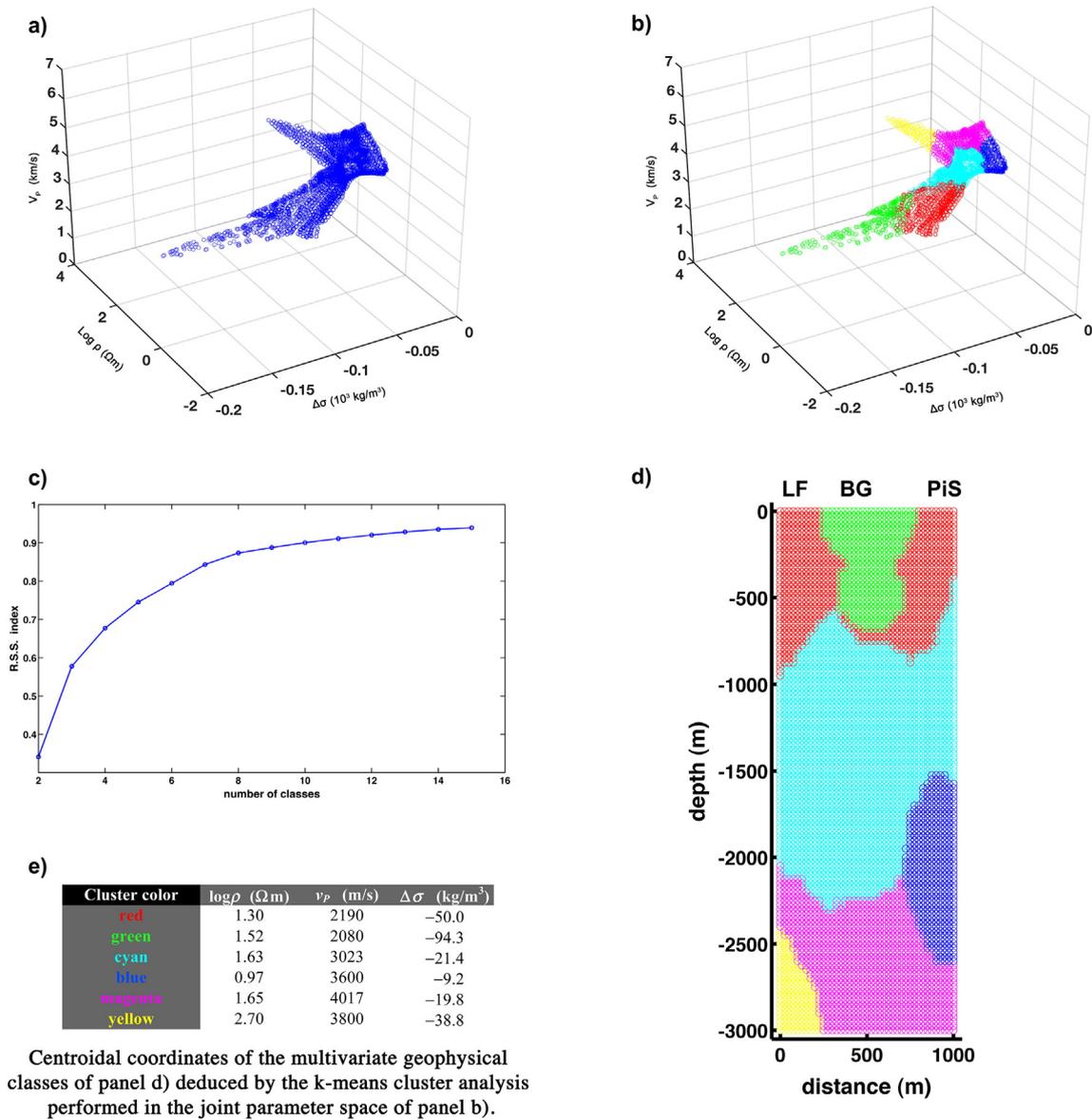


Fig. 3. Quantile-quantile (Q-Q) plots for similarity check between initial and estimated seismic ( $V_p$ ) versus density contrast ( $\Delta\sigma$ ) point distributions. Curve (a) is the Q-Q plot of  $V_p$  versus  $\Delta\sigma$  from the original 3D model distributions. Curve (b) is the Q-Q plot of  $V_p$  versus  $\Delta\sigma$  after interpolation along the 2D section across the common survey line of Fig. 1b.



**Fig. 4.** a) Scatterplot of the univariate  $\log\rho$ ,  $V_p$  and  $\Delta\sigma$  models, depicted in Fig. 2, projected onto the tri-rectangular joint parameter space. b)  $k$ CAT clustered scatterplot of the univariate  $\log\rho$ ,  $V_p$  and  $\Delta\sigma$  models, depicted in Fig. 2, projected onto the tri-rectangular joint parameter space. c)  $RS$  validation index with changing number of clusters,  $k$ , for the analysis of the resistivity, velocity and density models through the  $k$ CAT algorithm, depicted in Fig. 2. d)  $k$ CAT clustered trivariate model section across the common survey line of Fig. 1b, obtained by combining the univariate  $\log\rho$ ,  $V_p$  and  $\Delta\sigma$  models, depicted in Fig. 2. e) Centroidal coordinates of the trivariate geophysical classes of Fig. 4d, deduced by the  $k$ CAT analysis performed in the joint parameter space of Fig. 4b.

We have adopted the  $k$ -means cluster analysis technique ( $k$ CAT) that was independently introduced into various scientific fields by Steinhaus (1956), Ball and Hall (1965), MacQueen (1967), and Lloyd (1982). It is one of the most widely used algorithms for the classification of a given dataset through several clusters, as an a-priori fixed number,  $k$ . The motivations that lie behind the choice of the  $k$ CAT, altogether with the related issues will be discussed in the following. The main idea of this technique consists in the definition of  $k$  centroids, or means points, one for each cluster  $S_i$ , by minimizing a norm of the kind

$$V = \sum_{i=1}^k \sum_{z_j \in S_i} (z_j - \bar{z}_i)^2, \quad (2)$$

where  $\bar{z}_i$  are the mean points of all  $z_j \in S_i$ . Initially, the centroids are placed randomly. Each point belonging to the dataset is ascribed to the class linked to the closest centroid, thus obtaining the first data partition. Subsequently, the barycentre of each cluster is recalculated by

adopting a suitable norm (in our case, a simple Euclidean distance). Using this upgraded set of centroids, further binding is made to generate the new classes. By iterating the procedure, a loop is generated with the  $k$  centroids that change their location at each step. Accordingly, the clusters change their composition, until convergence is reached when the classes do not vary any more.

