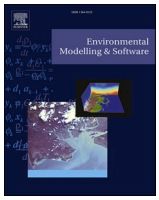
Environmental Modelling and Software 193 (2025) 106652

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



sGs UnMix: a web application for spatial prediction and mixture modeling with a case study on volcanic soil CO_2 fluxes

Giulio Bini *, Giancarlo Tamburello, Stefano Cacciaguerra, Paolo Perfetti

Istituto Nazionale di Geofisica e Vulcanologia, Viale Berti Pichat 6/2, 40127, Bologna, Italy

ARTICLE INFO

Keywords:
Spatial prediction
Geostatistics
Soil CO2 flux
Shiny web application

ABSTRACT

Spatial data analysis and prediction are fundamental in geoscience for mapping continuous variables and supporting decision-making. However, traditional geostatistical tools often require programming skills or involve manual, subjective steps. Here, we developed sGs UnMix, an interactive web application that simplifies spatial prediction workflows and reduces subjectivity in statistical analysis, making it accessible to the entire geoscience community. sGs UnMix (available online at https://apps.bo.ingv.it/sgs-unmix) is built with the shiny package for R and is organized into four main panels, which allow data loading and coordinate projection, data separation through mixture modeling, variogram modeling, and spatial prediction using sequential Gaussian simulation (sGs). Automated variogram fitting and mixture modeling reduce user bias, while dynamically updated heat maps enable real-time visualization of spatial patterns. sGs UnMix provides not only a standardized approach for estimating volcanic volatile fluxes (e.g., soil CO_2 emissions) but also applications in ore deposit mapping, hydrocarbon exploration, environmental monitoring, and climatology. Compared to existing geostatistical tools, it offers automation, interactivity, and a platform-independent, standalone web-based solution for geoscientists.

1. Introduction

The statistical analysis of spatial data is crucial for understanding spatial and spatiotemporal phenomena, playing an integral role across various fields in geosciences. Many of the foundational techniques for spatial data analysis were first developed in the context of mineral resource exploration. One such technique, kriging, was pioneered by Georges Matheron in the 1960s (Matheron, 1962, 1963), drawing inspiration from the work of South African mining engineer Daniel Krige (1951). Kriging is an interpolation technique widely used to estimate values at unsampled locations based on spatially distributed data. It is grounded in the concept of the variogram, which quantifies the spatial continuity of the data as a function of distance, thereby providing a mathematical framework for making spatial estimation and prediction (Matheron, 1963a,b). However, kriging tends to produce smoothed estimates that underestimate spatial variability (Journel and Huijbregts, 1978). To overcome this limitation, more sophisticated techniques, such as Sequential Gaussian Simulation (sGs from now on), have been developed (Journel and Alabert, 1989). sGs preserves both the variance and spatial continuity of the data by incorporating multiple realizations of spatially distributed values, making it more suitable for spatial prediction (Pyrcz and Deutsch, 2014).

Spatial prediction using sGs has found applications far beyond mining, becoming essential in reservoir modeling and several soil-science related fields, including environmental science, climatology, agriculture, public health, hydrology, and natural resource management. In environmental science, sGs has been applied to model the distribution of soil contaminants, such as cadmium, providing more realistic risk maps (Goovaerts, 1997). In agriculture, sGs has been used to map soil nutrients, such as potassium (Webster and Oliver, 2007), enabling precision farming practices and better resource management. Mapping of soil properties with sGs, such as pH, has been also performed to study the effect of urbanization on land (Sun et al., 2013). Public health applications include predicting heavy metal concentrations in topsoil and calculating the carcinogenic health-risk level for each node of the grid (Özen et al., 2022). sGs has been applied in hydrology to map the soil water content to identify areas prone to surface runoff and erosion (Delbari et al., 2009) and in petroleum reservoir modeling to estimate porosity and permeability for hydrocarbon recovery (Pyrcz and Deutsch, 2014).

Spatial prediction using sGs has also become indispensable for volcano monitoring and geothermal exploration, especially for mapping

* Corresponding author.
E-mail address: giulio.bini@ingv.it (G. Bini).

https://doi.org/10.1016/j.envsoft.2025.106652
Received 12 March 2025; Received in revised form 27 June 2025; Accepted 14 August 2025
Available online 16 August 2025

soil CO_2 fluxes (Cardellini et al., 2003). As CO_2 is the second most abundant gas emitted by volcanoes (after H_2O), and a significant portion of it escapes through soil diffusion (Fischer et al., 2019; Werner et al., 2019), accurate measurements and spatial prediction is essential for quantifying volcanic CO_2 emissions and understanding the degassing extent. Furthermore, the spatial correlation between soil temperature and CO_2 flux maps constructed with sGs has proven particularly useful for locating upflow zones in geothermal systems, providing a valuable tool for geothermal energy exploration (Chiodini et al., 2005, 2021a). Cardellini et al. (2003) first applied sGs to soil CO_2 emission data to map volcanic structures, such as faults and craters, which serve as pathways for gas escape, and estimate CO_2 emissions from various volcanoes, including Campi Flegrei (Italy), Vesuvius (Italy), and Nisyros (Greece). Despite its advantages, sGs has not been entirely adopted within the volcanological and geological communities, mainly due to the solid understanding of geostatistical principles needed. As a result, interpolation methods like inverse distance weighting (IDW) or kriging have sometimes been preferred (e.g., Carapezza et al., 2011; Tarchini et al., 2019; Gurrieri et al., 2023; Taussi et al., 2023; Tardani et al., 2024). Although these methods find useful applications in some circumstances–e.g., IDW is commonly used when data do not show spatial correlation (i.e., the variogram only displays the nugget effect)–, they fail to accurately reproduce data variance and spatial continuity, leading to the underestimation of volcanic CO_2 flux and degassing extent (e.g., Lewicki et al., 2005). This inconsistent application of methods underscores the need for a standardized approach to processing soil CO_2 emission data, ensuring that total flux estimates are comparable across different volcanoes.

While open-source geostatistical software packages such as GSLIB (Deutsch and Journel, 1998) offer powerful tools for spatial modeling, they can require programming knowledge in languages like Fortran, Python, or R. Even graphical user interface (GUI)-based tools, while somewhat more user-friendly, can be cumbersome for those without a deep understanding of geostatistics. Additionally, many of these platforms lack automated features, such as auto-fitting variograms, which introduces operator subjectivity and can affect the accuracy of results. These barriers have made advanced geostatistical tools, including sGs, inaccessible to part of the geoscience community. To address these challenges, we developed sGs UnMix, a user-friendly, interactive web application that simplifies the process of spatial prediction using sGs. While the app does not eliminate the need for critical interpretation, it is designed to guide users—regardless of programming experience—through the modeling process in an intuitive and transparent way, helping them to better explore and valorize their datasets. The web app is structured around five interactive panels, guiding users from data upload to the generation of mean and probability maps. Users can choose between an automated variogram fitting procedure based on least-squares or a manual adjustment, with real-time updates displayed on variogram and density plots. The app also provides interactive maps overlaid on GIS-layers, enabling users to visualize the spatial distribution of predicted values, identify anomalies, and gain deeper insights into the processes underlying spatial data patterns. This tool offers a faster, more intuitive approach to decision-making in resource exploration, volcano monitoring, environmental risk assessment, and beyond.

2. Methods

sGs UnMix has been designed with shiny (https://shiny.posit.co), an online platform that enables the creation of web applications based on the R programming language (https://www.r-project.org). sGs UnMix is available online at https://apps.bo.ingv.it/sgs-unmix. The web app is based on four modulus, and each of these relies on fundamental concepts of geostatistics needed to unmix different populations of data within the total spatial distribution, modeling the spatial continuity of the data with the variogram, and performing spatial prediction with sGs. To these scopes, sGs UnMix uses several R packages, which are listed in Table 1.

Table 1
R packages used by sGs UnMix.

