



# OPEN Principal component analysis on twenty years (2000–2020) of geochemical and geophysical observations at Campi Flegrei active caldera

Zaccaria Petrillo<sup>1</sup>, Simona Tripaldi<sup>2</sup>, Annarita Mangiacapra<sup>1</sup>✉, Sergio Scippacocola<sup>1</sup>, Stefano Caliro<sup>1</sup> & Giovanni Chiodini<sup>3</sup>

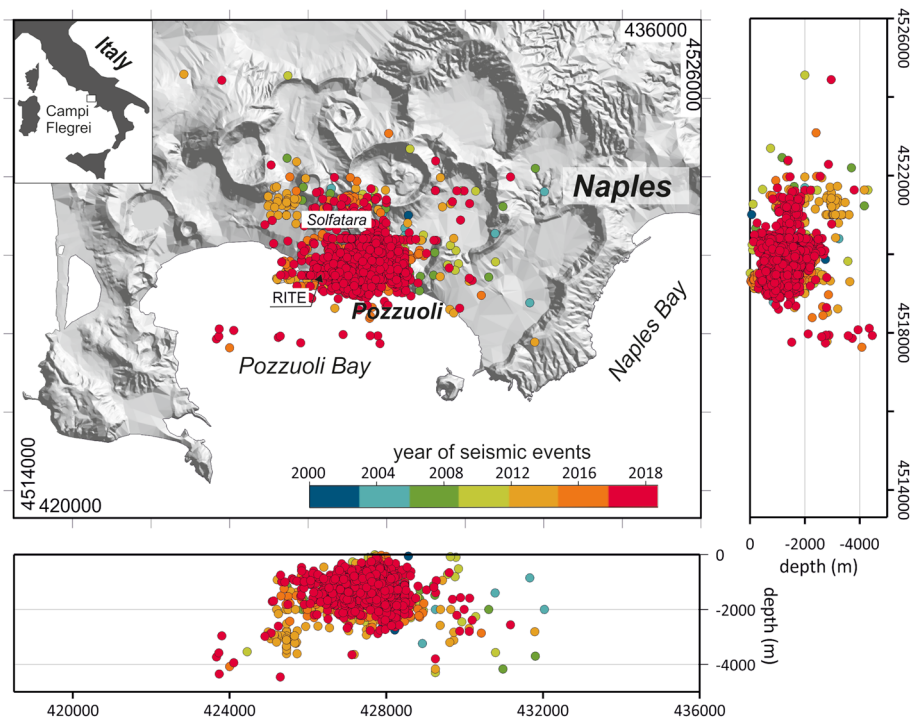
Campi Flegrei (CF) is an active and densely populated caldera in Southern Italy, which has manifested signs of significant unrest in the last 50 years. Due to the high volcanic risk, monitoring networks of the most sensitive unrest indicators have been implemented and improved over time. Precious database constituted by geophysical and geochemical data allowed the study of the caldera unrest phases. In this paper we retrace the caldera history in the time span 2000–2020 by analyzing displacement, seismicity and geochemical time series in a unified framework. To this end, Principal Component Analysis (PCA) was firstly applied only on geochemical data because of their compositional nature. The retrieved first three components were successively analyzed via PCA together with the geophysical and thermodynamical variables. Our results suggest that three independent processes rely on geochemical observations: a heating/pressurizing of the hydrothermal system, a process related to magmatic fluids injection at the hydrothermal system roots, and third process probably connected with a deeper magmatic dynamic. The actual volcano alert state seems mainly linked to the variation of the hydrothermal system activity. Our approach made it possible to explore the interrelation among observations of different nature highlighting the importance of the relative driving processes over time.

In the last decades great attempts have been made to understand the complexity of the volcanic systems structures<sup>1–3</sup> and to define the chemical and physical processes characterizing the associated hydrothermal systems<sup>4,5</sup>, united with their evolution through time. These attempts have been mainly addressed to the mitigation of the volcanic hazard in densely populated areas, where explosive catastrophic eruptions are expected. Campi Flegrei caldera (CFc, Fig. 1) in Southern Italy, represents a suitable example being a particularly dangerous active volcanic site, inhabited by more than 1.5 million people.

During its history, the caldera has experienced two very large explosive events that led to the formation of its primary structure: the Campanian Ignimbrite and the Neapolitan Yellow Tuff eruptions<sup>6–8</sup> then modified by the more recent Agnano Monte Spina and Astroni explosive eruptions, as well as by the numerous successive eruptions, which generated pyroclastic deposits spread over an extremely large area<sup>8,9</sup>. The last eruption in CF occurred in 1538 AD<sup>10</sup>. Currently the caldera magmatic system is still active, as testified by bradyseismic episodes and by a widespread fumarolic and thermal springs<sup>11</sup>, leading this area to a very high volcanic risk.

Over the years the scientific community dealing with CF has mainly focused their attempt on improving the monitoring systems devices<sup>12–20</sup> and developed numerous models to simulate volcanic and hydrothermal processes<sup>4,21–29</sup>. These efforts have been made in order to obtain reliable data and sophisticated simulators able to improve models confidence and predictive capabilities addressed to mitigate the volcanic risk and to prevent and estimate possible catastrophic scenarios.

<sup>1</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Vesuviano, Via Diocleziano 328, Napoli, Italy. <sup>2</sup>Dipartimento di Scienze della Terra e Geoambientali, Università degli Studi di Bari Aldo Moro, via Orabona 4, Bari, Italy. <sup>3</sup>Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Via Donato Creti, 12, 40128 Bologna, Italy. ✉email: annarita.mangiacapra@ingv.it



**Figure 1.** Campi Flegrei caldera map (UTM) showing the geochemical fluids sampling points (BG: zone 33T, Longitude 427649, Latitude 4519918), the ground lifting benchmark (RITE: zone 33T, Longitude 426319, Latitude 4519509), and earthquakes epicentres (1866 events with  $M \geq -0.5$ ) in the period 2000–2020. On the right a depth-NS earthquakes projection and at the bottom the depth-EW earthquakes projection. Colours scale represents the earthquakes time occurrence. The software used to create the figure is Grapher 18.4.420 (<https://www.goldensoftware.com/products/grapher>).

Objectively, there is a certain difficulty to treat and interpret large sets of multiple variables data resulting from long-term monitoring. The basic strategy generally used is the univariate statistical analysis, which can cause, however, uncertainty and error when dealing with huge and inhomogeneous dataset<sup>30</sup>. In order to avoid this problem, multivariate statistical techniques can be used, as they are unbiased methods which can help indicate natural associations between samples and/or variables<sup>31</sup>, thus highlighting information not available at first glance.

In this study, we treat a large dataset of geophysical and geochemical data resulting from long-term (2000–2020) monitoring of the Campi Flegrei caldera. In particular, we examine, vertical displacements, earthquakes occurrences, geochemical composition of fumaroles and the estimated temperature and pressure time series using the Principal Component Analysis (PCA).

PCA is a multivariate statistical procedure widely applied for data processing and dimension reduction, when large multivariate datasets are analyzed. The PCA simplifies the data structure and helps data interpretation<sup>32</sup>. The great advantage<sup>31</sup> lies in giving the ability to detect broader patterns of interrelationships among data than given by individual univariate analyses<sup>30</sup>. The main goal of this study is to identify and distinguish the different processes which in the last twenty years have been the protagonists of the ongoing crisis in the CF caldera and how these processes statistically interact. The geochemical fumaroles fluids samples, being compositional data, require a prior appropriate transformation according to theory of compositional data<sup>33–38</sup>. Via this transformation, it is possible to perform an analysis in Principal components to extract components which contain the synthesis of the geochemical processes acting in the CF caldera. Successively, the integration of the geochemical data with the geophysical data is carried out through a joint principal component analysis from which it is possible to identify the relationships between the physical processes and the geochemical variables synthesis involved in the analysis. The relationships found between geophysical and geochemical processes, allowed us to highlight the main process responsible for the last crisis.

## Materials and methods

### Compositional data analysis

All the statistical measures (e.g. mean, standard deviation, correlation, etc.) are defined in Euclidean space and, therefore, the usual univariate and multivariate analysis can lead to erroneous results when applied to compositional data. The constant sum for each observation limits the Euclidean sampling space to a Simplex, subspace of  $\mathcal{R}^p$ . According to the current literature, only the transformation of compositional data by CoDA methodology<sup>33–38</sup> allows a correct approach to the statistical data analysis. The three main approaches for modeling compositional data analysis are an *additive log-ratio* (*alr*), a *centered log-ratio* (*clr*)<sup>34,39</sup> and the *isometric log-ratio* (*ilr*)<sup>40,41</sup>. The

clr transformation is isometric<sup>39</sup> and allows an interpretation of the relationships one-to-one among the compositional variables being a subspace of  $p$  dimensions<sup>42</sup>. Therefore, in this work, each compositional time series  $X(t)$  will be treated according to the *clr* methodology:

$$\mathbf{Z}(t) = \text{clr}[X(t)] = [\ln(x_{ij}(t)/g_i)] (i = 1, \dots, n; j = 1, \dots, p; t = t_0, \dots, T) \quad (1)$$

where  $n$  is the number of observations,  $p$  the number of variables and  $g_i$  is the geometric mean of each compositional data at  $t$  time ( $t \in T$ ).