One of the issues that must be addressed when using the  $k$ CAT is the determination of the optimal cluster number  $k$ . Unfortunately, there are no theoretical solutions to define  $k$  beforehand, and its estimation is a key question that arises during data mining. Usually, clustering algorithms are run with different values of  $k$ . The best value of  $k$  is then chosen based on a predefined criterion, aimed to obtain a compromise between good data clustering and acceptable physical interpretability. The procedure to estimate  $k$  using an external criterion is known as cluster validity. On this basis, a number of validity indices have been defined and proposed in the literature, including the Bayes information criterion and the Akaike information criterion (Xie and Beni, 1991; Sharma,

1996; Rezaee et al., 1998; Theodoridis and Koutroubas, 1999; Halkidi et al., 2000). In spite of the number of different objective criteria applied, it still remains difficult to decide which value of  $k$  leads to meaningful clusters. In the evaluation and selection of an optimal clustering scheme, for our application we stress two of the available criteria (Berry and Linoff, 1996), namely ‘compactness’ and ‘separation’. Compactness means that the members of each cluster should be as close to each other as possible. A common measure of compactness is the variance, which should be minimized. Separation means that the clusters themselves should be widely spaced.

As validation index, in the present study we have used the  $R$ -squared ( $RS$ ) index (Halkidi et al., 2000; Sharma, 1996), which is defined as

$$RS = \frac{SS_t - SS_w}{SS_t} = \frac{SS_b}{SS_t}, \quad (3)$$

where  $SS_t$  is the total sum-of-squares distance between the points belonging to the whole dataset,  $SS_w$  is the sum-of-squares distance within each single group, and  $SS_b$  is the sum-of-squares distance between points belonging to the distinct groups.

The  $RS$  index value of a given cluster set is the ratio between  $SS_b$  and  $SS_t$ . As can be seen,  $SS_b$  is a measure of the difference between the groups. Since  $SS_t = SS_b + SS_w$ , the greater is  $SS_b$ , the smaller is  $SS_w$ , and vice versa. As a result, the greater the differences between the groups are, the more homogeneous each group is, and vice versa. Thus,  $RS$  can be considered as a measure of the degree of difference between the clusters. Furthermore,  $RS$  also measures the degree of homogeneity in each group. The values of  $RS$  vary between 0 and 1. If  $RS$  is zero, then there are no differences among the groups. On the other side, if  $RS$  is 1, this is an indication of significant differences among the groups. The optimum number of clusters is reached when the addition of another cluster does not substantially modify the information in the dataset. In our case, we iterated the clustering algorithm for  $k$  varying from 2 to 15 classes, and we obtained a clear bend in the  $RS$  curve. The crucial area is centered at  $k = 6$ , as shown in Fig. 4c. This is thus a suitable constraint for the clustering class number selection.

The resulting data clustering is shown in Fig. 4b for the non-normalized dataset. As each point in the JPS is related to a point in the physical space, every class can be mapped back from the JPS to the spatial domain. In this way, a whole joint depth section is obtained across the selected profile of Fig. 1b, as shown in Fig. 4d. Every single class is characterized by a set  $(\log\rho, V_p, \Delta\sigma)$ , which is defined by the coordinates of the relative centroid listed in Fig. 4e.

Although, as previously stated, there is no definitive criterion available to estimate the best number  $k$  of distinct classes, the result of the application of the  $k$ CAT algorithm to the S-P dataset, depicted in Fig. 4, can be considered an acceptable partitioning of the relative JPS. To this purpose, we note that the resulting classes are constrained in shape and are separated in the JPS representation of Fig. 4b, which is coherent with the  $RS$  index convergence towards 1, already at low class numbers, as shown in Fig. 4c. Moreover, and more significantly, the partition into distinct zones retains its compactness and sharpness even when the JPS classes are mapped back in the physical space, to compose the joint section shown in Fig. 4d.

#### 4. Discussion

##### 4.1. Significance of the $k$ CAT analysis

An interesting aspect of the  $k$ CAT arises when the existence of relationships between the physical parameters can directly be inferred from the scatterplot. Such a possibility was investigated by Di Giuseppe et al. (2014) in the study of a simple  $\log\rho$ – $V_p$  bivariate coupling. We extend now this approach to the present  $\log\rho$ – $V_p$ – $\Delta\sigma$  trivariate coupling.

The scatterplot in Fig. 4a and b shows a peculiar distribution of the clusters in two distinct branches, in each of which the three parameters involved appear to be linearly related to each other. The first branch includes the red, green and cyan clusters, and the second one the yellow, magenta and blue clusters. Considering the indications of this visual inspection, a more quantitative way to substantiate such subdivision has to be looked for. To such aim, a Gaussian mixture approach can be introduced, which approximate the probability density functions (pdf) of our trivariate variables to the sum of an assigned number of trivariate Gaussian (Bedrosian et al., 2007; Ward et al., 2014). Once identified the Gaussian best-fit combination for our dataset, characterized by specific mean values and variances, linking each point of the parameter space to the Gaussian of nearest mean value, a classification of the data has been obtained, summarized in Fig. 5. In such way the two branches previously identified through a non-quantitative visual inspection have been defined in a more rigorous manner. For each one of the two branches, a least-squares trivariate linear regression have been subsequently performed, thus obtaining the result shown in Fig. 6, where the two distinct interpolating planes are represented. The goodness of fit parameters  $R^2$  values indicate that the 80% and the 62% of the variance is taken into account for the two branches, respectively.

The equation of the plane reported in Fig. 6 permits 6 partial derivatives, which occupy off-diagonal positions in the following matrix

$$\begin{bmatrix} 1 & (\partial \log\rho / \partial V_p)_{\Delta\sigma} & (\partial \log\rho / \partial \Delta\sigma)_{V_p} \\ (\partial V_p / \partial \log\rho)_{\Delta\sigma} & 1 & (\partial V_p / \partial \Delta\sigma)_{\log\rho} \\ (\partial \Delta\sigma / \partial \log\rho)_{V_p} & (\partial \Delta\sigma / \partial V_p)_{\log\rho} & 1 \end{bmatrix} \quad (4)$$

In the matrix above, the subscripts denote the variable that is held fixed. It is easy to prove that the mutual relations between each pair of variables are fully reversible, hence each term below the leading diagonal is the reciprocal of the corresponding term above that diagonal. In this sense, the matrix is phenomenologically symmetric. Each point of the scatterplot diagram in Fig. 4a can be thus considered as a three-phase elementary cell in the JPS, namely the  $\Delta\sigma$ -constant phase, the  $V_p$ -constant phase and the  $\log\rho$ -constant phase.

