A multivariate non-parametric approach for estimating probability of exceeding the local natural background level of arsenic in the aquifers of Calabria region (Southern Italy)

A multivariate non-parametric approach for estimating probability of exceeding the local natural background level of arsenic in the aquifers of Calabria region (Southern Italy)

C. Apollaro a, D. Di Curzio b, I. Fuoco a,∗, A. Buccianti d,e, E. Dinelli f, G. Vespasiano a, A. Castrignanò b, S. Rusi b,

D. Barca a, A. Figoli c, B. Gabriele g, R. De Rosa a

a Department of Biology, Ecology and Earth Sciences (DIBEST), University of Calabria, via P. Bucci 15/B, 87036 Rende, CS, Italy

b Department of Engineering and Geology (InGeo), University "G. d'Annunzio" of Chieti-Pescara, via dei Vestini 31, 66013 Chieti, Italy

c Institute on Membrane Technology (ITM-CNR), via P. Bucci 17/C, 87036 Rende, CS, Italy

d Department of Earth Sciences, University of Florence (UniFI), Via G. La Pira 4, I-50121 Florence, Italy

e Institute of Geosciences and Earth Resources (CNR-IGG), Via G. La Pira 4, I-50121 Florence, Italy

f Department of Biological, Geological and Environmental Sciences (BiGeA), Alma Mater Studiorum - Università di Bologna, Piazza di Porta San Donato 1, 40126 Bologna, Italy g LISOC Group, Department of Chemistry and Chemical Technology, University of Calabria, via P. Bucci 12/C, 87036 Arcavacata di Rende, CS, Italy

∗ Corresponding author.

E-mail address: ilaria.fuoco@unical.it (I. Fuoco).

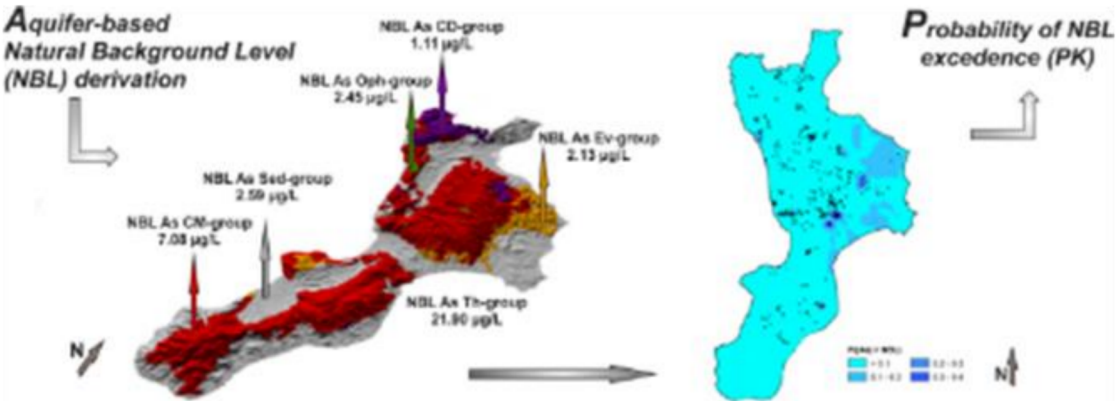
HIGHLIGHTS

- Geochemical processes causing the As enrichment in groundwaters
- As distribution in groundwater is affected by scale-dependent hydrogeochemical processes.
- Probability Kriging allowed performing a more accurate mapping of probability of NBL exceedance.
- Maps based on aquifer-based NBL values can be actual instruments for contamination risk management.

Keywords:

Arsenic, Aquifer-based NBLs, Groundwater, Water-rock interaction, Probability Kriging, Calabria region

GRAPHICAL ABSTRACT



abstract

The concept of natural background level (NBL) aims at distinguishing the natural and anthropogenic contributions to concentrations of specific contaminants, as groundwater management and protection tools. This is usu-ally defined as a unique value at a regional scale, even when the hydrogeological and geochemical features of a certain territory are far from homogeneous. The concentration of target contaminants is affected by multiple hydrogeochemical processes. This is the case of arsenic in the Calabria region, where concentrations are definitely variable in groundwater.