### Principal component analysis and biplot

PCA linearly transforms the original variables into a smaller new set of uncorrelated variables<sup>43</sup> each new variable is a linear combination of the old. The Principal components (PCs) are ordered so that the first few retain most of the information present in all of the original dataset. PCA can be implemented in two modalities: linear or via Decomposition into Singular Values (SVD)<sup>44,45</sup>. The first principal component is a linear combination of all variables with the greatest variability, the second principal component represents the greatest variability after the effect of the first has been removed, and so on. Since only the first components explain a significant part of the total variance, the remaining PCs can also be ignored. SVD is the most used method especially when one is interested in representing results with few dimensions. In this work we use SVD, following Aitchison's suggestion, since the compositional data have an adequate representation in the biplot. The Decomposition into Singular Values (SVD) of the  $X$  data matrix, centered or standardized is  $X = U\Lambda V'$ , where  $U(n, p)$  are left eigen-vectors of  $XX'$ ,  $\Lambda(p, p)$  are the eigen-values, and  $V'(p, p)$  are the right eigen-vectors of  $X'X$  ( $U$  and  $V$  are orthogonal). Splitting  $\Lambda$  into  $\Lambda^\alpha \Lambda^{\alpha-1}$ , where  $(0 \leq \alpha \leq 1)$ . The SVD of matrix  $X$  becomes:

$$X = (U\Lambda^\alpha)(\Lambda^{1-\alpha}V') = GH' \quad (2)$$

where  $G = U\Lambda^\alpha$  and  $H = \Lambda^{1-\alpha}V'$ .

Biplot<sup>34,46</sup> is a graphical tool that provides, in the factorial plane, the results of analysis in Principal Components (PCA) of a matrix  $X(n, p)$ . The *bi* prefix indicates that the plot contains information on the  $n$  observations and on the  $p$  variables. In a biplot, the  $n$  rows of the matrix  $G(n, 2)$  are represented as points-units (scores), corresponding to the observations, and the  $p$  rows of the matrix  $H(p, 2)$  are represented as vectors (rays), corresponding to the variables (loadings). The length of each vector starting from the origin of the axes, approximates the variance of the respective variable; the angle between two rays (cosine of the angles between the rays) approximates the correlation between the variables they represent; the projection of a generic data on a specific vector approximates the value of that data with respect to the variable represented by the vector. The  $\alpha$  value in (2) can range from 0 to 1. In this work, we are interested to the covariance-biplot ( $\alpha=0$ ) that preserves the covariance structure<sup>47</sup> and privileges the display of the variables.

In the case of compositional data matrix, SVD can be applied after a *clr*, *alr* or *ilr* transformation<sup>34</sup>. After this application, the approximation structure of the correlations among data can be deduced. The biplot of compositional data (relative variance biplot), shows the rays (projection of the original variables onto this orthogonal space) and the links between two rays (the ratios among geochemical species). The links, that is the approximation of the geochemical ratio by the new subspace, are compositionally invariant, this mean they are good physical observables. If two links are orthogonal, an independent relationship between the two sub-compositions (equivalently the two sub-compositional ratios) is estimated; if the link between two rays is short or their vertices almost coincide, the relationship between the two log-ratios is constant and the two log-ratio are proportional<sup>37,48</sup>. The barycentre of Biplot represent the geometric mean, used in the *clr* transformation, and can be considered as a reference to evaluate the increase or the decrease of a variable respect to the whole composition. The ray length of the compositional data (for example the length of CO in Fig. 7) indicates the behaviour of the variable with respect to the centre of gravity (barycentre of Biplot) and provides an estimate of the standard deviation of a variable (approximately proportional to the own standard deviation, note that we are in a subspace, that is we have approximated the processes).

### Dataset and preliminary analysis

The raw geochemical dataset consists of time series of nine geochemical variables sampled with a frequency of about a sample/month; this dataset is constituted of 243 chemical analyses<sup>49</sup> obtained from samples collected at the main fumarolic vent in Solfatarata, Bocca Grande (BG, Fig. 1), between August 2000 and October 2020. Sampling techniques and analytical procedures are reported in Caliro et al.<sup>12</sup>. Chemical data of fumarolic fluids are expressed as micromole/mole ( $\mu\text{mol}/\text{mol}$ ) for  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{S}$ , Ar,  $\text{N}_2$ ,  $\text{CH}_4$ ,  $\text{H}_2$ , He, CO gas species. Fumarolic gases do not show any detectable  $\text{SO}_2$ , HCl, and HF, due to the scrubbing of magmatic gases within the hydrothermal system<sup>12,50,51</sup>. The gas chemical compositions exhibit significant changes over time due to the periodic contributions of hotter and more oxidizing magmatic fluids entering the hydrothermal system<sup>4,17,28,49,52-54</sup>.

The examination of the behaviour of a single geochemical time series is distortive because the data are of a compositional nature<sup>33</sup>. Each statistical analysis is correct considering the ratios between two gas species (ratio is compositional invariant) although the number of independent ratios  $[(p-1)^2/2]$  makes complex a global analysis of the geochemical data based on the ratios. We will use the PCA analysis to explore the structure of the compositional dataset and to obtain a new and reduced number of variables for further analysis.

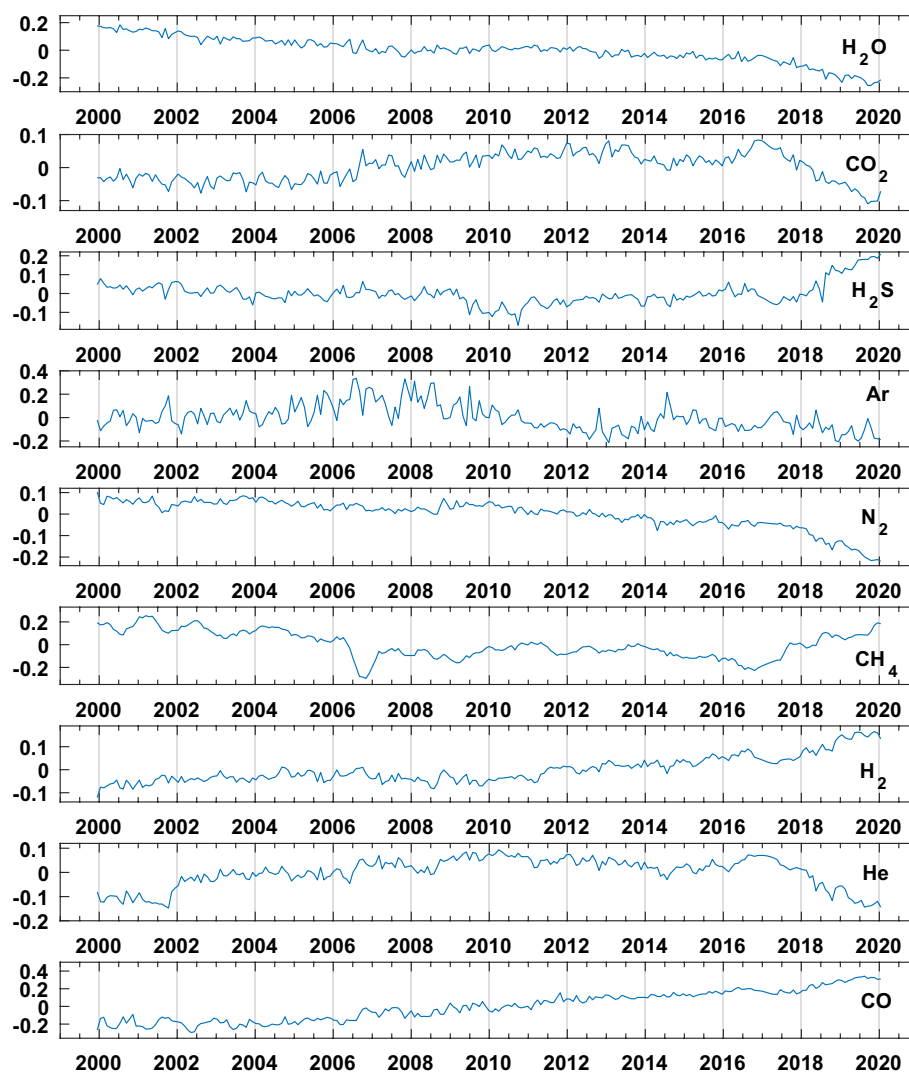
Before applying the PCA, we used the *ilr* transformation to cut-out outliers and to resample geochemical data at a constant step of a sample/month. Then we back transformed data to the simplex space. In fact, the *ilr* subspace, having orthogonal axes and  $p-1$  dimensions, allows the interpolations of each variable independently from the others<sup>33,55,56</sup>. After the interpolation, the *clr* transformation was adopted.