R Package	Authors	Link
shiny	Chang et al. (2024)	https://cran.r-project.org/package=shiny
sp	Pebesma and Bivand (2005); Bivand et al. (2013)	https://cran.r-project.org/package=sp
splancs	Rowlingson and Diggle (2024)	https://cran.r-project.org/package=splancs
fields	Nychka et al. (2021)	https://github.com/dnychka/fieldsRPackage.
raster	Hijmans (2024a)	https://cran.r-project.org/package=raster
terra	Hijmans (2024b)	https://cran.r-project.org/package=terra
gstat	Pebesma (2004)	https://cran.r-project.org/package=gstat
leaflet	Cheng et al. (2024)	https://cran.r-project.org/package=leaflet
leaflet.extras	Gatscha et al. (2024)	https://cran.r-project.org/package=leaflet.extras
leaflet.providers	Huang (2023)	https://cran.r-project.org/package=leaflet.providers
DT	Xie et al. (2024)	https://cran.r-project.org/package=DT
mixtools	Benaglia et al. (2009)	https://cran.r-project.org/package=mixtools

2.1. Unmixing lognormal populations of data

The distribution of the variable we want to predict at unsampled locations in the space can reflect the mixing of two or more populations of data, potentially indicating different data sources. For example, soil CO_2 flux data often reflect the overlapping of two lognormal populations, characterized by low and high flux values, reflecting the background, biogenic gas release and the magmatic-hydrothermal CO_2 coming from depth. Unmixing of these populations has been treated in the literature with the method of Sinclair (1974), which was used in the Graphical Statistical Approach (GSA; Chiodini et al., 1998) to disentangle biogenic and magmatic soil CO_2 flux populations and compute the respective total emission into the atmosphere (t d^{-1}). In sGs UnMix, we implemented the *normalmixEM* function of the mixtools package for R (Benaglia et al., 2009). This function uses an Expectation-Maximization (EM) algorithm, which alternates between estimating the probability of each data point belonging to each population (E-step) and updating the parameters describing each population (means, variances, and mixing proportions) to maximize the likelihood (M-step). This iterative process is able to automatically separate the data into two or more individual populations. In case of lognormal populations, sGs UnMix then uses a Monte Carlo simulation to calculate the mean and standard deviation of each population (e.g., Chiodini et al., 2015). In particular, this process involves drawing a random sample, equal in size to the original dataset, from each population identified by the EM algorithm. Each value in the sample is raised to the power of 10, and the average of these transformed values is computed. This operation is repeated 5000 times, generating 5000 mean values. From these, the overall mean and standard deviation of the population are calculated.

2.2. Modeling the spatial continuity of the data with the variogram

Since spatial prediction is performed through sGs, sGs UnMix first transforms the data from real space to Gaussian space. All the following calculations are performed on the transformed data. The spatial continuity of the data is measured through the semivariance (Matheron, 1963a,b)

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z(u_\alpha) - z(u_\alpha + h))^2 \quad (1)$$

which expresses half of the average squared difference of values separated by a distance equal to *h*, the lag distance, where $z(u_\alpha)$ is the value at location α and $N(h)$ is the number of pairs for lag distance *h*. The lag distance is automatically calculated by taking into account the geometry of the grid (Pebesma, 2004), but it can also be manually chosen to ensure a variogram calculation with a large number of pairs for each lag. sGs UnMix uses the *variogram* function of the gstat package (Pebesma, 2004) and plots not only the variogram (Fig. 1) but also the variogram cloud that represents all the possible square differences of values separated by distance *h*. This diagram can be particularly informative in locating potential outliers showing high semivariance at short distances, which results in a significant nugget effect, affecting the spatial continuity of the data (see Example #2 in the Supplementary Material). In such circumstances, a robust variogram estimator is preferred (Cressie, 1993)

$$\gamma(h) = \frac{\frac{1}{2}\left(\frac{1}{N(h)}\sum_{\alpha=1}^{N(h)}\sqrt{|z(u_\alpha) - z(u_\alpha + h)|}\right)^4}{0.457 + \frac{0.494}{N(h)}} \quad (2)$$

The spatial continuity of the data is then modeled by fitting the empirical variogram with a specific variogram model (Table A.1) that minimizes a weighted sum of squared errors (SSE; Pebesma, 2004)

$$\sum_{j=1}^{n} w_j\left(\hat{\gamma}(h_j) - \gamma(h_j)\right)^2 \quad (3)$$

where *j* is the number of lag distance at which $\gamma(h)$ is calculated, w_j is the weight equals to $N(h)_j/h_j^2$, and $\hat{\gamma}(h)$ is the semivariance of the fitted variogram model. For this operation, sGs UnMix uses the *fit.variogram* function of the gstat package (Pebesma, 2004) and returns the calculated nugget, range, and sill, which reflect the structure of the variogram model (Fig. 1). Such components may be adjusted in sGs UnMix by switching from automatic to manual mode.

2.3. Spatial prediction through sequential Gaussian simulations (sGs)

sGs UnMix incorporates the variogram model to perform spatial prediction through sGs using the *krige* function of the gstat package (Pebesma, 2004). The transformed data are distributed over the grid, and a random path is defined. The algorithm searches for 40 nearby measured and previously simulated data at each location along this path. These data are processed with ordinary kriging, an interpolation technique that relies on linear weights attributed to the data, which account for spatial continuity, data closeness, and redundancy. Ordinary kriging (OK) was chosen over simple kriging (SK) because it does not assume a known global mean, instead estimating the local mean from neighboring data. This is advantageous for datasets where the mean varies spatially or is poorly known, such as CO_2 emissions from volcanic soils. This step enables the algorithm to construct a normal distribution with a mean equal to the kriging estimate and variance equal to the kriging variance. The simulated value is drawn from this normal distribution with a Monte Carlo simulation and added to the random location of the grid. Extracting the simulated value from this distribution accurately reproduces the data variance, which would be underestimated if we relied solely on the kriging estimate. This procedure is then repeated across all locations of the random path, progressively including the simulated values in the simulations. By considering previously simulated values as data, the algorithm preserved the covariance structure of the entire dataset. After all grid nodes are populated with simulated values, all the data are back-transformed to the real space, producing the first realization. The algorithm repeats this procedure for *n* realizations, following a different random path each time, resulting in different simulated values. All realizations are equiprobable, and averaging them provides an estimate of uncertainty.

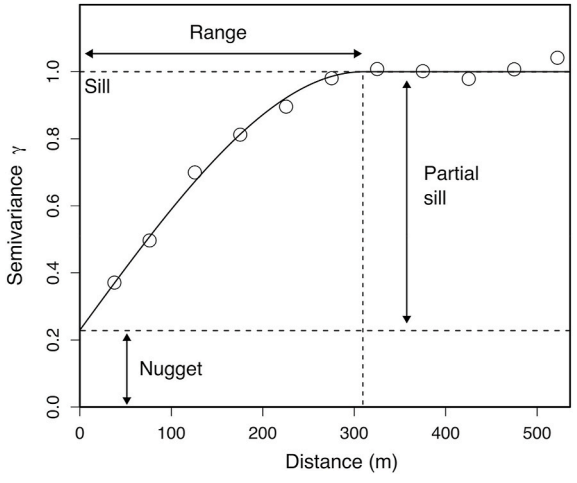


Fig. 1. Empirical variogram (circles) and spherical variogram model (solid line) fitted through weighted least squares. The variogram model is composed of a nugget, sill, and range. The nugget is the semivariance at zero distance, accounting for measurement errors or spatial variation at scales smaller than the sampling distance. The range is the distance beyond which observations are no longer correlated. The sill reflects the total variance, where the spatial correlation ceases–typically at the range. In this example, the nugget and sill are 0.23 and 1, respectively, and the range is 310 m.

3. sGs UnMix data processing

sGs UnMix can be used online at https://apps.bo.ingv.it/sgs-unmix. It consists of five tab panels: *Load data*, *Data*, *Mix model*, *Variogram*, and *sGs*. The *Data* panel shows data in a table on the left and a summary and histogram of a selected variable on the right. The remaining four panels consist of a left sidebar for inputs and a large main area for outputs, such as figures and tables. The sidebars contain several fields and a *Help* button showing a brief description to guide the user.

3.1. Load data tab panel

In the sidebar of this panel (Fig. 2), the user needs to upload a .csv file with coordinates (longitude and latitude) and a continuous variable, separated by comma, semicolon, or tab. The choice of separator needs to be made before uploading the file. To ensure user privacy, all uploaded data are processed locally and are not stored on any server; data are automatically discarded when the session ends or the browser is refreshed. Then, through the *Select columns* field, the user must specify the longitude, latitude, and the continuous variable (in this order) used for spatial prediction. The geodetic datum of the coordinate system must be specified as an EPSG code (available at https://epsg.io). If longitude and latitude are decimal WGS84, only the *Output EPSG* field has to be filled with the corresponding EPSG code. Otherwise, if the coordinates are not decimal WGS84, only the *Input EPSG* field must be compiled. The spatial points are plotted on a GIS layer (Esri World Imagery by default) and colored with a viridis gradient to have a first glance at the spatial variability of the data (Fig. 2a). The color scale can be switched from logarithmic to raw data through a radio button in the panel on the right side of the map, along with others color palette. The user can navigate into the interactive world map and choose the preferred GIS-layer between ESRI World Imagery and OpenStreetMap through a layer selection button on the left side of the map. In addition, one can measure distances or areas through the rectangle above the layer selection on the left side of the map. Last, the user has to draw a perimeter on the map to enclose the data to process (Fig. 2b). This can be done by drawing a rectangle or a polygon through the buttons below the layer selection. Alternatively, a .csv with the vertices of a polygon can be uploaded (and downloaded). The *Remove polygon* button can remove the uploaded polygon to define a new perimeter.