To overcome the limitation of a traditional approach and to include the intrinsic hydrogeological and geochemical heterogeneity into the definition of the natural contribution to As content in groundwater, an integrated probabilistic approach to the NBL assessment combining aquifer-based preselection criteria and multivariate non-parametric geostatistics was proposed. In detail, different NBL values were selected, based on the aquifer type and/or hydrogeochemical features. Then, these aquifer-based NBL values of arsenic were used in the Prob-ability Kriging method to map the probability of exceedance and to provide contamination risk management tools. This multivariate geostatistical approach that takes advantage of the physico-chemical variables used in the aquifer-based NBL values definition allowed mapping the probability of exceedance of As in a physically-based way. The hydrogeochemical diversity of the study area and all the processes affecting As concentrations in the aquifers have been considered too.

As a result, the obtained map was characterized by a short-range and long-range variability due to local hydro-geochemical anomalies and water-rock interaction and/or atmospheric precipitation. By this approach, the NBL exceedance probability maps proved to be less “noisy”, because the local hydrogeochemical conditions were filtered, and more capable of pointing out anthropogenic inputs or very anomalous natural contributions, which need to be investigated more in detail and properly managed.

1. Introduction

The hydrogeochemical characterization of groundwater bodies is the first step to investigate and assess water quality status. This information enables policy actions to be implemented for the reservoirs' protection, the proper management, and the safe supply of water resources. More than 98% of groundwater solutes derives from natural processes; however, anthropogenic activities can add minor to large amounts of ions in

89 the waters, so modifying their geochemistry and causing pollution (e.g., Huang et al.,
90 2013; Jiang et al., 2009; Kumar, 2014; Aiuppa et al., 2003).

91 Among polluting elements, arsenic (As) is one of the most monitored elements
92 worldwide due to its harmful effects on human health and environmental ecosystems
93 (IARC, 2012).

94 Based on these peculiarities the World Health Organization (WHO) has fixed 10 µg/L
95 as threshold value for drinking water (WHO, 2017 and reference therein). The As
96 content in the groundwaters may vary in a wide range of concentrations (<0.05-5000
97 µg/L) occurring as a trace element in many rocks and derived soils (Smedley and
98 Kinniburgh, 2002), or conversely at high concentrations in arsenic-rich pyrite
99 ($Fe(As,S)_2$) and arsenopyrite (FeAsS) (e.g. Tisserand et al., 2014; Pfeifer et al., 2007).
100 However, the mixing with As-rich geothermal waters and the desorption of As species
101 from Fe(III)-oxy-hydroxides (HFO) under alkaline or reducing conditions can also
102 affect arsenic mobility (Ravenscroft et al., 2009).

103 Although the concentration of this element is strictly related to hydrogeochemical
104 processes taking place in different kinds of aquifers, As can enter groundwater through
105 a by-product of anthropogenic activities such as the agriculture, mining and chemical
106 industry (Li et al., 2021). Determining the sources of pollution, is fundamental for
107 planning mitigation or remediation actions. The first step is defining the natural
108 background levels (NBLs) and the origin of solutes to accomplish this task.

109 The concept of NBLs was initially applied in geochemical exploration for
110 distinguishing among natural concentrations and anomalies due to ore occurrence or
111 anthropogenic input (Hawkes and Webb, 1962). Nowadays, NBLs represent "the range
112 of concentration of a given element, isotope or chemical compound in solution, derived
113 entirely from natural, geological, biological or atmospheric sources, under conditions
114 not perturbed by anthropogenic activity" (Edmunds and Shand, 2008). For
115 groundwater bodies, NBLs depend on the atmosphere and rainfall com-position, water-
116 rock interaction, chemical, and biological processes in both vadose and saturated zone,
117 interactions with other water bodies, residence time, and dissolution rate (Edmunds and
118 Shand, 2008; Appelo and Postma, 2005; Wendland et al., 2005; Marini, 2006; Critelli
119 et al., 2014; Palmucci et al., 2016a, 2016b; Rusi et al., 2018).