The geochemical time series resampled and transformed in *clr* are shown in Fig. 2, while their values are in the “Supplemental material”. In particular, it seems that, from a geochemical point of view, the crisis starts around the year 2006 with a clear fall of the  $\text{CH}_4$  content in the volcanic gasses with consequent increase in the typical ratios used to monitor the volcanic deep activity ( $\text{CO}_2/\text{CH}_4$  and  $\text{He}/\text{CH}_4$ ).

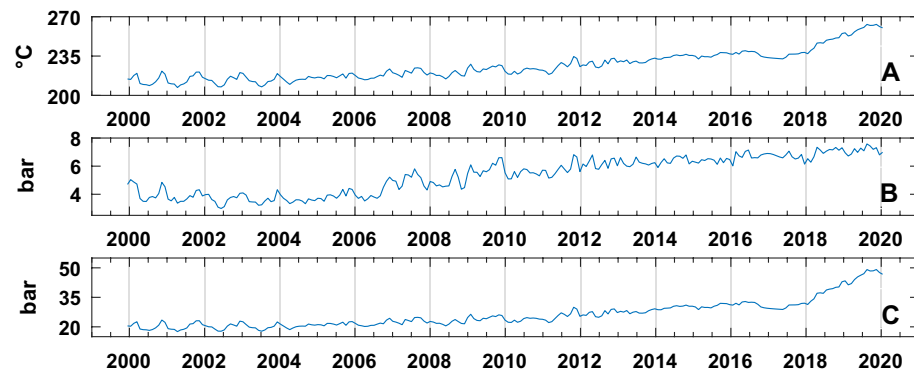
Moreover, geochemical variables derived from gas equilibria have also been considered in the multivariate approach. Equilibrium temperatures and pressures, in the  $\text{CO}-\text{CO}_2$  gas system, are computed according to Chiodini<sup>49</sup> considering the water fugacity controlled by the steam-liquid coexistence<sup>57</sup> and redox conditions fixed by the D’Amore and Panichi<sup>58</sup> empirical buffer.

Temperature is a function of ratio  $\text{CO}/\text{CO}_2$ ;  $\text{H}_2\text{O}$  pressure is a function of Temperature;  $\text{CO}_2$  pressure is a function of ratio  $\text{H}_2/\text{CO}$  (all the derived quantities depend on compositional invariant variable); the derivation of the geothermometric and geobarometric function was computed according to Chiodini et al.<sup>49</sup>, considering (i)  $f_{\text{H}_2\text{O}}$  fixed by the vapour-liquid coexistence and (ii)  $f_{\text{O}_2}$  as a function of the temperature. Redox conditions of Solfatara gases were assumed to be controlled by the DP buffer ( $\log f_{\text{O}_2} = 8.20 - 23,643/T$ ). The correspondent geothermometric relations are:  $T = 3133.5 / (0.933 - \log(X_{\text{CO}}/X_{\text{CO}_2}))$ . The geobarometric functions are:  $\log P_{\text{H}_2\text{O}} = 5.510 - 2048/T$ , where the water pressure is assumed equal to water fugacity of saturated vapor (i.e. vapor-liquid coexistence for pure water<sup>57</sup>),  $\log P_{\text{CO}_2} = 3.025 + 201/T - \log(X_{\text{H}_2}/X_{\text{CO}})$  and  $P_{\text{tot}} \sim P_{\text{CO}_2} + P_{\text{H}_2\text{O}}$  (Fig. 3).

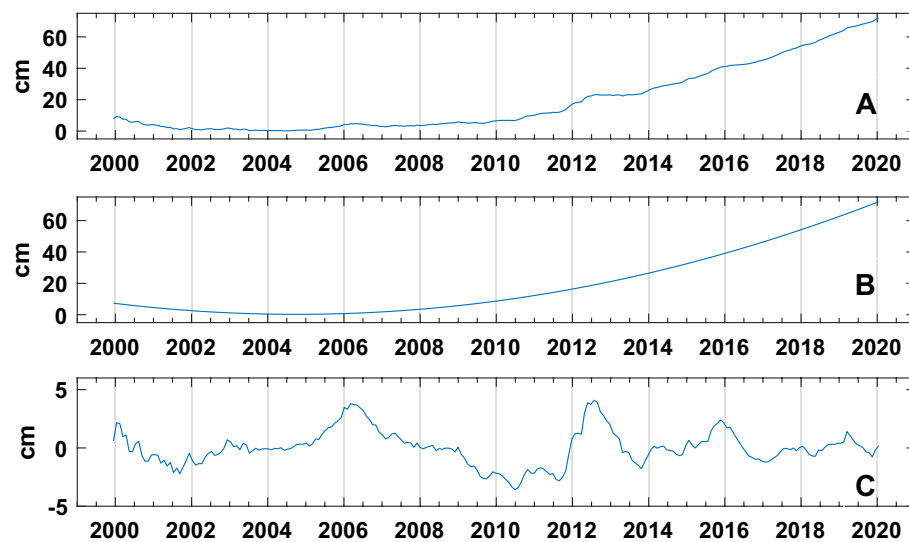
The vertical displacements dataset is composed of monthly averaged measurements recorded at the RITE station by the Neapolitan Volcanoes Continuous GPS (NeVoCGPS) network from 2000 to 2020 (Fig. 4A). The daily original recordings are available in Tramelli et. al.<sup>59</sup>. During the analyzed period, a subsidence phase switched to a slow uplift around 2005 and rose to a fast uplift phase in 2012 that is still ongoing. The maximum value was reached in the last analyzed year (2020) and was of 68.62 cm. Following the hypothesis that the complex displacement pattern is generated by the superposition of deformation processes separated in frequency (first intuited by Chiodini et al.<sup>5</sup> and then formalized in Petrillo et al.<sup>60</sup>), the time series has been separated into two distinct



**Figure 2.** Gas species after centered log ratio (*clr*) transformation.



**Figure 3.** Estimated T–P from gas equilibria in the CO–CO<sub>2</sub> system: (A) Temperature (in °C); (B) CO<sub>2</sub> pressure (in bar); (C) H<sub>2</sub>O pressure (in bar) (see text for details).

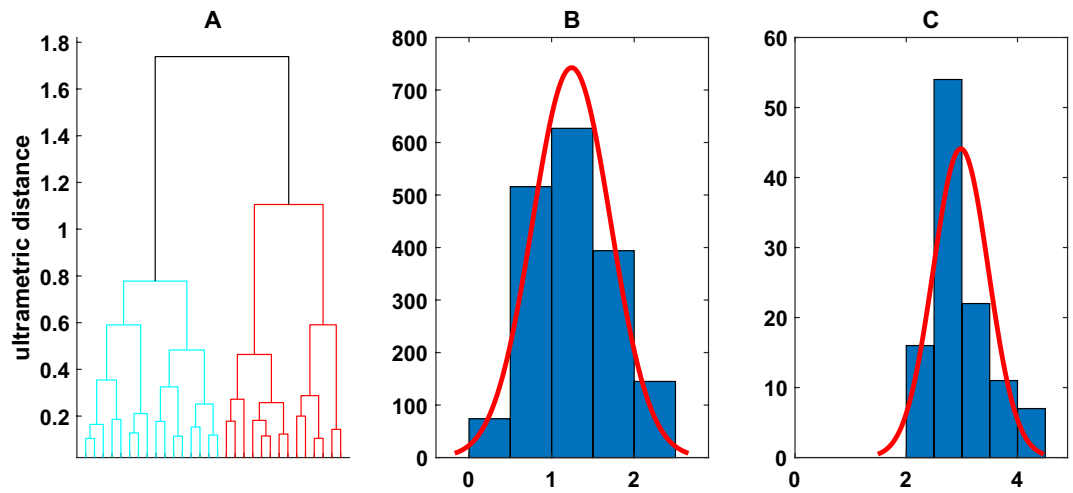


**Figure 4.** (A) Ground displacement in cm; (B) First geophysical variable (z-trend): Ground Displacement trend; (C) Second geophysical variable (z-osc): Ground displacement subtracted for the z-trend.