Inserting the numerical values of the coupling coefficients given in Fig. 6, the explicit expressions of the partial derivatives are obtained for the cluster branches **g1** (red dots in Figs. 5 and 6) and **g2** (blue

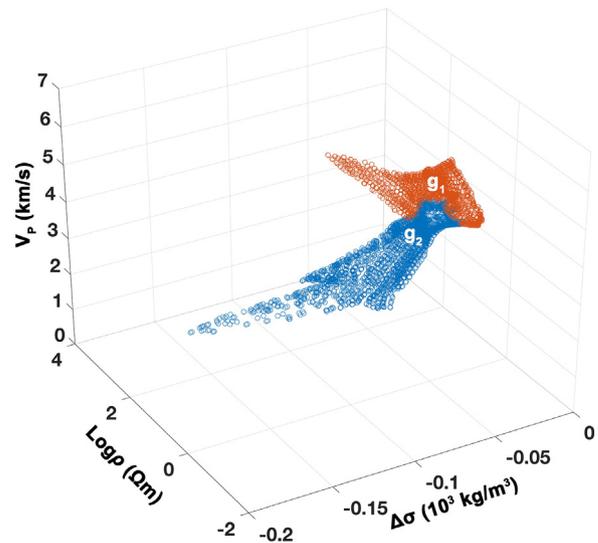
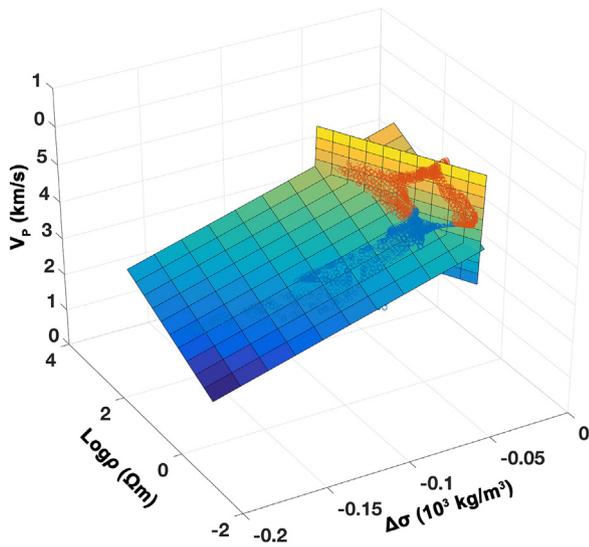


Fig. 5. Gaussian mixture clustered scatterplot of the univariate  $\log\rho$ ,  $V_p$  and  $\Delta\sigma$  models, depicted in Fig. 2, projected onto the tri-rectangular joint parameter space. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)



**Fig. 6.** Representation of the planes interpolating the Gaussian Mixture cluster groups **g1** (brown cluster) and **g2** (blue cluster) of Fig. 5 in the joint parameter space. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

dots in Figs. 5 and 6) as reported in Table 1. Due to the reciprocity property, we report only the terms above the leading diagonal of the matrix in Eq. (4).

In the following, we investigate the significance of the partial derivatives, taking in due consideration the volcano-geothermal context with which the original dataset is related. We assume the convention that the numerator is the effect, caused by the denominator.

4.1.1.  $(\partial \log \rho / \partial V_p)_{\Delta \sigma} > 0$

We refer to the  $\sigma$ -constant phase. Since it is  $V_p = [(\lambda + 2\mu)/\sigma]^{1/2}$ , with  $\lambda$  and  $\mu$  the Lamé parameters, a variation of  $V_p$  is the result of a variation of the elastic properties of the medium. At constant  $\sigma$ ,  $V_p$  in rocks has been shown to decrease with increasing effective pressure, defined as overburden pressure minus pore pressure (Brace, 1972). Moreover, for rocks where conduction is primarily through minerals, at low effective pressure,  $\rho$  decreases with increasing effective pressure (Brace and Orange, 1968). While, where conduction is primarily through a continuous network of electrolyte-filled cavities (pores, cracks, fractures), only rocks partially saturated with electrolytes become less resistive as the effective pressure increases (Brace and Orange, 1968). In both cases, the sign of the changes in  $V_p$  and  $\rho$  is the same, hence the corresponding partial derivative is always positive.

In the S-P area, both types of conduction have been shown to occur along the section. Recalling Troiano et al. (2014), the zone below the Solfatarata mud pool, was ascribed to a water-saturated, high-pressure geothermal reservoir. In the zone below Piscarelli, metal-like conduction was, instead, supposed to be dominant, likely due to widespread (up to 5–10%) pyrite, as observed in deep drillings at

comparable depths (Chelini and Sbrana, 1987; Mormone et al., 2011). The high conductivity in this zone can also be determined by the presence of highly saline fluids at the temperature of 300 °C detected at –1840 m of depth at the CF23 deep hole (Fig. 1a), drilled near the NE extremity of the profile under analysis (green line in Fig. 1b) (Carlino et al., 2012; Piochi et al., 2014).

It is worth mentioning that the partial derivative in question is in **g1** more than one order of magnitude greater than in **g2** (see 3rd column in Table 1). We remark that in **g2**, the condition of low effective pressure can still be admitted, because the increase of the confining pressure is compensated by an increase of the pore pressure with depth, due to the less space available for cavities. On its turn, the tendency to pore/crack closure with depth reduces the continuity of the ionic paths. Therefore in **g2** the result can be a lower decrease of  $\rho$  with decreasing  $V_p$ .

4.1.2.  $(\partial \log \rho / \partial \Delta \sigma)_{V_p} < 0$

We refer now to the  $V_p$ -constant phase, which requires a constant  $(\lambda + 2\mu)/\sigma$  ratio. Under such assumption, variations of  $\sigma$  can likely be due to varying weight percentages of neogenic minerals in the rock cavities due to hydrothermal alteration. Resistivity is thus expected to be directly related to density, moving downward through the various geothermal facies, unless we are in presence of heavy metallic minerals deposited along the fluid pathways, in which case resistivity decreases with increasing density. In the CF area, this last case is supported by cored rocks that, as already stated, evidence widespread pyrite, whose conductivity (e.g. Pridmore and Shuey, 1976) is some orders of magnitude higher than that of the other minerals filling the fractures (e.g. Mormone et al., 2015).

We observe in Table 1 that in **g1** such variation is lower compared to **g2**, where a slightly greater amount of metallic mineral particles should therefore be found.

4.1.3.  $(\partial V_p / \partial \Delta \sigma)_{\log \rho} > 0$

We refer at last to the  $\rho$ -constant phase. Such a phase can be allowed when the rock cavities contain volatiles (steam/gas) and non-metallic mineral particles. An increase of the non-metallic mineral deposition, at the expense of the volatile component, causes an increase of both density and velocity. Conversely, an increase of the volatile component in the cavities, at the expense of the non-metallic solid particles, generates a decrease of both density and velocity. Under these circumstances, the above inequality of the partial derivative is always satisfied.

We observe in Table 1 that in **g1** such variation is two order of magnitude lower compared to **g2**, where the cavities should thus contain a greater amount of volatiles.

A further important aspect that characterizes this last phase in both cluster groups, is that  $V_p$  and  $\Delta \sigma$  are linearly related to each other, which implies that  $(\lambda + 2\mu) \propto \sigma^3$ .