120 In the last decades, the concept of NBL and natural origin pollutants has been the
121 subject of International water management policies (e.g., the European Water
122 Framework Directive - WFD 2000/ 60/EC, and Groundwater Directive - GWD
123 2006/118/EC). In fact, NBLs are commonly used to define the threshold values, to
124 which the above mentioned policies refer to establish the actual chemical status of
125 aquifers (Quevauviller, 2005; De Caro et al., 2017; Marandi and Karro, 2008; Müller

et al., 2006). Accordingly, groundwater bodies have to be examined in detail through field investigation and chemical monitoring.

From a methodological point of view, the EC Working Group on Groundwater, through the BRIDGE project (Background CRiteria for the IDentification of Groundwater thresholds—BRIDGE 2006), defined two approaches for NBLs derivation: (i) statistical elaboration of data from different countries for the same type of aquifer, and (ii) use of pre-selection criteria which exclude contaminated samples based on anthropogenic indicators.

In literature, NBL assessment of a pollutant is performed using both approaches (e.g. Rahman et al., 2021, Gao et al., 2020; Panno et al., 2006; Matschullat et al., 2000; Parrone et al., 2019; Sellerino et al., 2019; De Caro et al., 2017), even though the probability distribution approach was widely used (e.g. Preziosi et al., 2014). However, defining differentiated NBLs, as a function of the main occurring water-rock interaction processes, remains a priority aspect (e.g. Preziosi et al., 2010; Wendland et al., 2005; Edmunds et al., 2003).

For instance, Ducci and Sellerino (2012) adopted the preselection criteria approach, merging those of ISPRA Protocol (ISPRA, 2009) and BRIDGE Project to assess the NBLs of some solutes, including As, in some groundwater bodies of the Campania region (Southern Italy). Preziosi et al. (2010) have assessed the NBLs for some pollutants, including As, in central Italy's groundwater bodies adopting the BRIDGE preselection criteria with modifications, stressing the relevant role played by hydrogeological/hydrogeochemical knowledge. Thus, the geo-knowledge of water bodies plays a key role in the accurate determination of NBLs, mostly in articulated geological settings like those of Calabria Region (Southern Italy), where As contamination was already revealed (Apollaro et al., 2003; Figoli et al., 2020).

Besides representing reference values for groundwater quality assessment and aquifer classification, accurate NBL values are also used in risk analysis of groundwater pollution caused by anthropogenic activities. Risk assessment usually requires a probabilistic approach due to the intrinsic uncertainty of the considered variables and the limited number of data generally available (Castrignanò et al., 2008; Manzione et al., 2021). It consists of recording the spatial occurrence with which specific criteria are met or not met and then mapping the exceedance probability (Dowd and Pardo-Igúzquiza, 2002; Passarella et al., 2020).

To delineate areas at equal risk of aquifer degradation, several re-searchers have dealt with the NBL exceedance probability and used geostatistical techniques based on the Regionalized Variable Theory (Matheron, 1971). The most common approach to exceedance probability mapping is Indicator Kriging (IK), a univariate non-parametric

technique that estimates the probability distribution by interpolating a binary indicator variable (Ducci et al., 2016; Avila-Sandoval et al., 2018; Sellerino et al., 2019; Parrone et al., 2020). Dalla Libera et al. (2018) adopted a multivariate approach to exceedance probability mapping of arsenic by applying Indicator Co-Kriging to different indicator variables corresponding to distinct NBL thresholds, in order to im-prove the probability prediction. Other Authors used more advanced geostatistical methods, such as Bayesian Kriging (Molinari et al., 2019) or Stochastic Simulation (Guadagnini et al., 2020), with arsenic and ammonium concentrations.

All the previous scientific contributions to the topic of NBL exceedance probability mapping took into account spatial autocorrelation. However, none considered the essential role of the hydrogeochemical processes by assessing the impact of physico-chemical variables (i.e., covariates) on the concentration of certain elements (i.e. pollutants) in groundwater. To this end, exceedance probability can be predicted with Probability Kriging (PK), a non-parametric multivariate technique (Journel, 1989; Carr and Mao, 1993), which uses auxiliary variables to improve probability estimation in the considered domain. PK has proved to perform better than IK in probability mapping (Juang and Lee, 2000; Adhikary et al., 2011; Shaddad et al., 2020), but the choice of auxiliary information is crucial.