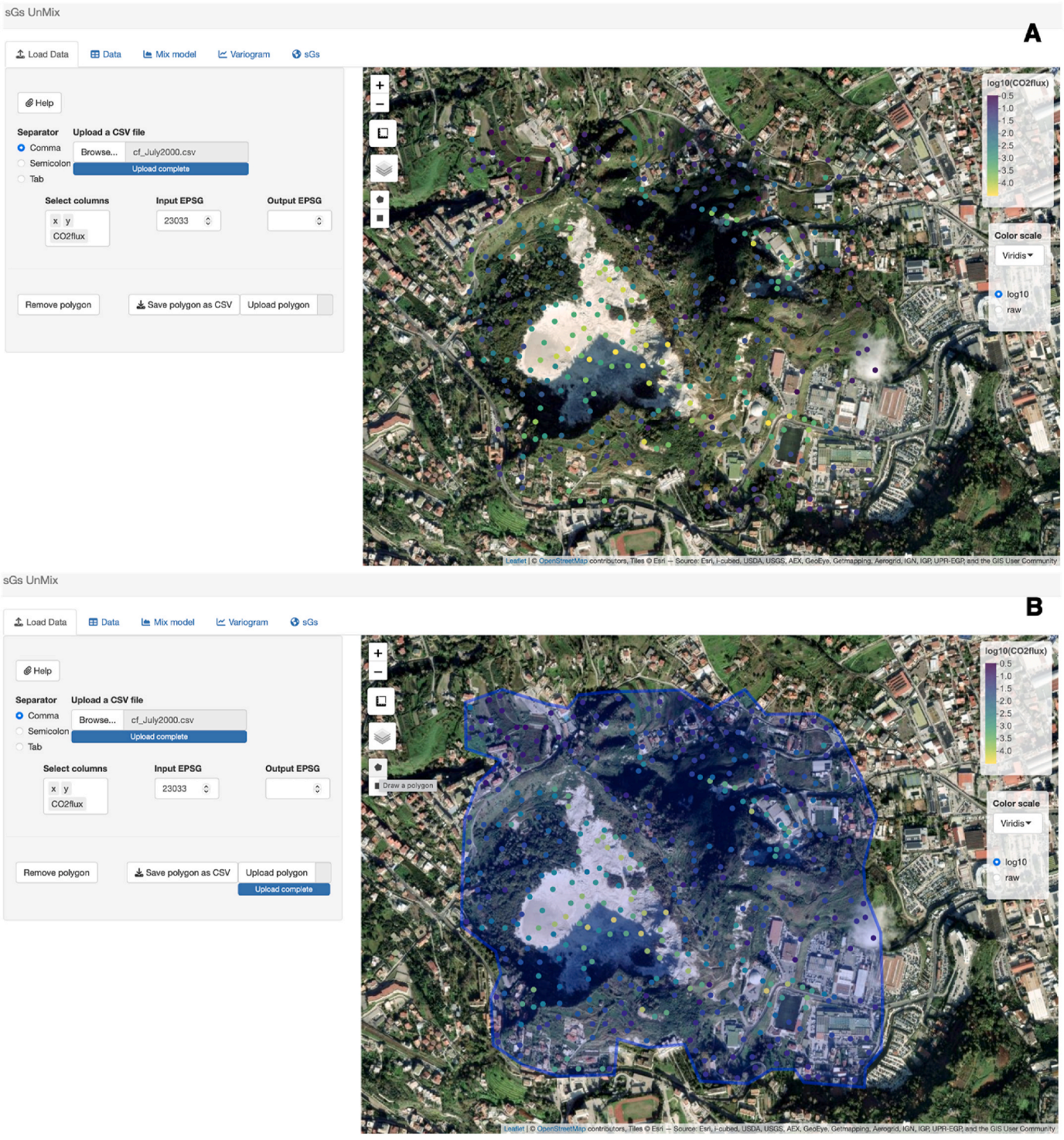


Fig. 2. Screenshot of the *Load data* tab panel of sGs UnMix showing (a) soil CO_2 fluxes (in g m^{-2} d^{-1}) from Campi Flegrei (Italy) over the ESRI World Imagery layer and (b) the polygon drawn to enclose data for spatial analyses.

3.2. Mix model tab panel

This panel (Fig. 3) enables the total density distribution of the data to be unmixed into two or more individual populations, which could reflect different data sources. By default, the sidebar panel is set on an *automatic* fitting procedure, that is sGs UnMix considers the distribution of the data as made by two lognormal populations and their means, standard deviations, and proportions are calculated using the EM-algorithm by maximizing the log-likelihood (*section 2.1*). Data distribution can be changed from lognormal to normal, enabling one to work with negative values. The user can change the *Number of populations* in *automatic* mode or can decide to manually fit the distribution by switching to *Manual* fit and specifying the parameters defining each individual population in the *Mean* and *St. dev.* fields, and pressing the *Calculate populations* button. Alternatively, the user can specify initial parameters using the *Initial Guess* mode, which is then accommodated by the web app to find the best solution. The log-likelihood output in the sidebar panel provides a measure of the goodness of fit, whose higher the value, the better the fit.

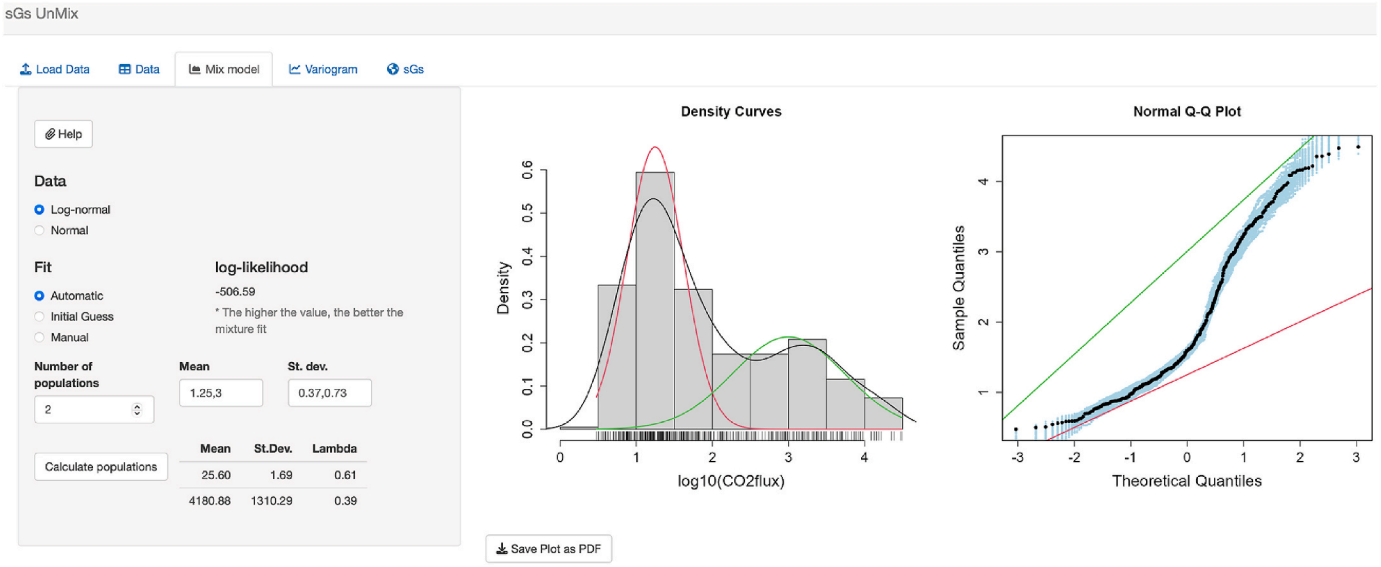


Fig. 3. Screenshot of the *Mix model* tab panel of sGs UnMix showing the unmixing of the data enclosed within the polygon (Fig. 2b) into two lognormal populations. These two populations returned from the automatic unmixing mode are plotted in density and Q-Q plots, and their means and standard deviations are plotted at the bottom of the sidebar. *Lambda* refers to the proportion of each population contributing to the total data distribution.

To help select the proper number of populations while avoiding overfitting, the user can calculate information criteria, such as the Akaike Information Criterion (Akaike, 1973). The parameters defined for each population are printed at the bottom of the sidebar and the results are displayed in the main area of the tab panel as density and Q-Q (Quantile-Quantile) plots. The Q-Q plot is a graphical tool used to compare the distribution of a dataset to a theoretical distribution. A single normal distribution appears as a 45° line in the Q-Q plot, whereas *n* normal populations result in *n* lines combined by *n–1* inflection points. Both density and Q-Q plots can be downloaded as PDF by selecting the *Save Plot as PDF* button.