4.2. A few more considerations about kCAT

As stated before, the kCAT represents a well consolidate technique, already applied in geophysics. However, a few cautions should be adopted every time a statistical approach is considered. In particular the reasoning behind selection of the clustering method should be clearly outlined. In our case, high density regions of points appear to follow the boundaries of the scatterplot. Being our kCAT implementation constructed on a Euclidean-based norm, issues arise when the data distribution is anisotropic, which is clear in the results shown. So kCAT could not be totally appropriate for the data itself and some alternative density-driven clustering may lead to more statistically viable results. In order to check such question, the commonly-used mean shift clustering (Comaniciu and Meer, 2002), which assigns points to a cluster based the gradient density of the dataset has been performed and the corresponding results are reported in Fig. 7. As first consideration, the number of clusters need not be predetermined in the mean shift framework,

**Table 1**  
Numerical values of the partial derivatives of the variables in the equation of the plane interpolating the cluster branches **g1** and **g2**. The third column (**g1/g2**) reports the values of the ratios of the homologous partial derivatives.

<b>g1</b>	<b>g2</b>	<b>g1/g2</b>
$\frac{\partial \log \rho}{\partial V_p}_{\Delta \sigma} = 1.9 \cdot 10^{-3} [\Omega s]$	$\frac{\partial \log \rho}{\partial V_p}_{\Delta \sigma} = 6.3 \cdot 10^{-5} [\Omega s]$	30.16
$\frac{\partial \log \rho}{\partial \Delta \sigma}_{V_p} = -22.94 [\Omega \text{kg}^{-1} \text{m}^4]$	$\frac{\partial \log \rho}{\partial \Delta \sigma}_{V_p} = -59.07 [\Omega \text{kg}^{-1} \text{m}^4]$	0.39
$\frac{\partial V_p}{\partial \Delta \sigma}_{\log \rho} = 1.2 \cdot 10^4 [\text{kg}^{-1} \text{m}^4 \text{s}^{-1}]$	$\frac{\partial V_p}{\partial \Delta \sigma}_{\log \rho} = 9.3 \cdot 10^5 [\text{kg}^{-1} \text{m}^4 \text{s}^{-1}]$	0.013

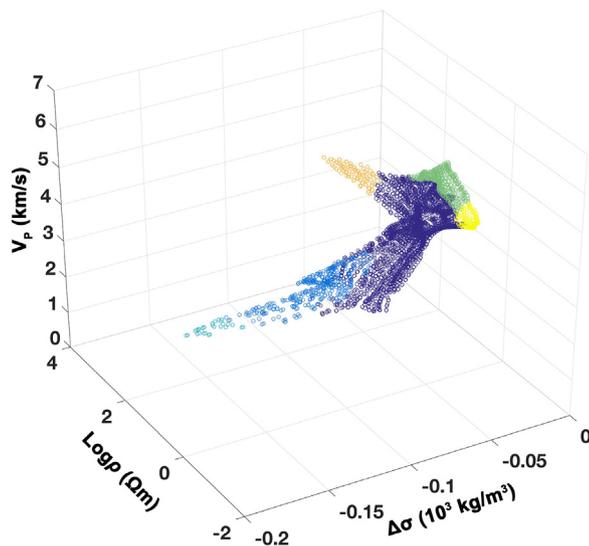


Fig. 7. Mean Shift clustered scatterplot of the univariate  $\log\rho$ ,  $V_p$  and  $\Delta\sigma$  models, depicted in Fig. 2, projected onto the tri-rectangular joint parameter space.

once a bandwidth parameter  $bw$  has been settled. In our case a choice of  $bw = 1$  has been imposed, due to the fact that data have been normalized following Eq. (1) e.g. have zero mean and unitary covariance. In such a way the best class number results of six, similarly to what has been found analyzing the knee-curve of the RS index performed on  $k$ CAT clustering, shown in Fig. 4c. The data partitioning itself retain a rather similar structures, apart for some minor differences. As a conclusion, the result of  $k$ CAT seems not to be influenced from the anisotropic distribution of the data in a significant way.

A further comment concerns the meaning that can be attributed to the various clustering techniques, as  $k$ CAT or mean shift, when treating geophysical dataset. Essentially two approaches can be retrieved in Literature. Clustering has been considered as a simplification tool towards a full joint inversion of the data (Paasche and Troncke, 2007; Moorkamp et al., 2013; Carter-McAuslan et al., 2015; Sun and Li, 2016), which still remains an extremely complex task not only for computational reason but firstly for the lacking of relationships linking the various physical parameter of the rocks. Moreover, it has been considered as a tool to classify in a schematic manner the various lithotypes present in each single univariate section (Bedrosian et al., 2007; Ward et al., 2014). Considering also the findings of Di Giuseppe et al. (2014), a significant amount of original information could arise from the clustering analysis of multivariate geophysical dataset, which could be totally hidden when the single univariate sections are considered. In other terms, the clustering sections combine the data in a way that represents more than a simple superimposition of the univariate sections, being not a true image of the subsoil but a combination of the structures in a joint framework. In the present case, mapping back in the physical space the two groups  $g_1$  and  $g_2$  evidenced in Fig. 5, the resulting model section indicates a sequence of two blocks,  $L_1$  and  $L_2$ , as reported in Fig. 8. The  $L_1$ - $L_2$  interface can be retrieved also considering the means shift technique described before, without any particularly change due to the different statistical approach. Even considering the Gaussian mixture related clustering, the interface still remains substantially unchanged. It is extremely worth to note that one could be lead to think that a similar effects could be maintained also in the case of  $k = 2$  manual selection for the number of classes in the  $k$ CAT, but this is not the case. In our opinion this represents a particularly significant results that evidences how the clustering represents a supplementary source of information. A physical key of reading of such statement lay in the fact that the classes identified through the  $k$ CAT are associated to some lithological or physical properties distinguishing the subsoil rocks. This result presents a direct volcanological significance, as will

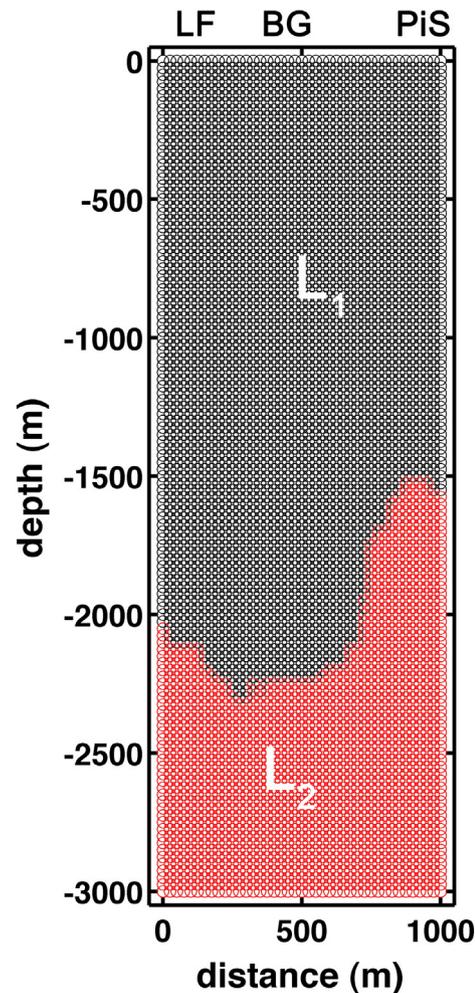


Fig. 8. a) Trivariate model section across the common survey line of Fig. 1b, obtained mapping back into the physical space the two groups evidenced in Fig. 5.  $L_1$  is the top layer corresponding with the cluster group  $g_1$ ;  $L_2$  is the bottom layer, corresponding with the cluster group  $g_2$ .