In the light of the previous considerations, this research is aimed at: (1) defining a differentiated NBL of arsenic in the Calabria Region aquifers, based on a combined approach obtained by integrating prese-lection criteria method with probability distribution approach, (2) map-ping the local NBL exceedance probability taking into account all the variables used in the preselection stage with Probability Kriging, which will be compared to a univariate approach (Indicator Kriging), to assess the impact of ancillary information.

2. Geological and hydrogeological setting

The Calabrian Peloritan Orogen (CPO) represents a fragment of the European margin that, during the Europe-Apulia collision (Oligocene-Early Miocene), overthrust on the Maghrebian-Sicilian and Appennine thrust-and-fold belts (Cirrincione et al., 2015). The CPO has been formally divided into two sectors (Northern and Southern) separated by a strike-slip tectonic line localized in the Catanzaro trough's proximity (Boccaletti et al., 1984; Tansi et al., 2007). Different paleogeographic evolutions characterize the sectors during the middle of the Miocene. In particular, the Northern one is characterized by the superposition of three major structural elements (nappe pile) identifying different paleogeographic domains.

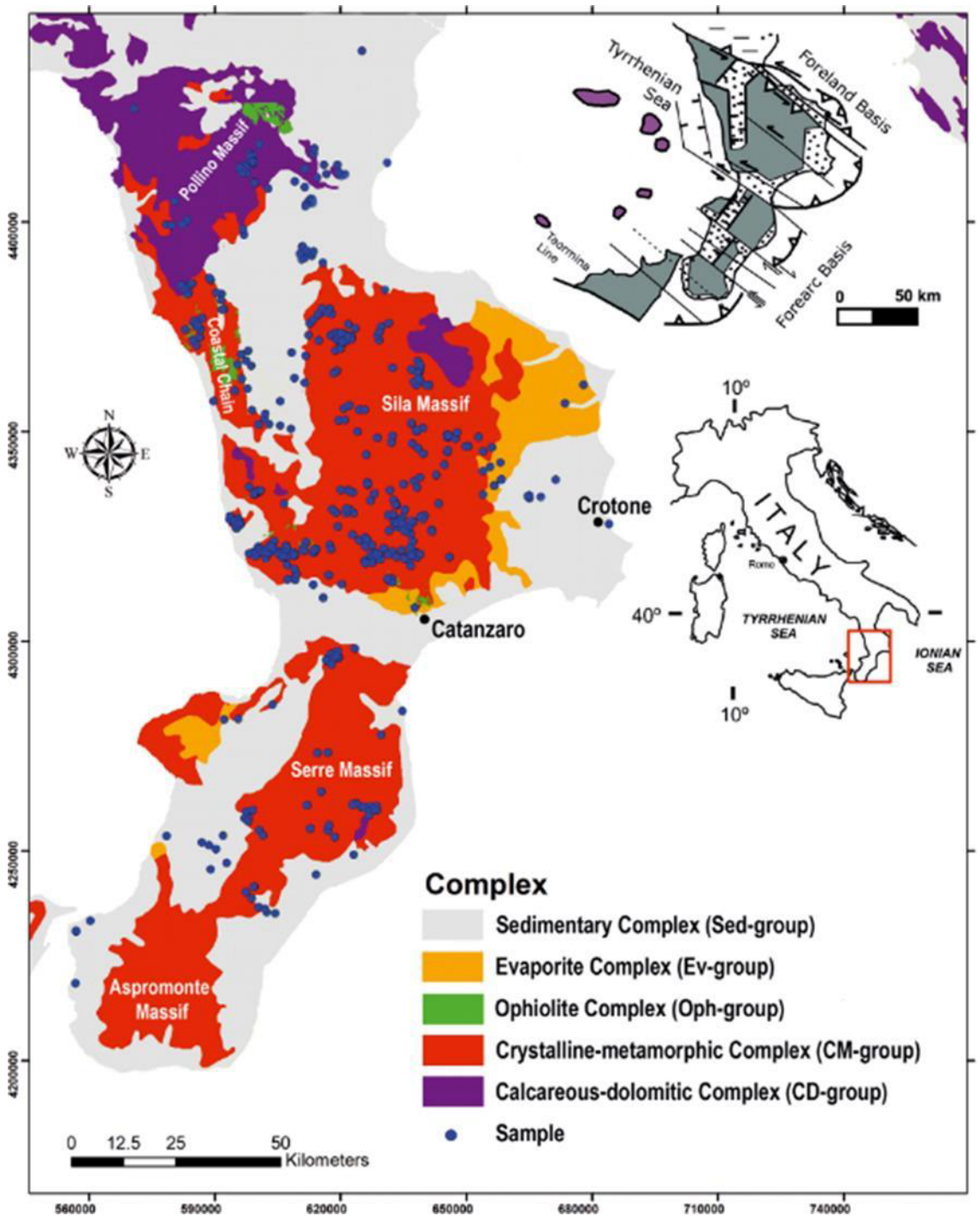


Fig. 1. Simplified hydrogeological map of Calabria region. (Modified by Tripodi et al. (2018) and Apollaro et al. (2019a)).

The lower Apennine Unit (Trias-Miocene) consisting of Mesozoic calcareous sedimentary and calcareous metasedimentary successions (Iannace et al., 2007). These lithotypes are generally very fractured, be-cause of their syn- and post-orogenetic tectonic evolution. In these de-posits, dissolving rainwater's action tends to expand and improve the network of pre-existing fractures favouring the development of karst phenomena. This domain represents a vast reservoir in which water circulation is conditioned by the geometric relationships with the sur-rounding geological units and

209 the well-developed inner structural discontinuities. Due to its peculiar hydrogeological
210 characteristics, it represents the wide “carbonate-dolomitic complex” (CD-group in
211 Fig. 1), one of the primary sources of drinking water supplies in South-ern Italy
212 (Allocca et al., 2007).

213 The intermediate Alpine Liguride Unit (Liberi and Piluso, 2009; Bloise et al., 2009,
214 2017) comprises Cretaceous-Paleogene metapelites-ophiolitic-carbonate assemblage
215 (Tithonian-Neocomian).