3.3. Variogram tab panel

The spatial continuity of the transformed data is modeled in this panel (Fig. 4), which is divided in two subpanels for modeling both omnidirectional (Fig. 4a) and anisotropic variograms (Fig. 4b). By default, the variogram is omnidirectional and calculated using a spherical model through an autofitting procedure, which is based on the minimization of a weighted sum of square residuals (section 2.2). Results are displayed as a table and variogram and variogram cloud in the main area of the tab panel (Fig. 4a). The user is required to choose a model from the *variogram model* field (see Figure A.1 and Table A.1 for a list of the available models) that minimizes the weighted sum of squared errors (SSE, printed in the sidebar), which provides a measure of the goodness of fit. The automatic lag distance is calculated using a formula depending on the density of the data over the grid and its geometry (principal diagonal of the polygon divided by 45; Pebesma, 2004) and is printed in the sidebar panel. The app uses it as a starting input. Then, the average distance at which semivariances are calculated and the respective number of pairs (*N(h)*) used are displayed in the table on the right side of the main panel. If the number of pairs (*N(h)*) for the lag distance (*h*) is small (*N(h)* > 30 is recommended; Journel and Huijbregts, 1978), the user can manually insert it in the *Lag distance* field while remaining in autofit mode. If outliers exist within the data set, the variogram cloud shows data clusters with high semivariance at short distances. These outliers would drastically raise the nugget of the variogram. In such a case, the user is advised to enable the *Robust* variogram field or manually remove the outliers (see Example #2 in the Supplementary Material). Finally, the user can manually fit the variogram by switching from the *Automatic* to *Manual* fit field and adjusting the structure of the variogram by providing numeric inputs into the *Nugget*, *Partial sill*, and *Range* fields in the sidebar. The variogram plots are reactive to changes in the input of the variogram parameters, updating the empirical variogram fit in real-time. In addition to modeling a single variogram, the app offers a *Nested* option that allows users to model data structures resulting from the sum of two variograms. As for the *Single* selection, the *Nested* option supports both *Automatic* and *Manual* modes. Both table and variogram plot can be downloaded as .csv and PDF, respectively, by clicking the *Save Table as CSV* and *Save Plot as PDF* buttons.

The *anisotropic* subpanel (Fig. 4b) allows users to model directional variograms in manual mode. By default, N (0°), NE (45°), E (90°), and SE (135°) directions are selected. A variogram map is plotted in the main panel to help identify preferential directions, which can then be entered into the *Direction* input field (separated by commas). Variograms computed along these four directions (using an angle tolerance of 22.5°) are displayed below the variogram map. Users must specify the principal direction (*Main dir*), defined as the direction with the largest range, and the anisotropy ratio (*Ratio*), calculated as the ratio between the range in the direction perpendicular to the principal direction (minor range) and the range in the principal direction (major range). By default, *Main Dir* is set to 45° and *Ratio* is set to 0.4. Finally, the variogram parameters—nugget, partial sill, and range (defined with respect to the main direction)—has to be adjusted to minimize the Sum of Squared Errors (SSE).

An example of directional variogram modeling is provided in the Supplementary Material (Example #4). This example also demonstrates the modeling of nested variograms.

3.4. sGs tab panel

This panel enables the user to perform spatial prediction through sequential Gaussian simulation (sGs; section 2.3; Fig. 5). The sidebar of this panel shows a summary of the grid extent in m (*X min, X max, Y min,* and *Y* max fields), the extent of the cell (in m) over which perform sGs (*Delta X* and *Delta Y* fields), and the *Number of simulations* field. By default, the cells have 10 × 10 m spacing, and the algorithm performs 100 sGs, but the user can adjust these values as desired. Once these parameters have been set up, the simulations start by clicking the *Run sGs* button. After a few seconds, the result is plotted in the main panel as a heat map of the mean values of the *n* number of simulations,

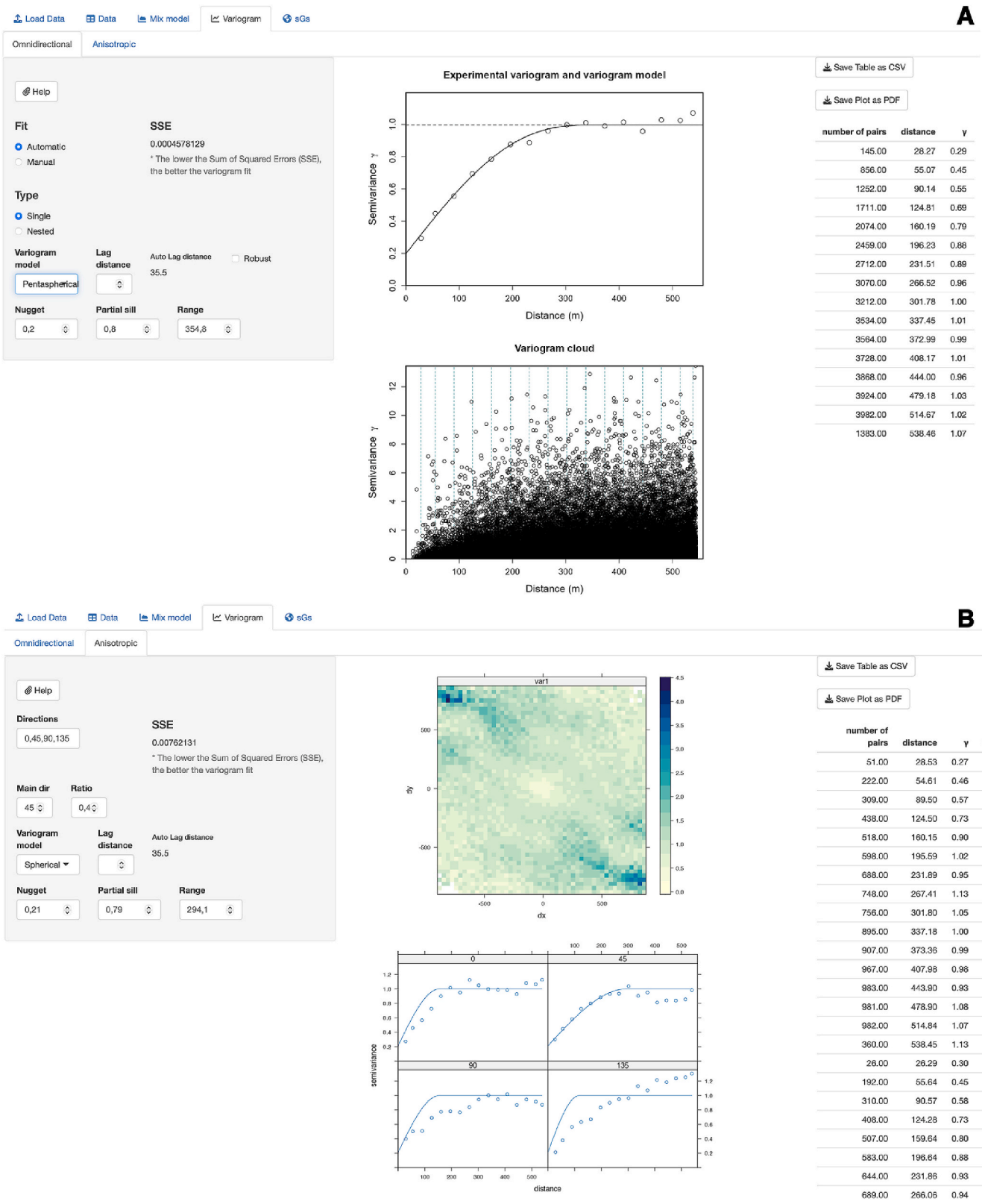


Fig. 4. Screenshot of the *Variogram* tab panel of sGs UnMix showing the (a) omnidirectional and (b) anisotropic subpanles. (a) The omnidirectional subpanel shows the variogram and variogram cloud of Campi Flegrei soil CO_2 fluxes, automatically fitted with a pentaspherical model through weighted least squares. (b) The anisotropic subpanel shows the variogram map and the variograms calculated along the N, SE, E, and SW directions (0°, 45°, 90°, and 135°).