be discussed in the next section. This rationale results strictly linked to the different geophysical behavior of the distinct properties of rocks. There is, in fact, a wide range of structures presenting different signature when interrogated by different geophysical methodologies. As example, it is possible to consider the literature concerning the saline domes geophysical imaging, a kind of structure presenting, as known, a different electromagnetic and seismic behavior (Key et al., 2006). In the section of Fig. 4d some of the univariate features apparently disappear. This is the case of the interface between the Solfatarata and Pisciarelli identified by the CSAMT-MT survey (Fig. 2b), the density change between 700 and 1300 m (Fig. 2a) or the drastic change of  $V_p$  after 2000 m depth (Fig. 2c). On other hand, the two blocks sequence of Fig. 8 simplifies the structures even more. However, the sections of Figs. 4d and 8 group together very diversified structures and this grouping leak on a totally quantitative statistical basis founded on some common properties of rocks, which physical and lithological meaning has to be guessed. In the understanding of the nature of this element lay the gain of knowledge that can be obtained through the clustering approach.

#### 4.3. Volcanological insight

As stated in the end of the previous section, our analysis allows to map the two groups  $g_1$  and  $g_2$ , evidenced in Fig. 5, back in the physical space, as shown in Fig. 8. The resulting model section indicates a sequence of two blocks ( $L_1$  and  $L_2$ ). The  $L_1$ - $L_2$  interface, located at 2 km

mean depth b.g.l. beneath the Solfatara crater, becomes gently shallower to the SW, and abruptly rises up to ca. 1500 m depth b.g.l. to the NE, i.e. beneath the Pisciarelli area.

The **L1-L2** interface may represent either (1) the limit of the marine sediments, underlain by the pyroclastic sequence, or (2) the transition towards deep geothermal mineralization facies, or (3) the brittle-ductile transition. To help assess the consistency of the three hypotheses, it is possible to consider the main volcanological features of the area, derived also from the nearest deep wells CF23 and SV1 (Fig. 1a), located at about 3.5 km NW from the S-P area.

The possibility that the **L1-L2** interface may represent a trace of the floor of the CF caldera is, indeed, poorly constrained. On the other hand, it must be stressed the **L1-L2** interface is the result of a geophysical clustering process, which groups very diversified geovolcanologic structures, strongly altered given the particular physical conditions of the area. Therefore, it seems difficult to imagine, at least in principle, that, in the examined area, the interface in question may be described as a very specific marker of a homogeneous geological discontinuity.

As it concerns the second hypothesis, as previously stated, the resistivity drop with increasing density within the upraised portion of the **L2** block beneath the Pisciarelli area can be justified by the presence of widespread conductive pyrite (Piochi et al., 2015). Thereby, the **L1-L2** transition could represent the increasing of geothermal alteration with depth from the argillic to the propylitic facies.

About the third hypothesis, we remember that the transition from brittle to ductile rock behavior is a substantial change in rock rheology in volcano-hydrothermal systems (Fournier, 1999; Rye, 2005; Hurwitz et al., 2007). It is normally placed where the temperature is at 350–450 °C (Hurwitz et al., 2007). The geothermal gradient measured inside the CF caldera, with value of 150–220 °C km<sup>-1</sup> (Chelini and Sbrana, 1987), indicates that the rheological transition may be located in a wide depth range from 1.6 km down to 3 km b.g.l. At the CF23 bottom hole (ca. 1840 m), the temperature is 300 °C (Carlino et al., 2012; Piochi et al., 2014). Based on field (e.g., Rosi and Sbrana, 1987; Piochi et al., 2015) and geochemical (Piochi et al., 2014, and references therein) evidences, the 300 °C isotherm, generally flat beneath the CF caldera at ca. 2000 m depth, it is likely that may rise up to shallower depths below the active S-P area.

A strong support to the third hypothesis is provided by the analysis of the location of the seismic events recorded during the recent CF caldera unrest (from 2004 to 2014). The majority of the long-period (LP) and volcano-tectonic (VT) seismic events happened in the neighborhood of the S-P complex, with the largest number of hypocentres (≈90%) clustered at 2 km of depth, and ≈50% at 2 km. D'Auria et al. (2011) have interpreted the LP events as the effect of fluid injection and pore pressure diffusion across the transition from a ductile to a brittle crust, and the VT events as the effect of the diffusion of deeper ductile strain (creep) along the shallower brittle crust. In both cases, a transition from a brittle to a ductile zone has been postulated at a depth fully compatible with our findings.

On the basis of such considerations, we feel now we can interpret the non-flat interface discontinuity retrieved by the *k*CAT analysis in the S-P area as a local evidence of the transition from the brittle zone to the underlying ductile zone. Such context can integrate the resistivity decrement determined by the geothermal alteration facies, rich in pyrite. In fact, sulphides-bearing ore mineralizations may develop above the ductile-brittle transition (Rye, 2005).

The **L1-L2** boundary dislocation of up 700 m relates very well with the surface traces of the SE-NW faults system (Rosi and Sbrana, 1987; Di Vito et al., 1999), cutting almost transversally the profile under analysis (see Fig. 1b). This dislocation should have been originated by the recent tectonics that produced the downthrow of the west side, consistently with evidence recovered offshore (Milia et al., 2000). It corresponds with the plume-like body detected across the CSAMT-MT profile (Fig. 2b) (Troiano et al., 2014). The dislocation, therefore, allows the

large hot gas discharges from the brittle zone, characterized by a near-hydrostatic pressure distribution, towards the overlain ductile zone, where there is a progressive decrease in permeability and pressures approach a near-lithostatic distribution (Carlino and Somma, 2010), depending on the strain rate (Turcotte and Schubert, 2014).

From the previous analysis of the physical properties of the **g1** and **g2** cluster branches that characterize the blocks **L1** and **L2**, respectively, it seems plausible to admit that the transition from ductile to brittle would act as a sort of permeability barrier. We are admitting, in fact, that in the ductile, less permeable **L2** layer, fluids, in non-vapour phase, should migrate slowly, and heat should flow mainly by conduction. Conversely, in the brittle **L1** layer, much more permeable due to increase of fracturing, heat transport should occur mainly by advection, and the abrupt pressure fall would produce flashing of the fluid coming from **L2** through the interface, which enriches the vapour phase in the fractured rocks of **L1**. Therefore, the transition from **L2** to **L1** would prevent the liquid phase to be entirely found within the **L1** layer, thus acting like an impenetrable barrier for it. In this context, the dislocation well explains the gas discharge and the alteration features observed in the S-P area.