216 The Calabride Unit is constituted by Hercynian and pre-Hercynian gneiss, granite, and
217 metapelite (Van Dijk et al., 2000).

218 From a hydrogeological point of view, the Calabride and Liguride Units are similar.
219 These complexes are characterized by a less important hydrogeological system than
220 the carbonate-dolomitic complex. They show a very complex hydrogeological asset in
221 which porous shallow aquifers, identified in the weathered profile, coexist, and interact
222 with the intermediate aquifer, represented by the fissured rock basement. In these
223 aquifers, groundwater circulation and recharge are mainly conditioned by rock
224 alteration, varying according to the depth and fracturing degree.

225 The southern CPO sector includes the Serre, the Aspromonte, and the Peloritani
226 massifs. At the south of this sector, the following structural elements can be recognized
227 (Tortorici, 1982): (a) Longi-Taormina Unit including schists with intercalations of
228 quartzites, metarenites and metalimestones (Tortorici, 1982); (b) Mandanici Unit
229 which mainly consists in Hercynian Paleozoic phyllite sequence (Pezzino et al., 2008);
230 (c) Aspromonte Unit comprising amphibolite facies metamorphic rocks intruded by
231 peraluminous granites (Cirrincione et al., 2008); (d) Stilo Unit made up of low
232 greenschist to low amphibolite–facies Hercynian Paleozoic rocks, intruded by late-to-
233 post-orogenic plutonic body (Pezzino et al., 2008). Together with the Calabride Unit
234 of northern Calabria, these units can be incorporated into the “crystalline-metamorphic
235 complex” (CM-group in Fig. 1). In con-trast, the alpine Liguride Unit represents the
236 ophiolite complex (Oph-group in Fig. 1).

237 The sedimentary rock complex represents other important complexes and the two main
238 complexes (carbonate-dolomitic and crystalline-metamorphic). These complexes
239 include the porous-quaternary aquifers representative of the alluvial and coastal plains.
240 The aquifers' high interest mainly correlates with the high groundwater request for
241 anthropic activities (Vespasiano et al., 2015a, 2016). These terrains constitute
242 continuous/heterogeneous and anisotropic aquifers, the most represented in the
243 Calabrian Region's outcrop (Allocca et al., 2007).

The evaporite complex is a minor but important hydrogeological complex in Calabria region because of the pronounced geochemical im-print on the water chemistry leave by the high solubility of evaporite rocks, even where they directly don't crop up to the surface.