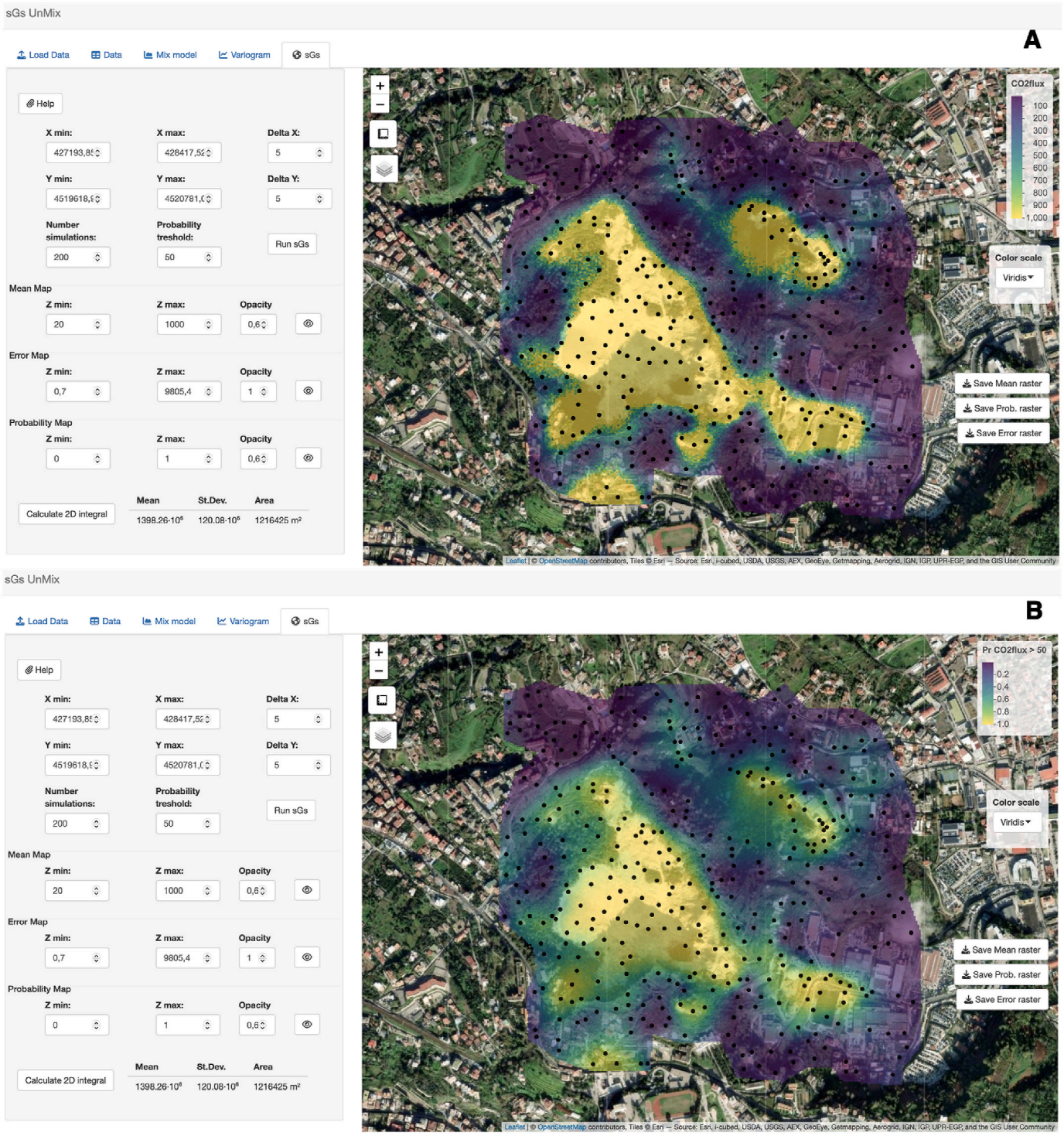


Fig. 5. Screenshot of the *sGs* tab panel of sGs UnMix showing the (a) mean and (b) probability heat maps of Campi Flegrei soil CO_2 fluxes returned by averaging 200 sGs. The total soil CO_2 emission from Campi Flegrei has been calculated through the *Calculate 2D integral* button at the bottom of the sidebar, accounting for 1398 ± 120 t d^{-1}.

overlaying a GIS-based layer (Esri World Imagery by default). The user can then change the color scale (on the right side of the map) or adjust its minimum and maximum values and the transparency of the heat map by adjusting the *Z min*, *Z max*, and *Opacity* fields in the sidebar and clicking the button with the eye icon. These adjustments can improve the localization of the spatial distribution of patterns and their relationship with the area morphology, helping to understand possible sources and the processes driving spatial anomalies. This panel also enables calculating the selected variable double integration over the area defined in the *Load data* panel and its uncertainty, which are printed at the bottom of the left side panel after pressing the *Calculate 2D integral* button. These values reflect the mean and standard deviation of the outputs estimated

for each *n* equiprobable realizations generated by sGs. In the case of soil CO_2 flux data, or any other input variable measured in g m^{-2} d^{-1}, sGs UnMix returns the total output and its standard deviation in g d^{-1}. In addition to a mean heat map, sGs UnMix calculates both error and probability heat maps. The error heat map shows the standard deviation calculated for each cell across the *n* number of simulations. The probability heat maps indicates the likelihood that the target variable exceeds a specific threshold, which must be entered in the *probability threshold* field before clicking the *run sGs* button. To help choose the threshold, the user can inspect the populations disentangled from the total data distribution in the *Mix model* tab. The user can visualize the error or probability heat maps by clicking the corresponding button with the eye icon and adjust the color scale and opacity as for the mean heat map. Finally, the *Save raster* button allows downloading mean, error, and probability heat maps as Tagged Image File Format (TIFF) for further elaborations on GIS software.

4. A case study: soil CO_2 emission from Campi Flegrei, Italy

The Neapolitan area of Italy is populated by about four million people and includes three active volcanoes: Campi Flegrei, Somma-Vesuvius, and Ischia. This volcanic area has shown activity for at least the past 50,000 years, characterized by two significant eruptive events forming a ~15-km-wide caldera (Acocella et al., 2015). Since 2006, Campi Flegrei has shown increased seismicity, ground deformation, and degassing, reflecting a phase of volcanic unrest (Chiodini et al., 2021b). Hence, the measurement of the soil CO_2 fluxes is fundamental to monitoring its activity (e.g., Cardellini et al., 2017) and has been part of the monitoring program of the Istituto Nazionale di Geofisica e Vulcanologia (INGV) since 1998.

In this example, we used the soil CO_2 fluxes measured during one of these campaigns in July 2000 (cf_july2000.csv in the Supplementary Material; from Cardellini et al., 2017) in the target area of Campi Flegrei, which includes the Solfatara di Pozzuoli, a tuff cone where the onshore degassing focuses. In the sidebar of the *Load data* panel (Fig. 2), we specified *comma* as separator, uploaded the cf_july2000.csv, and selected the *x*, *y*, and *CO2flux* variables of the data set. Since the coordinates are in WGS84 and derived from the ED50 system rather than decimal format, we provided the corresponding EPSG: 23033 in the input EPSG. After this configuration, the measured points are plotted on the ESRI World Imagery layer and color coded with a gradient from purple to yellow as the CO_2 flux increases (Fig. 2a). High flux measurements mainly concentrate in the Solfatara (grey area), with some also found in the northeast and southeast regions (Fig. 2a). Finally, we drew a polygon that tightly enclosed the measured data (Fig. 2b) over which the spatial prediction had to be performed.

The *Mix model* panel shows the distribution of the data enclosed within the polygon in density (black curve) and Q-Q plots (black circles; Fig. 3). The automatic fit reproduces this distribution through the mixing of two lognormal populations (red and green curves; Fig. 3). The low-CO_2 flux population (red curve) has an average value of 25.6 ± 1.7 g m^{-2} d^{-1}, contributing for the 61 % of the distribution (Fig. 3), whereas the high-CO_2 flux population (green curve) contributes for the 39 % of it, with an average value of 4181 ± 1310 g m^{-2} d^{-1}. Notably, these estimates closely match those obtained by Cardellini et al. (2017)–23.9 ± 1.50 g m^{-2} d^{-1} (61 %) and 4590 ± 1351 g m^{-2} d^{-1} (39 %)–by using a manual fitting based on the method of Sinclair (1974), demonstrating a strong concordance between the two approaches. The low-CO_2 flux population reflects the background emission from the soil, produced by biogenic activity, whereas the high-CO_2 flux population mainly reflect the volcanic CO_2 exsolved at depth from magmas. This step is fundamental to discriminating the sources of CO_2 and quantifying their average values (Chiodini et al., 1998), which can be used in the following steps to correct the volcanic output from nonvolcanic CO_2 contributions. In addition, the unmixing of populations helps us define the thresholds for the probability heat map calculation (Cardellini et al., 2003) and, hence, better constrain the spatial distribution of the sole volcanic CO_2 degassing. To this end, we can use as threshold 50 g m^{-2} d^{-1}, which corresponds approximately to the 90th percentile of the biogenic population. We can then calculate the probability of a value to be higher than this threshold, hence only considering the volcanic CO_2 emission – less than 10 % of the data belong to the biogenic population. It is important to underscore that there is not a standardized procedure to define thresholds above which calculating probability heat maps. Cardellini et al. (2003) indicated threshold values around the 90th percentile of the low-CO_2 flux population. However, the only effective method to quantify the actual biogenic CO_2 flux relies on the analyses of the C isotopes of the CO_2 (Chiodini et al., 2008; Bini et al., 2020; Viveiros et al., 2020). In fact, biological processes produce CO_2 with very negative values of C isotopes, which can be unequivocally distinguished from those produced by magmatic outgassing. In this respect, it is noteworthy that the biogenic threshold of 50 g m^{-2} d^{-1} defined here corresponds to that estimated by Chiodini et al. (2008) for Campi Flegrei using C isotope analyses of CO_2.

The variogram panel (Fig. 4a) displays the spatial continuity of the soil CO_2 flux data measured in July 2000, which is autofitted using a pentaspherical variogram model with nugget = 0.2, partial sill = 0.8, and range = 354.8 m. This variogram model is slightly different than that employed by Cardellini et al. (2017)–spherical model with nugget = 0.26, partial sill = 0.74, and range = 330 m–but it better fits the empirical variogram, that is, the SSE is lower. The number of pairs (145) calculated at the first lag distance (28.3 m) is sufficient to ensure a reliable estimate of the semivariance. Hence, we used the provided *Auto Lag distance* (35.5 m) without manually increasing it in the *Lag distance* field (Fig. 4a). The variogram cloud does not show any cluster at short distances (Fig. 4a)—distances less than the first lag—indicating potential outliers. In addition, the variogram map (Fig. 4b) does not exhibit clear anisotropy, suggesting that the omnidirectional variogram is an appropriate choice. Therefore, we proceed to the spatial prediction step.