## 5. Conclusion

The application of the *k*CAT method to the Solfatara-Pisciarelli volcano-geothermal area has shown that the clustering approach can be used as an effective method to characterize buried structures in presence of complex geological settings and to retrieve local relationships between distinct physical parameters. In the joint analysis of a trivariate geophysical dataset, the impact of the *k*CAT analysis should be appreciated also in situations showing negative responses. Indeed, a strongly irregular or even a random redistribution of clusters along the joint section would be indicative of a lack of physical meaning in the parameter correlations, and hence in the reconstructed section.

To enhance the predictive character of the *k*CAT approach, an interesting aspect is the search for relationships between the physical parameters directly inferred from the scatterplot. The trivariate regression tool applied to the resistivity, P-wave velocity and density scatterplot diagram at the Solfatara-Pisciarelli site has shown the possibility to assemble the clusters in two main groups. Projected onto the real space domain, the two cluster groups have been seen assume, within the section through the volcano-geothermal body, a two-layer geometry with the interface at the mean depth of 2000 m b.g.l., dislocated along the SE-NW fault system that interests the north-east crater rim. Each group results characterized by a different linear relationship between the mentioned three geophysical parameters, mathematically represented by the equation of a plane in the joint parameter space.

Moreover, the analysis of the partial derivatives has allowed the structural features to be distinguished between the two layers. The equations deduced from the *k*CAT analysis define the state of the rocks under the selected profile, under a given set of physical conditions. Accordingly, the interface between the two layers has been characterized as roughly marking the passage from a top brittle layer to a ductile substratum and a likely correspondence has also been retrieved with the lithological or geothermal facies transition detected in drilled cores. The equations are thus to be interpreted as mathematical relationships providing the local link between the resistivity, P-wave velocity and mass density. This relationship, reconstructed on a totally quantitative basis, represents also an a-priori knowledge useful to a full joint inversion of the parameters.

## Acknowledgments

The dataset analysed in this paper has been collected in the framework of the INGV-DPC UNREST. Data has been analysed using the commercial code Matlab®.