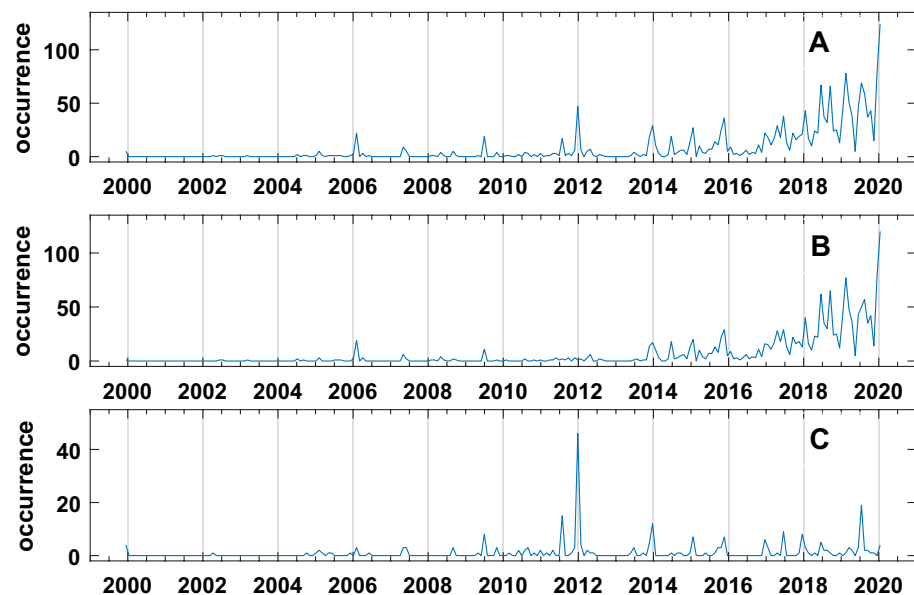
time series: the trend obtained by a second order polynomial fitting (Fig. 4B) and the oscillating component (Fig. 4C) obtained by subtracting the trend. Hereafter, these two time series will be mentioned respectively as z-trend and z-osc. It is noteworthy that in the z-osc series is evident a peak around the year 2006 (as in the CH<sub>4</sub> geochemical data series) as well as in the original up-lift time series where, around the same year, there is a step followed by an increasing trend (Fig. 4A).

The seismicity dataset is composed by monthly number of earthquakes located in the CFC area (Fig. 1). As reported by the catalogue of Osservatorio Vesuviano, Istituto Nazionale di Geofisica e Vulcanologia, between August 2000 up to October 2020, 2459 seismic events mainly occurred beneath the Solfatara–Pozzuoli area at depths from 0 to 4.46 km with the exception of a single (low quality and here not considered for further analysis) event at a depth of 7 km. The Gutenberg–Richter distribution<sup>61</sup> closely fit the data with magnitude  $M \geq -0.5$ . In this study we have selected 1866 volcano tectonic earthquakes with  $M \geq -0.5$  at which 93% of the observed data (events) are modelled by a straight line<sup>62</sup>. From 2000 to 2020, seismicity increased in time and clustered to shallow depth. Analyzing the recent seismic activity, Chiodini et al.<sup>54</sup> found that seismicity is distributed into low (swarms) and high (background) interarrival time populations<sup>63–65</sup>. The skewness of the hypocentral frequency distribution (in km) is 1.45, the mean is 1.34 with a standard deviation of 0.63, and the median is 1.27, which indicates a crowding of hypocentres towards shallow levels. We assume that there is a spatial discrimination between the two groups of earthquakes based on their depth. To test this hypothesis, a clustering algorithm (hierarchical linkage and Euclidian distance)<sup>66</sup> on the hypocentral depths was applied. The resultant dendrogram (cophenetic index = 0.79) is shown in Fig. 5A.

Two distinct depth classes are evidenced (Fig. 5B,C): one between 0 and 2.36 km (first cluster, Fig. 5B) and the other below 2.36 km (second cluster, Fig. 5C). In Fig. 6 the two earthquakes time series related to the two clusters are shown. The very interesting result is the low value of temporal correlation (0.20) between the shallow



**Figure 5.** Earthquakes occurrences (1866 events) between 2000 and 2020: (A) Dendrogram of Earthquakes classified by depth: the left descending branch relates to Earthquakes with a hypocentral depth  $\leq 2.36$  km, the right descending branch relates to the deeper Earthquakes (hypocentral depth  $> 2.36$  km); (B) Histogram of 1756 Earthquakes with hypocentral depths  $\leq 2.36$  km; (C) Histogram of 110 Earthquakes with hypocentral depths  $> 2.36$  km.



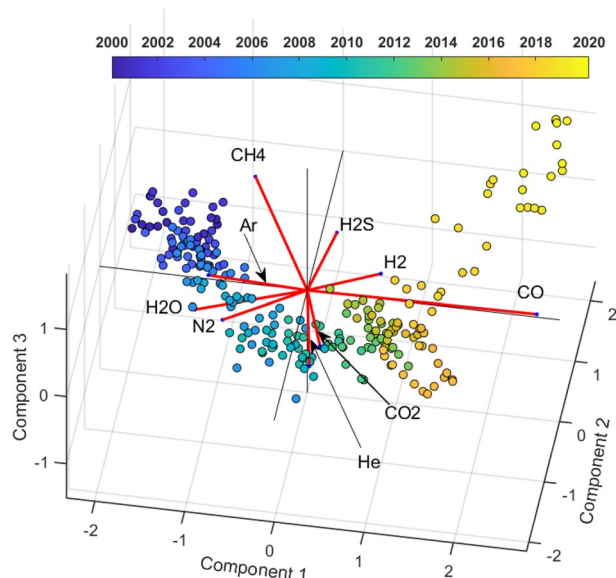
**Figure 6.** (A) Earthquakes occurrences; (B) Shallow Earthquakes occurrences (hypocentral depth  $\leq 2.36$  km); (C) Deep Earthquakes occurrences (hypocentral depth  $> 2.36$  km).

and deep earthquakes time series, which, with the spatial clustering, strongly supports our hypothesis that there is the presence of two distinct (time/space) physical seismic processes. The structure of the earthquakes occurrence again shows, in a coherent manner with the others geophysical and geochemical variables, a significant peak of the earthquakes occurrence around the 2006.

In particular, the correlation between the original ground displacement (Fig. 4A) and the earthquakes time series related to the first cluster is 0.72, while the correlation between the ground displacement and the earthquakes time series related to the second cluster is 0.21.

Hereafter the shallow and deep earthquakes occurrence time series will be mentioned respectively as: s-E and d-E.





**Figure 7.** Relative covariance biplot of Centered Log Ratio (clr) Gas Species (in red) on the first three factorial planes (see text for details). Note that the scores distribution (colored dots), representing the system state, is not chaotic, it draws an evolutionary path.

## Results and discussion

### Principal component analysis on geochemical species

To investigate the distribution of the geochemical variable in the reference periods, a PCA was applied on monthly sampled data, clr transformed by the Eq. (1) to obtain a relative covariance Biplot (Fig. 7) (Tables 1 and 2).

We look for hidden variables that control the changes of the fumarolic fluids compositions over time. To do this we will find the links which are best represented by the first three PCs (Tables 3, 4, 5).

Looking at the biplot of the first three components (Fig. 7) (95% explained variance), CO and CH<sub>4</sub> have the longest rays and, therefore, exhibit a much greater variability, in the time interval examined, respect to all the other geochemical species.

On the left side of the biplot (Fig. 7) there are the scores related to the period 2000–2005 (blue points) which reveal higher content in H<sub>2</sub>O, N<sub>2</sub> and CH<sub>4</sub>. If we follow the scores evolution through time, we can see that the system at  $t=2000$  is richer in CH<sub>4</sub> than at  $t>2005$ , when He and CO<sub>2</sub> enter the hydrothermal system and start to play a major role in the score distribution (projection). In the transition period (from 2005 to 2017) He and CO<sub>2</sub> mark a change in the hydrothermal system state evolution. From the year 2017 to 2020 we notice a decrease in the CO<sub>2</sub> and He content and an increase of CH<sub>4</sub> in the hydrothermal system. CO, which is the ray with higher variance, dominate the general trend.