To predict soil CO_2 fluxes at unsampled locations of Campi Flegrei, we used cells of 5 × 5 m (*Delta X* and *Delta Y*) and 200 simulations (Fig. 5). In addition, we set the probability threshold at 50 g m^{-2} d^{-1}, as estimated in the *Mix model* panel. Fig. 5a shows the map of the predicted soil CO_2 fluxes over the ESRI World Imagery layer, where we set a color scale ranging from 20 to 1000 g m^{-2} d^{-1} (*Z min and Z max*) and a transparency (*Opacity*) of 0.6. The total emission of CO_2 calculated by sGs UnMix accounted for 1398 ± 120 t d^{-1} (from 1,22 km^2) in July 2000 and most of the degassing spatially distributed over the Solfatara area (Fig. 5a). Two other anomalies are also present in the southeast and northeast areas of the map (Fig. 5a), in correspondence of the Pisciarelli area, which, as Solfatara, is characterized by emission of hydrothermal fluids from argillic, altered soil through diffusion and as fumarolic vents. Both spatial distribution and total output of the soil CO_2 emission estimated with sGs UnMix are practically identical to those reported by Cardellini et al. (2017). By adopting the parameters used by these authors–square polygon (cf_July2000_polygon_Cardellini_et_al_2017.csv), spherical variogram model with nugget = 0.26, partial sill = 0.74, and range = 330 m, and 100 sGs over 10 × 10 m cells–sGs UnMix yielded an estimate of 1501 ± 120 t d^{-1}, compared to 1513 ± 146 t d^{-1} reported by Cardellini et al. (2017). To better understand the spatial distribution of the volcanic outgassing and better constrain the extent of different structures, we can switch to the probability heat map visualization (Fig. 5b). In this case, Fig. 5b shows the probability of soil CO_2 flux higher than 50 g m^{-2} d^{-1} (from 0 to 1), that is the threshold defined using the 90th percentile of the biogenic population defined in the *Mix model* panel.

To estimate the volcanic emission of CO_2, we multiplied the average value of the biogenic population of soil CO_2 flux data (25.6 g m^{-2} d^{-1})

by the area of the grid (1,216,425 m^2), and subtracted it (31.1 t d^{-1}) from the total emission, resulting in 1367 ± 120 t d^{-1}. Finally, we downloaded the raster from sGs UnMix. By processing rasters in GIS softwares and incorporating structural features like faults and craters, one can investigate the control of these on the soil CO_2 flux (e.g. Fig. 4 in Cardellini et al., 2017).

Further examples of spatial data processing with sGs UnMix are reported in the Supplementary Material. The first example shows a case of zinc pollution in the topsoil of a flood plain of the river Meuse, using the meuse dataset of the sp package for R (Pebesma and Bivand, 2005; Bivand et al., 2013). The second and third examples illustrate the processing of other soil CO_2 flux datasets. In particular, the Krafla case, Iceland (Bini et al., 2024), shows an example of how to deal with the presence of outliers by choosing robust variograms or removing them and how to manually unmix data populations. The case of the Nisyros caldera, Greece (Bini et al., 2019), shows how to search for different populations of data when the total distribution appears to follow only one lognormal distribution. In addition, this case provides a good spatial distribution of different structures of volcanic degassing that contribute to the total CO_2 emission. The fourth example shows heat maps of CO_2 soil emission from Latera, Italy (Chiodini et al., 2007), constructed using both nested and anisotropic variogram models.

5. Discussion

Through a straightforward four-step procedure, we developed an intuitive application to process spatial data and predict their distribution at unsampled locations across space. These steps are implemented in four interactive panels, which, helped by detailed descriptions, guide users through loading datasets and reprojecting coordinates, separating different data populations, modeling spatial continuity through variograms, and performing spatial predictions using sequential Gaussian simulations (sGs). Each panel has a left sidebar for input parameters and a central display area for output visualizations, such as plots and tables, which face the user to a user-friendly, interactive graphical interface. Tables and plots, such as variograms and maps of predicted values overlaid on GIS-layers, dynamically update in response to changes in input values. Additionally, the application allows users to download tables, high-resolution graphs suitable for scientific publications, and raster files compatible with GIS software like QGIS (https://qgis.org).

A key advantage of sGs UnMix is its automated approach to data population separation and variogram fitting. This implementation not only reduces the users' subjectivity in statistical analysis, such as determining data sources and thresholds or mapping the spatial distribution of continuous variables, but also enhances the reproducibility of results. However, in some cases—such as when some populations are not automatically detected (see Example #2 in the Supplementary Material) or when variogram fitting overestimates data variance (i.e., when the sill exceeds 1)—it may be beneficial to provide users with a higher degree of control. This can be easily accomplished with sGs UnMix by switching from *automatic* to *manual* mode in the *Mix model* and *Variogram* panels.

The most significant advantage of sGs UnMix is that it does not require programming skills, software installation, or a high-performance computer, making it accessible to the broader geoscience community through an internet connection. This web app is a valuable alternative for variogram modeling and spatial prediction, comparable to other available codes such as FORTRAN (GSLIB; Deutsch and Journel, 1998), Python (GeostatsPy; Pyrcz et al., 2021; PyGSLIB; https://opengeostat.github.io/pygslib/; GSTools; Müller et al., 2022), and R (geoR; Ribeiro and Diggle, 2001). Although some GUIs exist, such as WinGslib (http://www.statios.com/WinGslib/) and SGeMS (Remy et al., 2009), these often involve numerous manual steps from data processing to visualization, making spatial prediction time-consuming and challenging for those without extensive knowledge of geostatistics. A lack of automated features, like variogram auto-fitting, may introduce user subjectivity and reduce accuracy. The ease of use, reactivity, interactivity, and automated features of sGs UnMix minimize user bias and significantly reduce operational complexity while maintaining accuracy, thus reflecting a valuable alternative to these programs. In particular, the app's reactive outputs, such as maps that update instantly with input changes, enable users to quickly identify spatial patterns and investigate anomalies at multiple scales through interactive (rectangular and polygonal) data selection. Additionally, sGs UnMix facilitates the automated separation of data groups with the *Mix model*, which is typically lacking in current geostatistical tools. This feature enables users to set thresholds for detecting anomalies in spatial geochemical data distributions.

sGs UnMix facilitates spatial data analysis and prediction in multiple geoscience fields. In volcanology, it provides a consistent approach for assessing and comparing volatile emissions from volcanoes locally and globally. By simplifying the application of sGs, the web app eliminates the need for alternative interpolation techniques, such as inverse distance weighting and kriging, which can underestimate volatile fluxes. At the local scale, the web app can be applied not only to investigate the degassing and its structural control, such as faults, but also to monitor their variations over time for volcanic surveillance (e.g., Cardellini et al., 2017). Locally, sGs UnMix can also support geothermal exploration by coupling soil CO_2 emissions and temperature maps to estimate heat flow (e.g., Chiodini et al., 2005, 2021a). The app standardized approach allows for comparable volatile flux estimates across volcanoes globally, contributing to more reliable regional and global outflux assessments. This is critical for understanding the relationship between Earth degassing and climate change (e.g., Lee et al., 2016; Brune et al., 2017) and for mitigating environmental risks. Beyond volcanology and climate remediation, the app can find many other important applications to diverse fields, such as studying the distribution of ore deposits, hydrocarbons, and pollutants in soils and freshwater systems, facilitating targeted interventions.

6. Conclusions

We presented sGs UnMix, a web application written in shiny for R (freely available online at https://apps.bo.ingv.it/sgs-unmix) designed to make spatial data analysis and prediction accessible to a broad geoscience audience. Its user-friendly interface, automated workflows, and interactive outputs reduce user bias and technical complexity, supporting reproducible results. sGs UnMix could be used as a standardized method for studying volcanic volatile fluxes (e.g., soil CO_2 emissions), but could also be applied across diverse geoscience fields, including environmental monitoring, climate research, and resource exploration. Although sGs UnMix has been shaped with a straightforward structure and interface, it offers potential for future enhancements to meet the needs of advanced users. The source code and future updates will be available on its Git repository (https://github.com/giancarlotamburello/sGs_UnMix.git) ensuring ongoing development and accessibility.

CRediT authorship contribution statement

Giulio Bini: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Giancarlo Tamburello:** Writing – review & editing, Visualization, Software, Methodology, Conceptualization. **Stefano Cacciaguerra:** Software. **Paolo Perfetti:** Software.

Code availability

sGs UnMix is available online at: https://apps.bo.ingv.it/sgs-unmix.

The source codes are available for downloading at: https://github.com/giancarlotamburello/sGs_UnMix.

Contact: giulio.bini@ingv.it.

Program language: R.

Declaration of generative AI in scientific writing

The authors used ChatGPT to enhance readability and language clarity in preparing this work. Following this tool, they carefully reviewed and edited the content as necessary, assuming full responsibility for the publication's content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Giovanni Chiodini (INGV-OV, Napoli, Italy) for useful discussions during the development of the web application. G. B. was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 858092 (IMPROVE project). Finally, we thank two anonymous reviewers for their constructive comments.