## References

- Aldorf, D., Dietrich, P., 2010. Cluster analysis of geophysical field data: an approach for reasonable partitioning of sites. Proc. 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1–6 August 2010, Brisbane, Australia ([www.iuss.org](http://www.iuss.org)).
- Audebert, M., Clément, R., Touze-Foltz, N., Günther, T., Moreau, S., Duquennoy, C., 2014. Time-lapse ERT interpretation methodology for leachate injection monitoring based on multiple inversions and a clustering strategy (MICS). *J. Appl. Geophys.* 111, 320–333.
- Ball, G.H., Hall, D.J., 1965. A novel method of data analysis and pattern classification. Technical Report. Stanford Research Institute, Menlo Park, CA ([www.google.it](http://www.google.it)).
- Battaglia, J., Zollo, A., Virieux, J., Dello Jacono, D., 2008. Merging active and passive data sets in travel-time tomography: the case study of Campi Flegrei caldera (southern Italy). *Geophys. Prospect.* 56:555–573. <http://dx.doi.org/10.1111/j.1365-2478.2007.00687.x>.
- Bedrosian, P.A., Maercklin, N., Weckmann, U., Bartov, Y., Ryberg, T., Ritter, O., 2007. Lithology-derived structure classification from the joint interpretation of magnetotelluric and seismic models. *Geophys. J. Int.* 170:737–748. <http://dx.doi.org/10.1111/j.1365-246X.2007.03440.x>.
- Berry, Linoff, 1996. *Data Mining Techniques for Marketing, Sales and Customer Support*. Wiley (ISBN: 978-0-471-17980-1).
- Bosch, M., 1999. Lithologic tomography: from plural geophysical data to lithology estimation. *J. Geophys. Res.* 104:749–766. <http://dx.doi.org/10.1029/1998JB900014>.
- Bosch, M., Zamora, M., Utama, W., 2002. Lithology discrimination from physical rock properties. *Geophysics* 67:573–581. <http://dx.doi.org/10.1190/1.1468618>.
- Brace, W.F., 1972. Pore pressure in geophysics. In: Heard, H.C., Borg, I.Y., Carter, N.L., Raleigh, C.B. (Eds.), *Flow and Fracture of Rocks*. Geophysical Monograph Series vol. 16:pp. 265–273. <http://dx.doi.org/10.1029/GM016>.
- Brace, W.F., Orange, A.S., 1968. Electrical resistivity changes in saturated rocks during fracture and frictional sliding. *J. Geophys. Res.* 73:1433–1445. <http://dx.doi.org/10.1029/JB073i004p01433>.
- Bruno, P.G., Ricciardi, G.P., Petrillo, Z., Di Fiore, V., Troiano, A., Chiodini, G., 2007. Geophysical and hydrogeological experiments from a shallow hydrothermal system at Solfatara volcano, Campi Flegrei, Italy: response to caldera unrest. *J. Geophys. Res.* 112, B06201. <http://dx.doi.org/10.1029/2006JB004383>.
- Byrdina, S., Vandemeulebrouck, J., Cardellini, C., Legaz, A., Camerlynck, C., Chiodini, G., ... Carrier, A., 2014. Relations between electrical resistivity, carbon dioxide flux, and self-potential in the shallow hydrothermal system of Solfatara (Phlegrean fields, Italy). *J. Volcanol. Geotherm. Res.* 283, 172–182.
- Capuano, P., Russo, G., Civetta, L., Orsi, G., D'Antonio, M., Moretti, R., 2013. The active portion of the Campi Flegrei caldera structure imaged by 3-D inversion of gravity data. *Geochem. Geophys. Geosyst.* 14:4681–4697. <http://dx.doi.org/10.1002/ggge.20276>.
- Carlino, S., Somma, R., 2010. Eruptive versus non-eruptive behavior of large calderas: the example of Campi Flegrei caldera (southern Italy). *Bull. Volcanol.* 72:871–886. <http://dx.doi.org/10.1007/s00445-010-0370-y>.
- Carlino, S., Somma, R., Troise, C., De Natale, G., 2012. The geothermal exploration of Campanian volcanoes: historical review and future development. *Renew. Sust. Energ. Rev.* 16:1004–1030. <http://dx.doi.org/10.1016/j.rser.2011.09.023>.
- Carter-McAuslan, A., Lelivre, P.G., Farquharson, C.G., 2015. A study of fuzzy c-means coupling for joint inversion, using seismic tomography and gravity data test scenarios. *Geophysics* 80 (1):W1–W15. <http://dx.doi.org/10.1190/geo2014-0056.1>.
- Chelini, W., Sbrana, A., 1987. *Subsurface geology*. In: Rosi, M., Sbrana, A. (Eds.), *Phlegrean Fields*. Quad. Ric. Sc. 114, CNR, 9:pp. 9–93 ([www.edizioni.cnr.it](http://www.edizioni.cnr.it)).
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (5), 603–619.
- D'Auria, L., Giudicepietro, F., Aquino, I., Borriello, G., Del Gaudio, C., Lo Bascio, D., Martini, M., Ricciardi, G.P., Ricciolino, P., Ricco, C., 2011. Repeated fluid-transfer episodes as a mechanism for the recent dynamics of Campi Flegrei caldera (1989–2010). *J. Geophys. Res.* 116, B04313. <http://dx.doi.org/10.1029/2010JB007837>.
- De Natale, G., Troise, C., Pinguet, F., Mastrolorenzo, G., Pappalardo, L., Battaglia, M., Boschi, E., 2006. The Campi Flegrei caldera: unrest mechanisms and hazards. In: Troise, C., De Natale, G., Kilburn, C.R.J. (Eds.), *Mechanics of Activity and Unrest at Large Calderas*. Geological Society, London, Special Publications Vol. 269:pp. 25–45. <http://dx.doi.org/10.1144/GSL.SP.2006.269.01.03>.
- De Vivo, B., Belkin, H.E., Barbieri, M., Chelini, W., Lattanzi, P., Lima, A., Tolomeo, L., 1989. The Campi Flegrei (Italy) geothermal system: a fluid inclusion study of the Mofete and San Vito fields. *J. Volcanol. Geotherm. Res.* 36:303–326. [http://dx.doi.org/10.1016/0377-0273\(89\)90076-0](http://dx.doi.org/10.1016/0377-0273(89)90076-0).
- Di Giuseppe, M.G., Troiano, A., Troise, C., De Natale, G., 2014. K-means clustering as tool for multivariate geophysical data analysis. An application to shallow fault zone imaging. *J. Appl. Geophys.* 101:108–115. <http://dx.doi.org/10.1016/j.jappgeo.2013.12.004>.
- Di Giuseppe, M.G., Troiano, A., Fedele, A., Caputo, T., Patella, D., Troise, C., De Natale, G., 2015. Electrical resistivity tomography imaging of the near-surface structure of the Solfatara crater, Campi Flegrei (Naples, Italy). *Bull. Volcanol.* 77:27. <http://dx.doi.org/10.1007/s00445-015-0910-6>.
- Di Vito, M.A., Isaia, R., Orsi, G., Southon, J., D'Antonio, M., De Vita, S., Pappalardo, L., Piochi, M., 1999. Volcanism and deformation since 12,000 years at the Campi Flegrei caldera (Italy). *J. Volcanol. Geotherm. Res.* 91:221–246. [http://dx.doi.org/10.1016/S0377-0273\(99\)00037-2](http://dx.doi.org/10.1016/S0377-0273(99)00037-2).
- Dietrich, P., Tronicke, J., 2009. Integrated analysis and interpretation of crosshole P- and S-wave tomograms: a case study. *Near Surf. Geophys.* 7:101–109. <http://dx.doi.org/10.3997/1873-0604.2008041>.
- Dietrich, P., Fechner, T., Whittaker, J., Teutsch, G., 1998. An integrated hydrogeophysical approach to subsurface characterization. In: Herbert, M., Kovar, K. (Eds.), *Groundwater Quality: Remediation and Protection*. IAHS Publication n.250, pp. 513–520 ([iahs/250-0513](http://iahs/250-0513)).
- Dvorak, J.J., Mastrolorenzo, G., 1991. The mechanisms of recent vertical crustal movements in Campi Flegrei caldera, southern Italy. *Geol. Soc. Am. Spec. Pap.* 263:1–47. <http://dx.doi.org/10.1130/SPE263-p1>.
- Fournier, R.O., 1999. Hydrothermal processes related to movement of fluid from plastic into brittle rock in the magmatic-epithermal environment. *Econ. Geol.* 94: 1193–1211. <http://dx.doi.org/10.2113/gsecongeo.94.8.1193>.
- Halkidi, M., Vazirgiannis, M., Batistakis, I., 2000. Quality Scheme Assessment in the Clustering Process. In: Zighed, D.A., Komorowski, H.J., Zytkow, J.M. (Eds.), *Principles of Data Mining and Knowledge Discovery*, Proc. 4th European Conference PKDD. Vol. 1910. Springer Lecture Notes in Computer Science, pp. 265–276 (ISBN: 3-540-41066-X).
- Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. *J. Intell. Inf. Syst.* 17:107–145. <http://dx.doi.org/10.1023/A:1012801612483>.
- Hurwitz, S., Lowenstern, J.B., Heasler, H., 2007. Spatial and temporal geochemical trends in the hydrothermal system of Yellowstone National Park: inferences from river solute fluxes. *J. Volcanol. Geotherm. Res.* 162:149–171. <http://dx.doi.org/10.1016/j.jvolgeores.2007.01.003>.
- Isaaks, E.H., Srivastava, R.M., 1990. *An Introduction to Applied Geostatistics*. Oxford University Press (ISBN: 978-0-19-505013-4).
- 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). *Geol. Soc. Am. Bull.* 127 (9–10), 1485–1504.
- Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recogn. Lett.* 31: 651–666. <http://dx.doi.org/10.1016/j.patrec.2009.09.011>.
- Key, K., Constable, S.C., Weiss, C., 2006. Mapping 3D salt using the 2D marine magnetotelluric method: case study from Gemini Prospect, Gulf of Mexico. *Geophysics* 71, B17–B27.
- Klemelä, J., 2009. *Smoothing of Multivariate Data: Density Estimation and Visualization*. Wiley (ISBN: 978-0-470-29088-0).
- Lees, J.M., VanDecar, J.C., 1991. Seismic tomography constrained by bouguer gravity anomalies: applications in western Washington. *Pure Appl. Geophys.* 135:31–52. <http://dx.doi.org/10.1007/BF00877007>.
- Letort, J., Roux, P., Vandemeulebrouck, J., Coutant, O., Cros, E., Wathelet, M., Cardellini, C., Avino, R., 2012. High-resolution shallow seismic tomography of a hydrothermal area: application to the Solfatara, Pozzuoli. *Geophys. J. Int.* 189:1725–1733. <http://dx.doi.org/10.1111/j.1365-246X.2012.05451.x>.
- Lloyd, S.P., 1982. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* IT-28:129–137 ([www.cs.nyu.edu](http://www.cs.nyu.edu)).
- MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J. (Eds.), *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*. Statistics vol. I. University of California Press, pp. 281–297 (LCCN: 49-8189).
- Milia, A., Torrente, M.M., Giordano, F., 2000. Active deformation and volcanism offshore Campi Flegrei, Italy: new data from high-resolution seismic reflection profiles. *Mar. Geol.* 171:61–73. [http://dx.doi.org/10.1016/S0025-3227\(00\)00111-0](http://dx.doi.org/10.1016/S0025-3227(00)00111-0).
- Moorkamp, M., Roberts, A.W., Jegen, M., Heincke, B., Hobbs, R.W., 2013. Verification of velocity-resistivity relationships derived from structural joint inversion with borehole data. *Geophys. Res. Lett.* 40 (14):3596–3601. <http://dx.doi.org/10.1002/grl.50696> (ISSN 1944-8007).
- Mormone, A., Tramelli, A., Di Vito, M.A., Piochi, M., Troise, C., De Natale, G., 2011. Secondary hydrothermal minerals in buried rocks at the Campi Flegrei caldera: a possible tool to understand the rock-physics and to assess the state of the volcanic system. *Period. Mineral.* 80:385–406. <http://dx.doi.org/10.2451/2011PM0027>.
- Mormone, A., Troise, C., Piochi, M., Balassone, G., Joachimski, M., De Natale, G., 2015. Mineralogical, geochemical and isotopic features of tuffs from the CFDDP drill hole: hydrothermal activity in the eastern side of the Campi Flegrei volcano (southern Italy). *J. Volcanol. Geotherm. Res.* 290:39–52. <http://dx.doi.org/10.1016/j.jvolgeores.2014.12.003>.
- Muñoz, G., Bauer, K., Moeck, I., Schulze, A., Ritter, O., 2010. Exploring the Groß Schönebeck (Germany) geothermal site using a statistical joint interpretation of magnetotelluric and seismic tomography models. *Geothermics* 39:35–45. <http://dx.doi.org/10.1016/j.geothermics.2009.12.004>.
- Oliveri del Castillo, A., Palumbo, A., Percolo, E., 1968. Contributo allo studio della Solfatara di Pozzuoli (Campi Flegrei) mediante osservazione gravimetriche. *Ann. Oss. Vesuv.* 22, 217–225.
- Paasche, H., Tronicke, J., 2007. Cooperative inversion of 2d geophysical data sets: a zonal approach based on fuzzy c-means cluster analysis. *Geophysics* 72 (3), A35–A39.
- Parascandola, A., 1947. *I Fenomeni Bradisismici del Serapeo di Pozzuoli*, Tip. Genovese, Naples (English Translation by M.Capuano and J.J.Dvorak) (ISBN: SBL0731820).
- Petrosino, S., Damiano, N., Cusano, P., Di Vito, M.A., de Vita, S., Del Pezzo, E., 2012. Subsurface structure of the Solfatara volcano (Campi Flegrei caldera, Italy) as deduced from joint seismic-noise array, volcanological and morphostructural analysis. *Geochem. Geophys. Geosyst.* 13. <http://dx.doi.org/10.1029/2011GC004030>.
- Piochi, M., Kilburn, C., Di Vito, M.A., Mormone, A., Tramelli, A., Troise, C., De Natale, G., 2014. The volcanic and geothermally active Campi Flegrei caldera: an integrated multidisciplinary image of its buried structure. *Int. J. Earth Sci.* 103:401–421. <http://dx.doi.org/10.1007/s00531-013-0972-7>.
- Piochi, M., Mormone, A., Balassone, G., Strauss, H., Troise, C., De Natale, G., 2015. Sulfur and S-bearing minerals from the active Campi Flegrei volcano (southern Italy): genetic environments and degassing dynamics revealed by mineralogy and isotope geochemistry. *J. Volcanol. Geotherm. Res.* (submitted for publication).
- Pridmore, D.F., Shuey, R.T., 1976. Electrical-resistivity of galena, pyrite, and chalcopyrite. *Am. Mineral.* 61, 248–259 (doi: collectors\_corner/amtc/toc1976.htm).
- Rezaee, R., Lelieveldt, B.P.F., Reiber, J.H.C., 1998. A new cluster validity index for the fuzzy c-mean. *Pattern Recogn. Lett.* 19:237–246. [http://dx.doi.org/10.1016/S0167-8655\(97\)00168-2](http://dx.doi.org/10.1016/S0167-8655(97)00168-2).