In the context of the compositional theory, the analysis of individual species, in the clr space, leads to general results, but in particular we will use, even, the principal ratios (links) which are adopted in volcanic environments

PC	H <sub>2</sub> O	CO <sub>2</sub>	H <sub>2</sub> S	Ar	N <sub>2</sub>	CH <sub>4</sub>	H <sub>2</sub>	He	CO
PC1 (GF1)	-0.89	0.31	0.25	-0.53	-0.89	-0.46	0.87	0.15	0.99
PC2 (GF2)	0.02	-0.58	0.62	-0.60	-0.24	0.86	0.27	-0.74	0.00
PC3 (GF3)	-0.35	-0.67	0.43	0.59	-0.35	0.09	0.16	-0.51	0.01

**Table 1.** Correlations among variables and PCs of the nine geochemical species in Fig. 7.

PC	H <sub>2</sub> O	CO <sub>2</sub>	H <sub>2</sub> S	Ar	N <sub>2</sub>	CH <sub>4</sub>	H <sub>2</sub>	He	CO
PC1 (GF1)	0.80	0.10	0.07	0.28	0.79	0.22	0.76	0.02	0.99
PC2 (GF2)	0.00	0.34	0.38	0.37	0.06	0.73	0.08	0.55	0.00
PC3 (GF3)	0.12	0.45	0.18	0.35	0.12	0.00	0.03	0.26	0.00

**Table 2.** Squared cosines referring to the PCA of the nine geochemical species in Fig. 7.

	H <sub>2</sub> O	CO <sub>2</sub>	H <sub>2</sub> S	Ar	N <sub>2</sub>	CH <sub>4</sub>	H <sub>2</sub>	He	CO
Numerator									
H <sub>2</sub> O		-0.93	-0.79	-0.17	-0.53	-0.22	-0.92	-0.80	-0.99
CO <sub>2</sub>	0.93		-0.02	0.57	0.94	0.45	-0.50	0.11	-0.96
H <sub>2</sub> S	0.79	0.02		0.52	0.68	0.62	-0.60	0.05	-0.92
Ar	0.17	-0.57	-0.52		-0.01	-0.03	-0.72	-0.55	-0.92
N <sub>2</sub>	0.53	-0.94	-0.68	0.01		-0.03	-0.91	-0.86	-0.99
CH <sub>4</sub>	0.22	-0.45	-0.62	0.03	0.03		-0.75	-0.39	-0.91
H <sub>2</sub>	0.92	0.50	0.60	0.72	0.91	0.75		0.47	-0.95
He	0.80	-0.11	-0.05	0.55	0.86	0.39	-0.47		-0.92
CO	0.99	0.96	0.92	0.92	0.99	0.91	0.95	0.92	

**Table 3.** Correlations between the geochemical ratios and the first PC.

	H <sub>2</sub> O	CO <sub>2</sub>	H <sub>2</sub> S	Ar	N <sub>2</sub>	CH <sub>4</sub>	H <sub>2</sub>	He	CO
Numerator									
H <sub>2</sub> O		0.26	-0.28	0.56	0.41	-0.83	-0.10	0.41	0.01
CO <sub>2</sub>	-0.26		-0.68	0.34	-0.12	-0.83	-0.56	0.52	-0.16
H <sub>2</sub> S	0.28	0.68		0.74	0.48	-0.57	0.35	0.74	0.22
Ar	-0.56	-0.34	-0.74		-0.50	-0.94	-0.55	-0.19	-0.28
N <sub>2</sub>	-0.41	0.12	-0.48	0.50		-0.94	-0.27	0.37	-0.07
CH <sub>4</sub>	0.83	0.83	0.57	0.94	0.94		0.61	0.88	0.41
H <sub>2</sub>	0.10	0.56	-0.35	0.55	0.27	-0.61		0.69	0.13
He	-0.41	-0.52	-0.74	0.19	-0.37	-0.88	-0.69		-0.26
CO	-0.01	0.16	-0.22	0.28	0.07	-0.41	-0.13	0.26	

**Table 4.** Correlations between the geochemical ratios and the second PC.

	H <sub>2</sub> O	CO <sub>2</sub>	H <sub>2</sub> S	Ar	N <sub>2</sub>	CH <sub>4</sub>	H <sub>2</sub>	He	CO
Numerator									
H <sub>2</sub> O		-0.03	-0.47	-0.78	-0.22	-0.36	-0.29	-0.01	-0.14
CO <sub>2</sub>	0.03		-0.60	-0.74	-0.08	-0.26	-0.52	0.06	-0.20
H <sub>2</sub> S	0.47	0.60		-0.29	0.44	0.13	0.27	0.51	0.14
Ar	0.78	0.74	0.29		0.85	0.31	0.38	0.78	0.26
N <sub>2</sub>	0.22	0.08	-0.44	-0.85		-0.27	-0.27	0.11	-0.11
CH <sub>4</sub>	0.36	0.26	-0.13	-0.31	0.27		0.01	0.25	0.03
H <sub>2</sub>	0.29	0.52	-0.27	-0.38	0.27	-0.01		0.46	0.06
He	0.01	-0.06	-0.51	-0.78	-0.11	-0.25	-0.46		-0.20
CO	0.14	0.20	-0.14	-0.26	0.11	-0.03	-0.06	0.20	

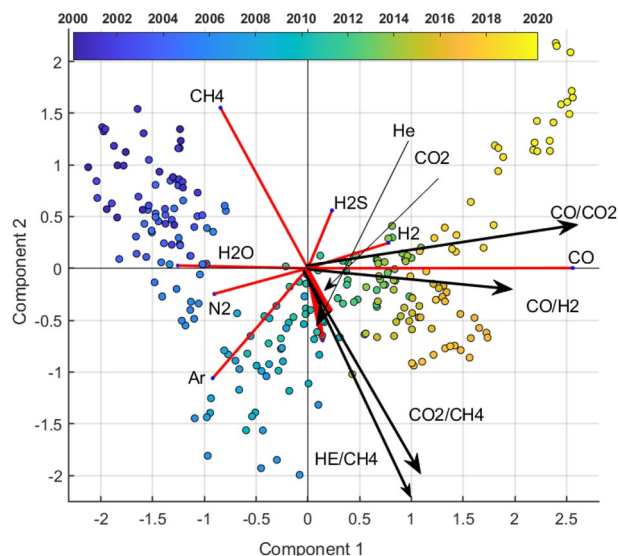
**Table 5.** Correlations between the geochemical ratios and the third PC.

that can be geochemically interpreted. Therefore, we consider the correlation between the main ratios and the first three PCs (Tables 3, 4, 5).

The links (ratios) with geochemical significance which have high correlation with the first PC are CO/CO<sub>2</sub> ( $r=0.96$ ) and CO/H<sub>2</sub> ( $r=0.95$ ) (Table 3 and Fig. 8). This associations and these ratios are notable because CO and H<sub>2</sub> are in general considered as gas species controlled by the Temperature and Pressure conditions of the hydrothermal systems at depth<sup>57</sup>. It should be emphasized that CO is related to all compositional species and all these relationships are strongly correlated with the first component ( $r \geq 0.91$ ). Considering that CO is the most representative species of the fumarole gas, delineating a warning trend<sup>53</sup> of the hydrothermal system conditions, we can hypothesize that the first axis (Figs. 7 and 8) can be interpreted as a latent process of the hydrothermal system heating/pressurizing. In the follow, we name this first principal component as GF1 (first geochemical factor).

The second component orthogonal to the previous one, opposes CH<sub>4</sub> and H<sub>2</sub>S to He, Ar and CO<sub>2</sub> (Fig. 8 and Tables 1, 2). In the period 2000–2006 the second principal component is positively dominated by CH<sub>4</sub> (which





**Figure 8.** Relative covariance biplot of Centered Log Ratio (clr) Gas Species on the first factorial plane with the representation of main links (ratios) having geochemical significance.

indicate a probably absence of deep magmatic fluids injection in the hydrothermal system) while, in the intermediate period 2007–2017, the contribution of He and CO<sub>2</sub> determines a transition with a relative decrease of CH<sub>4</sub>. The last period (2017–2020) is again characterized by a relative increase of CH<sub>4</sub>. In the factorial plane, Ar is present in the transition period, however, its occurrence is not further discussed due to likely significant air contamination (see the wide high frequency oscillation in the Ar signal of Fig. 2).

Even for the interpretation of the second axis we resort to the analysis of the correlations between the geochemical ratios and the second component (Table 4). The ratios representative of the hydrothermal dynamics He/CH<sub>4</sub> ratios ( $r = -0.88$ ) and CO<sub>2</sub>/CH<sub>4</sub> ( $r = -0.83$ ), are well projected on the second component.

These last ratios have been suggested by Chiodini et al.<sup>5,15,28,53,68</sup> as indicative of the arrivals of magmatic fluids in the hydrothermal system from depth. The CO<sub>2</sub>/CH<sub>4</sub> and He/CH<sub>4</sub> ratios have been interpreted as powerful indicators of magma degassing episodes<sup>65</sup>. The magmatic gases entering the hydrothermal system are, in fact, relatively rich in CO<sub>2</sub> and He and poor in CH<sub>4</sub>, a specie that is formed in the hydrothermal environment. Therefore, the second axis could represent a deep rooted hydrothermal process, that around 2006–2007 (see the projection of the data on the links CO<sub>2</sub>/CH<sub>4</sub> and He/CH<sub>4</sub> in Fig. 8) generated the pressure and temperature increase well represented by the first principal component.