The sampling sites fall into the main recognized hydrogeological complexes, exploited to regional water supply (Fig. 1). Each sampling point's chemical composition is linked to water-rock processes in each investigation area.

3. Materials and methods

3.1. Sampling and chemical analysis

536 water samples distributed throughout the Calabria region were collected and analyzed for major components and arsenic con-tent. The sampling and analytical methods were already described in previous works (e.g. Fuoco et al., 2021; Apollaro et al., 2020, 2021). However, a brief description is reported to make the paper self-consistent. In the field, the labile parameters like pH, temperature (T), redox potential (Eh), electrical conductivity (EC), were deter-mined with a previously calibrated multiparametric probe. In contrast, total alkalinity was measured with the acidimetric titration method using a micro-dosimeter, HCl 0.05 N as the titrating agent and methyl orange as the indicator.

Four aliquots for each sampling point were collected, filtered in the field via a 0.45 μm pore-size membrane filter and acidified using the addition of pure acid (1% HNO_3), except the aliquots for the determination of anions which were stored without further treatment. The concentrations of major cations and anions (Na^+, K^+, Mg^{2+}, Ca^{2+}, F^-, Cl^-, SO_4^{2-}, and NO_3^-) were determined with high-performance liquid chromatography (HPLC, Dionex DX 1100), while the concentration of SiO2 was measured using VIS spectrophotometry. Arsenic content was determined with a quadrupole inductively coupled plasma-mass spectrometer (ICP-MS, PerkinElmer/SCIEX, ELAN DRC-e) with a collision reaction cell capable of reducing or avoiding the formation of polyatomic spectral interferences. Data quality for major ions was evaluated with charge balance (±10%), while NIST 1643f standard reference solution was taken into account for arsenic analysis accepting a deviation from certified concentration below the ±10%.

3.2. Aquifer-based NBL derivation

Based on acquired hydrogeological and chemical knowledge, the whole dataset was split into six sub-datasets, corresponding to the main hydrogeological complexes they fall, except for the thermal waters group, interacting with a different type of lithotypes. The sub-datasets are the water belonging to: (i) the crystalline-metamorphic rock complex (CM-group), (ii) the ophiolite rock complex (Oph-group), (iii) the calcareous-dolomitic rock complex (CD-group), (iv) the sedimentary rock complex (Sed-group), (v) the evaporite rock com-plex (Ev-group) and (vi) the thermal waters group (Th-group).

The Oph-group and the CM-group were treated separately due to the different mineralogical assemblage which constitute the involving rocks. Considering each water group's chemical features, specific prese-lection criteria were adopted to exclude the samples influenced by anthropogenic pollution. Indeed, BRIDGE's preselection criteria (Müller et al., 2006) were adopted with modification based on groundwater bodies' peculiarities. The used preselection criteria are the following:

1. ion balance was set ≤10% for all groups;

2. NO_3^- concentration was set ≤10 mg/L for all groups except for the Ev- and Th-groups. For the latter, the threshold was increased to 50 mg/L since the nitrate salts could occur in evaporite deposits and they could be leached during the water-rock process (e.g. Holloway and Dahlgren, 2002);

3. the samples belonging to the Ev- and Th-groups characterized by NaCl >1000 mg/L were not removed;

4. the few samples characterized by a negative Eh were not separated from those of oxidizing aquifers. Although arsenic is a sensitive redox element, no significant correlation was observed between redox potential and As concentration in the studied samples (see Table 3a). Moreover, the waters representing the reducing aquifers in each group were statistically not representative.

All datasets contained values below the limit of detection (LOD). To allow their statistical processing, the arsenic values below the LOD were substituted through the "Simple substitution Method" with half of the LOD (US EPA, 2002). Furthermore, an outlier test was also performed to identify the not representative samples in each water group using the software ProUCL v. 5.1 (Singh and Maichle, 2015). Once deleted all samples potentially influenced by anthropogenic activities, the possible presence of two or more data populations, linked to different natural sources, can be identified.

In this study, the partitioning method proposed for the first time by Sinclair (1974) was applied for identifying the overlapping of more populations in a single water group. In the past, the method has been successfully used in geochemical mineral exploration

(Sinclair, 1974, 1991) for separating the baseline concentration to the "anomalies", which represented the ore deposits.