- Rosi, M., Sbrana, A., 1987. Introduction, geological setting of the area, stratigraphy, description of mapped products, petrography, tectonics. In: Rosi, M., Sbrana, A. (Eds.), Phlegrean Fields, Quad. Ric. Sc. 114. Vol. 9. CNR:pp. 9–93 ([www.edizioni.cnr.it](http://www.edizioni.cnr.it)).
- Rye, R.O., 2005. A review of the stable-isotope geochemistry of sulfate minerals in selected igneous environments and related hydrothermal systems. *Chem. Geol.* 215:5–36. <http://dx.doi.org/10.1016/j.chemgeo.2004.06.034>.
- Sharma, S.C., 1996. *Applied Multivariate Techniques*. Wiley (ISBN: 978-0-471-31064-8).
- Song, Y.C., Meng, H.D., O'Grady, M.J., O'Hare, G.M.P., 2010. The application of cluster analysis in geophysical data interpretation. *Comput. Geosci.* 14:263–271. <http://dx.doi.org/10.1007/s10596-009-9150-1>.
- Steinhaus, H., 1956. Sur la division de corps matériels en parties. *Bull. Acad. Pol. Sci.* 4, 801–804.
- Sun, J., Li, Y., 2016. Joint inversion of multiple geophysical data using guided fuzzy c-means clustering. *Geophysics* 81 (3), ID37–ID57.
- Theodoridis, S., Koutroubas, K., 1999. *Pattern Recognition*. Academic Press (ISBN: 978-1-59749-272-0).
- Troiano, A., Di Giuseppe, M.G., Petrillo, Z., Troise, C., De Natale, G., 2011. Ground deformation at calderas driven by fluid injection: modelling unrest episodes at Campi Flegrei (Italy). *Geophys. J. Int.* 187:833–847. <http://dx.doi.org/10.1111/j.1365-246X.2011.05149.x>.
- Troiano, A., Di Giuseppe, M.G., Patella, D., Troise, C., De Natale, G., 2014. Electromagnetic outline of the Solfatara-Pisciarelli hydrothermal system, Campi Flegrei (southern Italy). *J. Volcanol. Geotherm. Res.* 277:9–21. <http://dx.doi.org/10.1016/j.jvolgeores.2014.03.005>.
- Turcotte, D.L., Schubert, G., 2014. *Geodynamics*. Cambridge University Press (ISBN: 9780521186230).
- Vitale, S., Isaia, R., 2014. Fractures and faults in volcanic rocks (Campi Flegrei, southern Italy): insight into volcano-tectonic processes. *Int. J. Earth Sci.* 103: 801–819. <http://dx.doi.org/10.1007/s00531-013-0979-0>.
- Vozoff, K., 1987. The magnetotelluric method. In: Nabighian, M.N. (Ed.), *Electromagnetic Methods in Applied Geophysics: Application*, 2B. SEG, Tulsa <http://dx.doi.org/10.1190/1.9781560802686.ch8>.
- Ward, W.O., Wilkinson, P.B., Chambers, J.E., Oxby, L.S., Bai, L., 2014. Distribution-based fuzzy clustering of electrical resistivity tomography images for interface detection. *Geophys. J. Int.*, ggu006
- Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal.* 13:841–846. <http://dx.doi.org/10.1109/34.85677>.