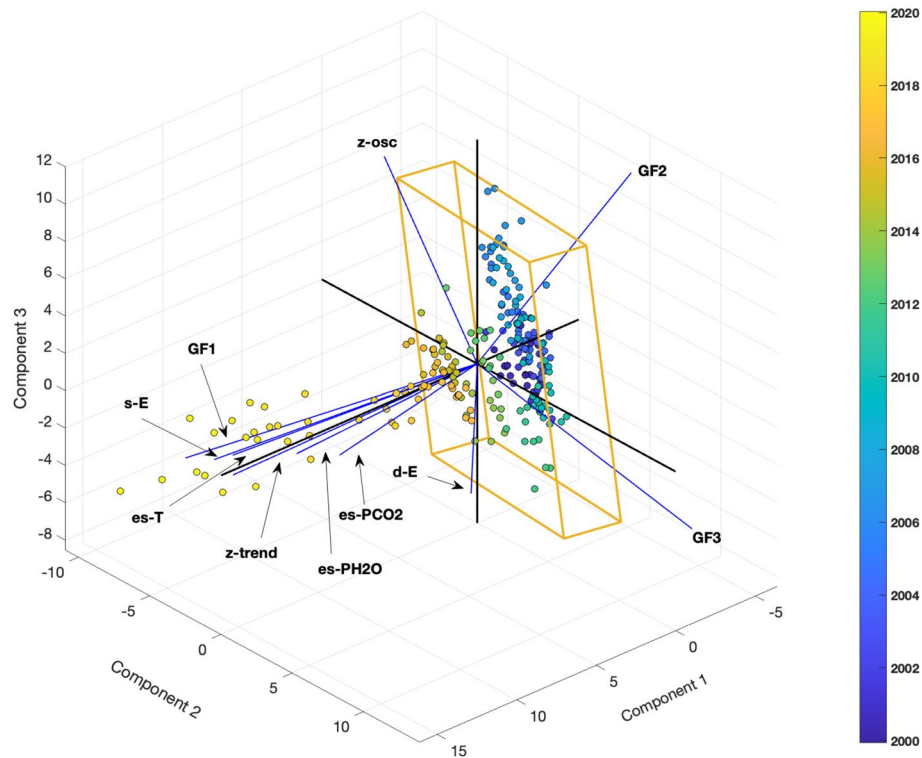
We underline that the first component is very strongly correlated (angles among the rays and the first axis very near zero, Fig. 7) with the CO increasing trend (Fig. 2) and H<sub>2</sub>O reverse trend (Fig. 2). The second component, orthogonal to the first one, represents an independent process dominated by CH<sub>4</sub>. Note, orthogonality implies that, even if the second component (deep magmatic fluid batches) could be the cause of the first, their direct statistical dependence must have been lost over time. In the follow, we call the second principal component GF2 (second geochemical factor).

The third component (12% explained variance) is dominated during the intermediate period (2007–2012) by CO<sub>2</sub> opposed to Ar and H<sub>2</sub>S (Fig. 7; Table 1). CO<sub>2</sub> is the best represented among the geochemical species on this axis (Table 2). We can hypothesize that the third principal component (hereafter GF3, third geochemical factor) can be considered a latent factor linked to the production of CO<sub>2</sub>, probably from deeper zone of the volcanic system.

### Joint PCA on geochemical and geophysical data

We were interested in the relationship over time among the geochemical data (Fig. 2), the geophysical data (Fig. 4B and C and Fig. 6B and C, and the derived T–P functions (Fig. 3). The Principal components analysis<sup>69</sup> was conducted on the multivariate dataset where the geochemical data are represented by GF1, GF2 and GF3 obtained from the previous analysis. We found that the 85% of the total variance was explained by the first three components. The graphical representation of the Biplot in Fig. 9 (together with the inferences obtained from the data listed in Tables 6 and 7) shows an impressive development of the scores trajectory, representing the state of the system. The dynamic, in the early year (2000–2008), is constrained substantially in a volume (parallelepiped in Fig. 9) mainly described by the second and the third PC, here the dynamics is governed by GF2, GF3, deep earthquakes and z-oscillation. From the year 2009 up to 2014 the system is losing its stability and tends to invade the first PC. Around the year 2015 up to the year 2020, the system has a violent impulse and the scores pattern moves away from the parallelepiped along the first PC.

The first PC (Fig. 9) is strongly and positively correlated with the z-trend, CO estimated temperature, H<sub>2</sub>O estimated pressure and GF1. More moderate, but always very significant, is the association of shallow earthquakes and CO<sub>2</sub> estimated pressure with the first component (Tables 6 and 7).



**Figure 9.** Biplot of the ground displacement trend (z-trend), ground displacement without trend (z-osc), shallow (s-E) and deep (d-E) Earthquakes occurrences, CO estimated Temperature (es-T), CO<sub>2</sub> estimated pressure (es-PCO<sub>2</sub>), H<sub>2</sub>O estimated pressure (es-PH<sub>2</sub>O), GF1, GF2 and GF3. The explained variance is 80% in the first three factorial planes.

PC	z-trend	z- oscillation	Shallow earthquakes	Deep earthquakes	Est. Temp. CO	Est. Press CO <sub>2</sub>	Est. Press H <sub>2</sub> O	GF1	GF2	GF3
PC1	0.96	0.01	0.84	0.36	0.99	0.88	0.98	<b>0.94</b>	- 0.21	- 0.19
PC2	- 0.03	- 0.40	- 0.38	0.37	0.00	0.36	- 0.08	0.24	0.47	<b>0.75</b>
PC3	0.04	- 0.55	0.17	0.16	- 0.05	- 0.09	- 0.00	- 0.14	- <b>0.79</b>	0.33

**Table 6.** Correlations among the first three PCs and the variables shown in Biplot of Fig. 9. Significant values are in [bold].

PC	z-trend	z-oscillation	Shallow earthquakes	Deep earthquakes	Est. Temp. CO	Est. Press CO <sub>2</sub>	Est. Press H <sub>2</sub> O	GF1	GF2	GF3
PC1	<b>0.92</b>	0.00	<b>0.70</b>	0.13	<b>0.97</b>	<b>0.78</b>	<b>0.97</b>	<b>0.89</b>	0.04	0.04
PC2	0.00	0.16	0.14	<b>0.14</b>	0.00	0.13	0.00	0.06	0.22	<b>0.56</b>
PC3	0.00	<b>0.30</b>	0.02	0.02	0.00	0.03	0.00	0.02	<b>0.62</b>	0.10

**Table 7.** Squared cosines of the variables referred to the Biplot in Fig. 9 (in bold the values that correspond. for each variable. to the factor for which the squared cosine is the largest). Significant values are in [bold].

These correlations, together with the spatial correspondence between shallow earthquakes (located between 0 and 2.36 km at depth) and the hydrothermal system, support a strong link between the two. According to Chiodini et al., 2021 heating/pressurizing of the hydrothermal system plays an active role in triggering low magnitude seismicity at shallow depth<sup>49</sup>. The group of variables which characterize the first component implies a very interesting and generally original interpretation of the deformation trend which could be led by the hydrothermal heating/pressurizing in the first 2–3 km of the volcanic apparatus. This part of the hydrothermal system should be responsible of the deformation trend (bradyseism) as well as of the shallow earthquake occurrence. In fact, the advective/convective fluids transport mechanism increases the stress/strain by fluid pressure and so the earthquakes occurrence. We interpret this first component, as representing fluids related processes occurring in the hydrothermal system.

The second component of PCA (Fig. 9) is strongly related only to GF3 (Tables 6 and 7) whose major variability is found around 2010–2012; note that the scores of this period are fully projected on the GF3 vector which, under the hypothesis that the GF3 factor is strongly correlated with the CO<sub>2</sub>, could be the trigger of the following crises, well represented by the first PCA (Fig. 9). This leads to interpret the second component (Fig. 9) as representing very deep volcanic processes related to the production of CO<sub>2</sub>. The second component has weaker negative correlations with z-osc and shallow earthquakes, while has weak positive correlations with GF2, deep earthquakes and CO<sub>2</sub> estimated pressure. Notably this component shows an absolute independence from the z-trend.

The third component of PCA (Fig. 9) is strongly related to GF2 (Tables 6 and 7) and to z-oscillations whose major variability is found around 2002–2012, from blue colour up to green in Fig. 10. Since GF2 could represent a process connected to the deep injection of magmatic fluids, we can suggest that the z-oscillations (in agreement with Chiodini et al.<sup>5</sup>) are generated by CH<sub>4</sub>—poor magmatic injections at the root of the hydrothermal system, probably below the region that hosts the most earthquakes (Fig. 1), as suggested by the very low correlations showed in Table 6.

## Conclusions

Twenty years of geochemical, thermodynamical and geophysical observations at CFC were analyzed by means of the PCA method. The goal was to reveal the basic phenomenon responsible for the ongoing state of a volcanic crisis which started around 2006 and perturbed the physical–chemical CFC magmatic state. The current volcanic hazard level (yellow) has been determined by the increase in intensity and frequency of the main unrest indicators: earthquakes occurrence, ground deformation and volcanic gas flux at surface. The multivariate statistical analysis, we have applied, suggests processes occurring both in the hydrothermal and in the magmatic system, describing how these processes evolve through time.