Appendix. ASupplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2025.106652.

Appendix A. Variogram models used in the sGs UnMix

sGs UnMix offers seven variogram models to fit the empirical semivariogram, that is spherical, exponential, pentaspherical, gaussian, circular, linear, and bessel (Pebesma et al., 2004). These models and their equations are reported in Figure A.1 and Table A.1.

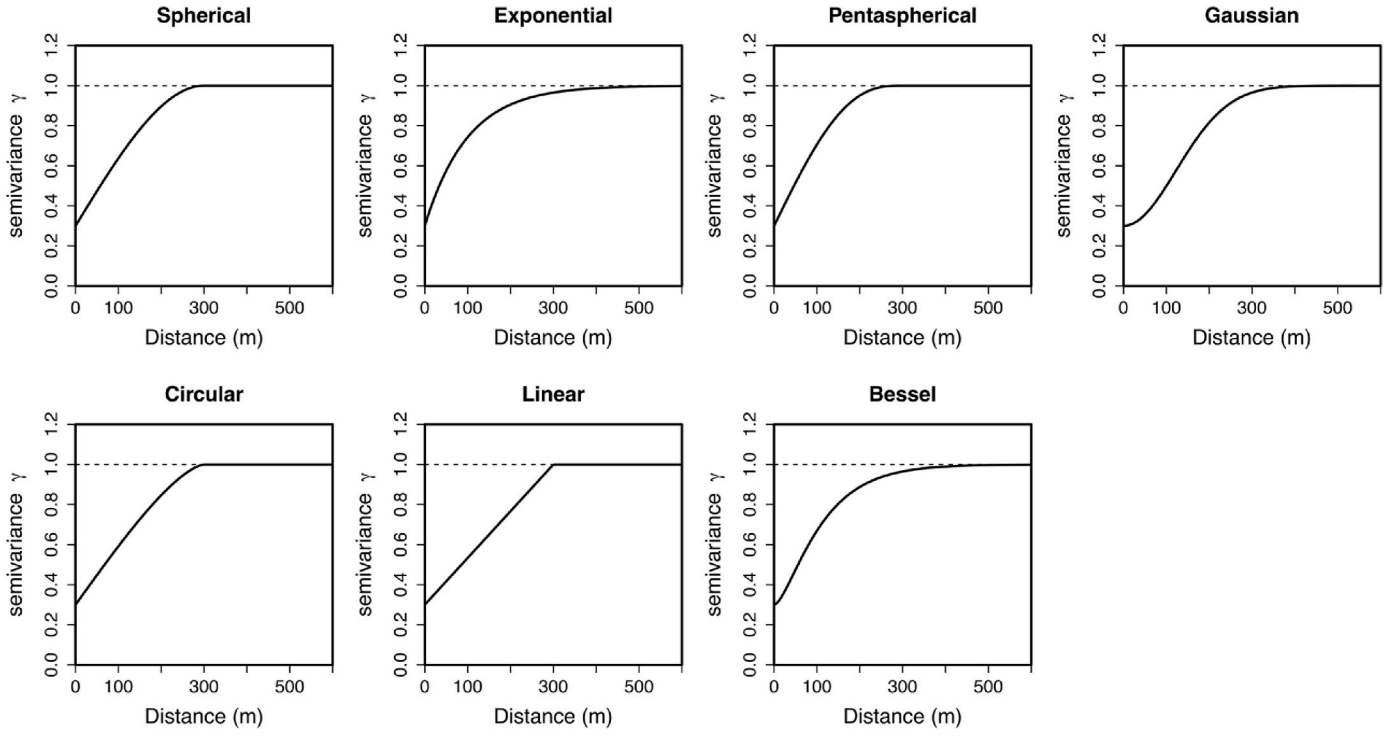


Figure A.1. Variogram models with nugget = 0.3, partial sill = 0.7, and range = 300 m

Table A.1
variogram models and their equations (after Pebesma et al., 2014) included in sGs UnMix. C_0, C, h, and a refer to the variogram nugget, sill, lag distance, and range, respectively.

Model	Equation
Spherical	$\gamma(h) = C_0 + (C - C_0)\left[\frac{3h}{2a} - \frac{h^3}{2a^3}\right],\ if\ h \le a$
	$\gamma(h) = C_0 + (C - C_0),\ if\ h \ge a$
Exponential	$\gamma(h) = C_0 + (C - C_0)\left[1 - exp\left(-\frac{h}{a}\right)\right]$
Pentaspherical	$\gamma(h) = C_0 + (C - C_0)\left[\frac{15h}{8a} - \frac{5h^3}{4a^3} + \frac{3h^5}{8a^5}\right],\ if\ h \le a$
	$\gamma(h) = C_0 + (C - C_0),\ if\ h \ge a$
Gaussian	$\gamma(h) = C_0 + (C - C_0)\left[1 - exp\left(-\left(\frac{h}{a}\right)\right)^2\right]$
Circular	$\gamma(h) = C_0 + (C - C_0)\left[\frac{2h}{\pi a}\sqrt{1 - \left(\frac{h}{a}\right)^2} + \frac{2}{\pi} arcsin\frac{h}{a}\right],\ if\ h \le a$
	$\gamma(h) = C_0 + (C - C_0),\ if\ h \ge a$
Bessel	$\gamma(h) = C_0 + (C - C_0)\left[1 - \frac{h}{a} J_1\left(\frac{h}{a}\right)\right]$, where J_1 is the Bessel function of the first kind
Linear	$\gamma(h) = C_0 + bh$, where b is the slope

Data availability

Test data are available at https://github.com/giancarlotamburello/sGs_UnMix

References

Acocella, V., Di Lorenzo, R., Newhall, C., Scandone, R., 2015. An overview of recent (1988 to 2014) caldera unrest: knowledge and perspectives. Rev. Geophys. 53 (3), 896–955.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S., 2009. Mixtools: an R package for analyzing mixture models. J. Stat. Software 32, 1–29.

Bini, G., Chiodini, G., Cardellini, C., Vougioukalakis, G.E., Bachmann, O., 2019. Diffuse emission of CO_2 and convective heat release at nisyros caldera (greece). J. Volcanol. Geoth. Res. 376, 44–53.

Bini, G., Chiodini, G., Lucchetti, C., Moschini, P., Caliro, S., Mollo, S., Selva, J., Tuccimei, P., Galli, G., Bachmann, O., 2020. Deep versus shallow sources of CO_2 and Rn from a multi-parametric approach: the case of the nisyros caldera (aegean arc, Greece). Sci. Rep. 10 (1), 13782.

Bini, G., Chiodini, G., Ricci, T., Sciarra, A., Caliro, S., Mortensen, A.K., Martini, M., Mitchell, A., Santi, A., Costa, A., 2024. Soil CO_2 emission and stable isotopes ($\delta^{13}C$, $\delta^{18}O$) of CO_2 and calcites reveal the fluid origin and thermal energy in the supercritical geothermal system of krafla, Iceland. J. Volcanol. Geoth. Res. 447, 108032.

Bivand, R., Pebesma, E., Gomez-Rubio, V., 2013. In: Applied Spatial Data Analysis with R, second ed. Springer, New York, NY, p. 405 https://asdar-book.org/.

Brune, S., Williams, S.E., Müller, R.D., 2017. Potential links between Continental rifting, CO_2 degassing and climate change through time. Nat. Geosci. 10 (12), 941–946.

Carapezza, M.L., Barberi, F., Ranaldi, M., Ricci, T., Tarchini, L., Barrancos, J., Fischer, C., Perez, N., Weber, K., Di Piazza, A., Gattuso, A., 2011. Diffuse CO_2 soil degassing and CO_2 and H_2S concentrations in air and related hazards at vulcano island (aeolian arc, Italy). J. Volcanol. Geoth. Res. 207 (3–4), 130–144.

Cardellini, C., Chiodini, G., Frondini, F., 2003. Application of stochastic simulation to CO_2 flux from soil: mapping and quantification of gas release. J. Geophys. Res. Solid Earth 108 (B9).

Cardellini, C., Chiodini, G., Frondini, F., Avino, R., Bagnato, E., Caliro, S., Lelli, M., Rosiello, A., 2017. Monitoring diffuse volcanic degassing during volcanic unrests: the case of campi flegrei (italy). Sci. Rep. 7, 6757.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B.. Shiny: web application framework for R R package version 1.10.0. https://CRAN.R-project.org/package=shiny.

Cheng, J., Schloerke, B., Karambelkar, B., Xie, Y., 2024. Leaflet: create interactive web maps with the JavaScript 'leaflet' library. R package version 2 (2.2). https://CRAN.R-project.org/package=leaflet.

Chiodini, G., Cioni, R., Guidi, M., Raco, B., Marini, L., 1998. Soil CO_2 flux measurements in volcanic and geothermal areas. Appl. Geochem. 13 (5), 543–552.

Chiodini, G., Granieri, D., Avino, R., Caliro, S., Costa, A., Werner, C., 2005. Carbon dioxide diffuse degassing and estimation of heat release from volcanic and hydrothermal systems. J. Geophys. Res. Solid Earth 110 (B8).