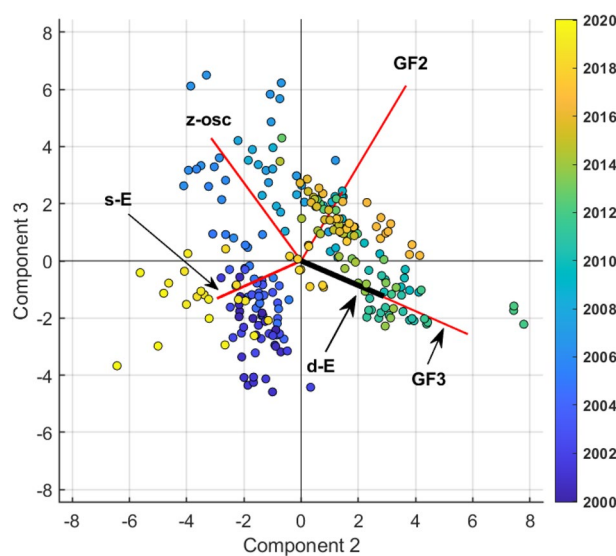
The heating/pressurizing processes strongly dominate the multivariate space together with the deformation trend and the shallow earthquakes occurrences (first geochemical/geophysical PCA) during the whole period; however, the primary contribution to this PCA is driven by the recent steeply increase of these processes. We interpret this first component, as representing processes occurring in the hydrothermal system and dominating the ongoing unrest.

The second and the third components that consider modulation of the processes through the variables oscillation, show substantially a significant projection of GF2, GF3, deep earthquakes and z-oscillation. These components, associated to deep volcanic processes, dominate the years 2000–2012. Injection of deep heated fluids, at the base of the hydrothermal system, could be responsible for the deformations pulses, as already discussed in Chiodini et al., 2015<sup>5</sup>. The trigger of the CFC volcanic unrest could be driven by the GF3 variability.

In conclusion, we can state that the unified and integrated approach on geochemical and geophysical indicators, applied in this study, has allowed to reveal the hidden and independent processes at the base of the CF volcanic crises, not clearly identifiable considering just a subset of them.

The results of this study are the basis for the identification of further and perhaps more effective geochemical relationships useful to improve the monitoring of the evolutionary volcanic processes which affect calderas similar to the Campi Flegrei one.

The adopted strategy, using the compositional theory applied on geochemical data from the CF caldera, offers a global interpretative framework with the confirmation that the geochemical processes are a keystone in



**Figure 10.** Biplot of only variables (in red and light blue) which have significant association with the second and third principal components (ground displacement without trend (z-osc), shallow (s-E) and deep (d-E) Earthquakes occurrences (the relative vector is dashed and in light blue), GF2 and GF3.

the interpretation of the volcanic phenomena at CF caldera. This methodology is certainly applicable to other calderas in the world in a similar state of hydrothermal activity.

## Data availability

The datasets analyzed during the current study are available in the section: “Supplementary Information”.

Received: 20 December 2022; Accepted: 16 October 2023

Published online: 27 October 2023

## References

- Acocella, V., Di Lorenzo, R., Newhall, C. & Scandone, R. An overview of recent (1988 to 2014) caldera unrest: Knowledge and perspectives: CALDERA UNREST. *Rev. Geophys.* **53**, 896–955 (2015).
- Sparks, R. S. J., Biggs, J. & Neuberg, J. W. Monitoring volcanoes. *Science* **335**, 1310–1311 (2012).
- Buono, G. *et al.* New insights into the recent magma dynamics under Campi Flegrei caldera (Italy) from petrological and geochemical evidence. *J. Geophys. Res. Solid Earth* **127**, e2021JB023773 (2022).
- Chiodini, G. *et al.* Magmas near the critical degassing pressure drive volcanic unrest towards a critical state. *Nat. Commun.* **7**, 13712 (2016).
- Chiodini, G. *et al.* Evidence of thermal-driven processes triggering the 2005–2014 unrest at Campi Flegrei caldera. *Earth Planet. Sci. Lett.* **414**, 58–67 (2015).
- De Vivo, B. *et al.* New constraints on the pyroclastic eruptive history of the Campanian volcanic Plain (Italy). *Mineral. Petrol.* **73**, 47–65 (2001).
- Deino, A. L., Orsi, G., de Vita, S. & Piochi, M. The age of the Neapolitan Yellow Tuff caldera-forming eruption (Campi Flegrei caldera – Italy) assessed by  $^{40}\text{Ar}/^{39}\text{Ar}$  dating method. *J. Volcanol. Geotherm. Res.* **133**, 157–170 (2004).
- Isaia, R., Marianelli, P. & Sbrana, A. Caldera unrest prior to intense volcanism in Campi Flegrei (Italy) at 4.0 ka BP: Implications for caldera dynamics and future eruptive scenarios. *Geophys. Res. Lett.* **36**, 213–217 (2009).
- Isaia, R., D’Antonio, M., Dell’Erba, F., Di Vito, M. & Orsi, G. The Astroni volcano: the only example of closely spaced eruptions in the same vent area during the recent history of the Campi Flegrei caldera (Italy). *J. Volcanol. Geotherm. Res.* **133**, 171–192 (2004).
- Di Vito, M., Lirer, L., Mastrolorenzo, G. & Rolandi, G. The 1538 Monte Nuovo eruption (Campi Flegrei, Italy). *Bull. Volcanol.* **49**, 608–615 (1987).
- Allard, P., Maiorani, A., Tedesco, D., Cortecchi, G. & Turi, B. Isotopic study of the origin of sulfur and carbon in Solfatara fumaroles, Campi Flegrei caldera. *J. Volcanol. Geotherm. Res.* **48**, 139–159 (1991).
- Caliro, S. *et al.* The origin of the fumaroles of La Solfatara (Campi Flegrei, south Italy). *Geochim. Cosmochim. Acta* **71**, 3040–3055 (2007).
- Weber, E. *et al.* Development and testing of an advanced monitoring infrastructure (ISNet) for seismic early-warning applications in the Campania region of southern Italy. In *Earthquake Early Warning Systems* (eds Gasparini, P. *et al.*) 325–341 (Springer, 2007).
- Iannaccone, G., Guardato, S., Vassallo, M., Elia, L. & Beranzoli, L. A new multidisciplinary marine monitoring system for the surveillance of volcanic and seismic areas. *Seismol. Res. Lett.* **80**, 203–213 (2009).
- Chiodini, G. *et al.* Long-term variations of the Campi Flegrei, Italy, volcanic system as revealed by the monitoring of hydrothermal activity. *J. Geophys. Res.* **115**, B03205 (2010).
- Somma, R. *et al.* Application of laser scanning for monitoring coastal cliff instability in the pozzuoli bay, coroglio site, posillipo hill, Naples. In *Engineering Geology for Society and Territory* Vol. 5 (eds Lollino, G. *et al.*) (Springer, 2014).
- Vilardo, G., Sansivero, F. & Chiodini, G. Long-term TIR imagery processing for spatiotemporal monitoring of surface thermal features in volcanic environment: A case study in the Campi Flegrei (Southern Italy). *J. Geophys. Res. Solid Earth* **120**, 812–826 (2015).
- Carlino, S. *et al.* Distributed-temperature-sensing using optical methods: A first application in the offshore area of Campi Flegrei caldera (Southern Italy) for volcano monitoring. *Remote Sens.* **8**, 674 (2016).
- Queißer, M., Granieri, D. & Burton, M. A new frontier in  $\text{CO}_2$  flux measurements using a highly portable DIAL laser system. *Sci. Rep.* **6**, 33834 (2016).
- Zaccarelli, L. & Bianco, F. Noise-based seismic monitoring of the Campi Flegrei caldera. *Geophys. Res. Lett.* **44**, 2237–2244 (2017).
- Pappalardo, L., Piochi, M., Dantonio, M., Civetta, L. & Petrinì, R. Evidence for multi-stage magmatic evolution during the past 60 kyr at Campi Flegrei (Italy) Deduced from Sr, Nd and Pb Isotope Data. *J. Petrol.* **43**, 1415–1434 (2002).
- Rossano, S., Mastrolorenzo, G. & De Natale, G. Numerical simulation of pyroclastic density currents on Campi Flegrei topography: A tool for statistical hazard estimation. *J. Volcanol. Geotherm. Res.* **132**, 1–14 (2004).
- Gottsmann, J., Camacho, A. G., Tiampo, K. F. & Fernández, J. Spatiotemporal variations in vertical gravity gradients at the Campi Flegrei caldera (Italy): A case for source multiplicity during unrest?. *Geophys. J. Int.* **167**, 1089–1096 (2006).
- Bodnar, R. J. *et al.* Quantitative model for magma degassing and ground deformation (bradyseism) at Campi Flegrei, Italy: Implications for future eruptions. *Geology* **35**, 791 (2007).
- Costa, A. *et al.* Tephra fallout hazard assessment at the Campi Flegrei caldera (Italy). *Bull. Volcanol.* **71**, 259–273 (2009).
- Lima, A. *et al.* Thermodynamic model for uplift and deflation episodes (bradyseism) associated with magmatic–hydrothermal activity at the Campi Flegrei (Italy). *Earth-Sci. Rev.* **97**, 44–58 (2009).
- Rinaldi, A. P., Todesco, M. & Bonafede, M. Hydrothermal instability and ground displacement at the Campi Flegrei caldera. *Phys. Earth Planet. Inter.* **178**, 155–161 (2010).
- Chiodini, G., Caliro, S., De Martino, P., Avino, R. & Gherardi, F. Early signals of new volcanic unrest at Campi Flegrei caldera? Insights from geochemical data and physical simulations. *Geology* **40**, 943–946 (2012).
- Petrillo, Z. *et al.* Defining a 3D physical model for the hydrothermal circulation at Campi Flegrei caldera (Italy). *J. Volcanol. Geotherm. Res.* **264**, 172–182 (2013).
- Ashley, R. P. & Lloyd, J. W. An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. *J. Hydrol.* **39**, 355–364 (1978).
- Wenning, R. J. & Erickson, G. A. Interpretation and analysis of complex environmental data using chemometric methods. *TrAC Trends Anal. Chem.* **13**, 446–457 (1994).
- Johnson, R. A. & Wichern, D. Multivariate analysis. In *Wiley StatsRef: Statistics Reference Online* (eds Balakrishnan, N. *et al.*) 1–20 (Wiley, 2015). <https://doi.org/10.1002/9781118445112.stat02623.pub2>.
- Filzmoser, P., Hron, K. & Reimann, C. Principal component analysis for compositional data with outliers. *Environmetrics* **20**, 621–632 (2009).
- Aitchison, J. & Greenacre, M. Biplots of compositional data. *J. R. Stat. Soc. Ser. C Appl. Stat.* **51**, 375–392 (2002).
- Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Methodol.* **44**, 139–160 (1982).
- Chayes, F. On correlation between variables of constant sum. *J. Geophys. Res.* **65**, 4185–4193 (1960).
- Pawlowsky-Glahn, V. & Buccianti, A. *Compositional Data Analysis: Theory and Applications* (John Wiley & Sons, 2011).

38. Pawlowsky-Glahn, V., Egozcue, J. J. & Tolosana-Delgado, R. *Modeling and Analysis of Compositional Data* (John Wiley & Sons, 2015).
39. Aitchison, J. *The Statistical Analysis of Compositional Data* (Chapman and Hall, 1986).
40. Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barcelo-Vidal, C. Isometric logratio transformations for compositional data analysis. *Math. Geol.* **35**, 279–300 (2003).
41. Egozcue, J. J. & Pawlowsky-Glahn, V. Groups of parts and their balances in compositional data analysis. *Math. Geol.* **37**, 795–828 (2005).
42. Van den Boogaart, K. G. & Van Tolosana-Delgado, R. *Analyzing Compositional Data with R* Vol. 122 (Springer, 2013).
43. Jolliffe, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
44. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417 (1933).
45. Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218 (1936).
46. Gabriel, K. R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58**, 453–467 (1971).
47. Greenacre Michael J. & Leslie G. Underhill Scaling a data matrix in a low-dimensional Euclidean space. Topics in Applied Multivariate Analysis, Hawkins, DM, ed., Cambridge UK: Cambridge University Press, 1982. Scaling a data matrix in a low-dimensional Euclidean space. In *Applied Multivariate Analysis* (Cambridge University Press, 1982).
48. Bacon-Shone, J., Buccianti, A., Mateu-Figueras, G. & Pawlowsky-Glahn, V. (eds) *Compositional Data Analysis in the Geosciences: From Theory to Practice* (2008).
49. Chiodini, G. et al. Hydrothermal pressure-temperature control on CO<sub>2</sub> emissions and seismicity at Campi Flegrei (Italy). *J. Volcanol. Geotherm. Res.* **414**, 107245 (2021).
50. Cioni, R., Corazza, E. & Marini, L. The gas/steam ratio as indicator of heat transfer at the Solfatara fumaroles, Phlegraean Fields (Italy). *Bull. Volcanol.* **47**, 295–302 (1984).
51. Chiodini, G. et al. CO<sub>2</sub> degassing and energy release at Solfatara volcano, Campi Flegrei. *Italy. J. Geophys. Res. Solid Earth* **106**, 16213–16221 (2001).
52. Chiodini, G. et al. Magma degassing as a trigger of bradyseismic events: The case of Phlegraean Fields (Italy). *Geophys. Res. Lett.* **30**, 1434 (2003).
53. Chiodini, G., Avino, R., Caliro, S. & Minopoli, C. Temperature and pressure gas geoindicators at the Solfatara fumaroles (Campi Flegrei). *Ann. Geophys.* **54**, 4 (2011).
54. Chiodini, G. et al. Clues on the origin of post-2000 earthquakes at Campi Flegrei caldera (Italy). *Sci. Rep.* **7**, 4472 (2017).
55. Filzmoser, P. & Hron, K. Outlier detection for compositional data using robust methods. *Math. Geosci.* **40**, 233–248 (2008).
56. Filzmoser, P., Hron, K. & Reimann, C. Interpretation of multivariate outliers for compositional data. *Comput. Geosci.* **39**, 77–85 (2012).
57. Guggenbach, W. F. Geothermal gas equilibria. *Geochim. Cosmochim. Acta* **44**, 2021–2032 (1980).
58. D'Amore, F. & Panichi, C. Evaluation of deep temperatures of hydrothermal systems by a new gas geothermometer. *Geochim. Cosmochim. Acta* **44**, 549–556 (1980).
59. Tramelli, A. et al. Statistics of seismicity to investigate the Campi Flegrei caldera unrest. *Sci. Rep.* **11**, 1–10 (2021).
60. Petrillo, Z. et al. A perturbative approach for modeling short-term fluid-driven ground deformation episodes on volcanoes: A case study in the Campi Flegrei caldera (Italy). *J. Geophys. Res. Solid Earth* **124**, 1036–1056 (2019).
61. Gutenberg, B. & Richter, C. F. *Seismicity of the Earth and Associated Phenomena* (Princeton University Press, 1954).
62. Wiemer, S. Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bull. Seismol. Soc. Am.* **90**, 859–869 (2000).
63. Petrosino, S., Cusano, P. & Madonia, P. Tidal and hydrological periodicities of seismicity reveal new risk scenarios at Campi Flegrei caldera. *Sci. Rep.* **8**, 13808 (2018).
64. Petrosino, S. & De Siena, L. Fluid migrations and volcanic earthquakes from depolarized ambient noise. *Nat. Commun.* **12**, 6656 (2021).
65. Tramelli, A., Giudicepietro, F., Ricciolino, P. & Chiodini, G. The seismicity of Campi Flegrei in the contest of an evolving long term unrest. *Sci. Rep.* **12**, 2900 (2022).
66. Gordon, A. D. *Classification* (CRC Press, 1999).
67. Chiodini, G. & Marini, L. Hydrothermal gas equilibria: the H<sub>2</sub>O-H<sub>2</sub>-CO<sub>2</sub>-CO-CH<sub>4</sub> system. *Geochim. Cosmochim. Acta* **62**, 2673–2687 (1998).
68. Chiodini, G. CO<sub>2</sub>/CH<sub>4</sub> ratio in fumaroles a powerful tool to detect magma degassing episodes at quiescent volcanoes: CO<sub>2</sub>/CH<sub>4</sub> in fumaroles and magma degassing. *Geophys. Res. Lett.* **36**, (2009).
69. Jolliffe, I. T. A note on the use of principal components in regression. *Appl. Stat.* **31**, 300 (1982).

## Acknowledgements

This study was partially supported by the Progetto Strategico Dipartimentale INGV 2019 “Linking surface Observables to sub-Volcanic plumbing-system: A multidisciplinary approach for Eruption forecasting at Campi Flegrei caldera (Italy)”, LOVE-CF and by the INGV Institutional funds: Ricerca Libera 2021. The data analysed in this study were provided by Unità Funzionale Geochimica dei fluidi, Istituto Nazionale di Geofisica e Vulcanologia (Osservatorio Vesuviano).

## Author contributions

Z.P., S.T., A.M. and S.S. contributed to the conception of the study; Z.P., S.T., A.M. and S.S. wrote the paper, S.C. and G.C. contributed to the data interpretation; S.C. contributed to the acquisition and to the data analysis; all authors revised the manuscript; all authors approved the submitted version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-45108-0>.

**Correspondence** and requests for materials should be addressed to A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023