Chiodini, G., Baldini, A., Barberi, F., Carapezza, M.L., Cardellini, C., Frondini, F., Granieri, D., Ranaldi, M., 2007. Carbon dioxide degassing at latera caldera (italy): evidence of geothermal reservoir and evaluation of its potential energy. J. Geophys. Res. Solid Earth 112 (B12).

Chiodini, G., Caliro, S., Cardellini, C., Avino, R., Granieri, D., Schmidt, A., 2008. Carbon isotopic composition of soil CO_2 efflux, a powerful method to discriminate different sources feeding soil CO_2 degassing in volcanic-hydrothermal areas. Earth Planet Sci. Lett. 274 (3–4), 372–379.

Chiodini, G., Cardellini, C., Lamberti, M.C., Agusto, M., Caselli, A., Liccioli, C., Tamburello, G., Tassi, F., Vaselli, O., Caliro, S., 2015. Carbon dioxide diffuse emission and thermal energy release from hydrothermal systems at copahue–caviahue volcanic complex (argentina). J. Volcanol. Geoth. Res. 304, 294–303.

Chiodini, G., Cardellini, C., Bini, G., Frondini, F., Caliro, S., Ricci, L., Lucidi, B., 2021a. The carbon dioxide emission as indicator of the geothermal heat flow: review of local and regional applications with a special focus on Italy. Energies 14 (20), 6590.

Chiodini, G., Caliro, S., Avino, R., Bini, G., Giudicepietro, F., De Cesare, W., Ricciolino, P., Aiuppa, A., Cardellini, C., Petrillo, Z., Selva, J., Siniscalchi, A., Tripaldi, S., 2021b. Hydrothermal pressure-temperature control on CO_2 emissions and seismicity at campi flegrei (italy). J. Volcanol. Geoth. Res. 414, 107245.

Cressie, N., 1993. Statistics for Spatial Data. John Wiley & Sons, New York, NY, USA.

Delbari, M., Afrasiab, P., Loiskandl, W., 2009. Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. Catena 79 (2), 163–169.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and Users Guide. Oxford University Press, New York, p. 369.

Fischer, T.P., Arellano, S., Carn, S., Aiuppa, A., Galle, B., Allard, P., Lopez, T., Shinohara, H., Kelly, P., Werner, C., Cardellini, C., Chiodini, G., 2019. The emissions of CO_2 and other volatiles from the world's subaerial volcanoes. Sci. Rep. 9, 18716.

Gatscha, S., Karambelkar, B., Schloerke, B., 2024. Leaflet.extras: extra functionality for 'leaflet' package. R package version 2.0.1. https://CRAN.R-project.org/package=leaflet.extras.

Goovaerts, P., 1997. In: Geostatistics for Natural Resources Evaluation, vol. 483. Oxford University Press.

Gurrieri, S., Di Martino, R.M.R., Camarda, M., Francofonte, V., 2023. Monitoring CO_2 hazards of volcanic origin: a case study at the island of vulcano (italy) during 2021–2022. Geosciences 13 (9), 266.

Hijmans, R., 2024a. Raster: geographic data analysis and modeling. R package version 3, 6–30. https://CRAN.R-project.org/package=raster.

Hijmans, R., 2024b. Terra: spatial data analysis. R package version 1, 8-5. https://CRAN.R-project.org/package=terra.

Huang, L., 2023. Leaflet.providers: leaflet providers. R package version 2.0.0. https://CRAN.R-project.org/package=leaflet.providers.

Journal, A.G., Alabert, F., 1989. Non gaussian data expansion in the Earth sciences. Terra Nova 1 (2), 123–134.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. J. S. Afr. Inst. Min. Metall 52 (6), 119–139.

Lee, H., Muirhead, J.D., Fischer, T.P., Ebinger, C.J., Kattenhorn, S.A., Sharp, Z.D., Kianji, G., 2016. Massive and prolonged deep carbon emissions associated with Continental rifting. Nat. Geosci. 9 (2), 145–149.

Lewicki, J.L., Bergfeld, D., Cardellini, C., Chiodini, G., Granieri, D., Varley, N., Werner, C., 2005. Comparative soil CO_2 flux measurements and geostatistical estimation methods on Masaya Volcano, Nicaragua. Bull. Volcanol. 68 (1), 76–90.

Matheron, G., 1962. In: Traité De Géostatistique Appliquée, vol. 1. Editions Technip, Paris.
Matheron, G., 1963a. In: Traité De Géostatistique Appliquée, vol. 2. Le krigeage. Editions Technip, Paris.
Matheron, G., 1963b. Principles of geostatistics. Econ. Geol. 58 (8), 1246–1266.
Müller, S., Schüler, L., Zech, A., Heße, F., 2022. GSTools v1. 3: a toolbox for geostatistical modelling in python. Geosci. Model Dev. (GMD) 15 (7), 3161–3182.
Nychka, D., Furrer, R., Paige, J., Sain, S., 2021. Fields: tools for spatial data. R package version 16.3,. https://github.com/dnychka/fieldsRPackage.
Özen, S.A., Yesilkanat, C.M., Özen, M., Başsarı, A., Taşkın, H., 2022. Health risk assessment of soil trace elements using the sequential gaussian simulation approach. Environ. Sci. Pollut. Control Ser. 29 (48), 72683–72698.
Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.
Pebesma, E.J., 2014. Gstat user's manual. https://www.gstat.org/gstat.pdf.
Pebesma, E., Bivand, R., 2005. Classes and methods for spatial data in R. R. News 5 (2), 9–13. https://CRAN.R-project.org/doc/Rnews/.
Pyrcz, M.J., Deutsch, C.V., 2014. Geostatistical reservoir modeling. Oxford University Press, New York, p. 433.
Pyrcz, M.J., Jo, H., Kupenko, A., Liu, W., Gigliotti, A.E., Salomaki, T., Santos, J., 2021. Geostatspy Python Package: Open-Source Spatial Data Analytics and Geostatistics. https://doi.org/10.5281/zenodo.13835444.
Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics with Sgems: a User's Guide. Cambridge University Press.
Ribeiro Jr, P.J., Diggle, P.J., 2001. geoR: a package for geostatistical analysis. R. News 1 (2), 14–18.
Rowlingson, B., Diggle, P., 2024. Splancs: spatial and space-time point pattern analysis. R package version 2.01-45. https://CRAN.R-project.org/package=splancs.
Sinclair, A.J., 1974. Selection of threshold values in geo- chemical data using probability graphs. J. Geochem. Explor. 3, 129–149.
Sun, X.L., Wu, S.C., Wang, H.L., Zhao, Y.G., Zhang, G.L., Man, Y.B., Wong, M.H., 2013. Dealing with spatial outliers and mapping uncertainty for evaluating the effects of urbanization on soil: a case study of soil pH and particle fractions in Hong Kong. Geoderma 195, 220–233.
Tarchini, L., Ranaldi, M., Carapezza, M.L., Di Giuseppe, M.G., Isaia, R., Lucchetti, C., Prinzi, E.P., D'Assisi Tramparulo, F., Troiano, A., Vitale, S., 2019. Multidisciplinary studies of diffuse soil CO_2 flux, gas permeability, self-potential, soil temperature highlight the structural architecture of Fondi di Baia craters (Campi Flegrei caldera, Italy). Ann. Geophys. 62 (1), 1–12.
Tardani, D., Taussi, M., Robidoux, P., Sánchez-Alfaro, P., Pérez-Flores, P., Serrano, G., Morales, G., Tassara, S., Grassa, F., Soler, V., Morata, D., 2024. Gas geothermometry, soil CO_2 degassing, and heat release estimation to assess the geothermal potential of the alpehue hydrothermal field (sollipulli volcano, southern Chile). Geothermics 122, 103092.
Taussi, M., Nisi, B., Brogi, A., Liotta, D., Zucchi, M., Venturi, S., et al., 2023. Deep regional fluid pathways in an extensional setting: the role of transfer zones in the hot and cold degassing areas of the larderello geothermal system (northern apennines, Italy). G-cubed 24, e2022GC010838.
Viveiros, F., Chiodini, G., Cardellini, C., Caliro, S., Zanon, V., Silva, C., Rizzo, A.L., Hipólito, A., Moreno, L., 2020. Deep CO_2 emitted at furnas do enxofre geothermal area (terceira island, azores archipelago). An approach for determining CO_2 sources and total emissions using carbon isotopic data. J. Volcanol. Geoth. Res. 401, 106968.
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons.
Werner, C., Fischer, T.P., Aiuppa, A., Edmonds, M., Cardellini, C., Carn, S., Chiodini, G., Cottrell, E., Burton, M., Shinohara, H., 2019. Carbon dioxide emissions from subaerial volcanic regions: two decades in review. In: Orcutt, B., Dasgupta, R., Daniel, I. (Eds.), Deep Carbon: past to Present. Cambridge University Press, pp. 188–236.
Xie, Y., Cheng, J., Tan, X., 2024. DT: a wrapper of the JavaScript library 'DataTables'. R package version 0.33. https://CRAN.R-project.org/package=DT.