Inflexion points on cumulative probability plots can recognize the different distributions belonging to two or more populations. This is be-cause, in cumulative probability plots, the values of a single normally or lognormally distributed population will form a straight line. In contrast, the overlapping of two or more populations will appear as a curved line with one or more inflexion points representing the populations' thresh-olds (Panno et al., 2006). Once the thresholds are identified, the populations can be separated.

The partitioning method was applied separately to the individual group of waters, and in some cases, two or more populations were identified. Generally, NBLs is defined as a percentile of the dataset distribution, between 90th and 97.7th percentile, according to the available data quality (Müller et al., 2006). In this study, the NBLs of arsenic were set at the 95th percentile of each specific water group's most representative probability distribution model.

3.3. Data analysis

3.3.1. Co-Kriging

Since As concentrations in the Calabria region aquifers are strictly related to water-rock interaction processes as well as to specific physico-chemical conditions, the estimation of the spatial distribution of As throughout its territory should be conditional to other hydrogeochemical variables, such as the main ions (i.e., Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, HCO_3^-, and SO_4^{2-}) and some physico-chemical properties (i.e., temperature, electrical conductivity, pH, and redox potential) of groundwater. For this purpose, the multivariate Co-Kriging method (CK) was preferred to the univariate approach (Wackernagel, 2003; Webster and Oliver, 2016, 2019; Di Curzio, 2019; Vessia et al., 2020a). LMC is then modelled as a linear combination of NS variogram models (gu), standardized to the unit sill for each spatial scale u:

$$\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{u=1}^{N_S} \mathbf{B}^u g^u(\mathbf{h}) \qquad (1)$$

where N is the number of variables, Γ (h) is an N × N matrix of direct variograms and cross-variograms, and Bu is a symmetric matrix of sills of the spatial structures $g^u(\mathbf{h})$ at each spatial scale u, while h is the lag vector. The CK system uses the estimated LMC to calculate the weights λ_α^i to estimate the target variable (zi_0^* (x_0)) in each node of the regular interpolation grid with mesh size of 2500 m overlapping Calabria region:

$$z^*_{i_0}(\mathbf{x}_0) = \sum_{i=1}^{n} \sum_{\alpha=1}^{n_i} \lambda^i_\alpha z_i(\mathbf{x}_\alpha) \quad (2)$$

where x_0 is the position where the target variable is estimated, $z_i(x_\alpha)$ are the observations in a neighbourhood of x_0, and I and α are the variable and position indices, respectively. Besides estimation (Eq. (2)), CK provides the estimation variance ($\sigma^2(x_0)$) as a measurement of uncertainty. With CK outcomes, it is possible to calculate the lower (LL, Eq. (3)) and upper(UL, Eq.(4)) limits of the 95% confidence interval of the estimated value (z(x)):

$$LL = z^*(x) - \frac{1.96 * \sigma}{\sqrt{N}} \quad (3)$$

$$UL = z^*(x) + \frac{1.96 * \sigma}{\sqrt{N}} \quad (4)$$

However, it is worth underlining that CK is the best linear unbiased predictor, i.e. with the minimum mean-squared prediction error, when estimation is linear in the data (Eq. (2)), which means that weights λ_α^i do not depend on the values zi(xα). When the data follow a multivariate Gaussian distribution, an estimation can be assumed linear, but for other distributions, the assumption of optimality may be a quite poor approximation (Schabenberger and Gotway, 2005). Statisticians solve the problem by transforming the data into a Gaussian distribution and then performing analyses with the transformed data. Thus, a highly skewed variable as As needs to be transformed into a Gaussian variable if one of the objectives is also the estimation of uncertainty and the chosen approach is parametric. To perform this transformation, the Gaussian Anamorphosis function (GA) has been used, which can convert a Gaussian variable (Y) into a variable with any distribution (Z = Φ(Y)) by a Hermite polynomial expansion (Chilès and Delfiner, 2012):

$$\Phi(Y) = \sum \Psi_i H_i(Y) \quad (5)$$

where H_i is Hermite polynomials and Ψ_i coefficients.

By inverting the GA (Y = Φ−1(Z)), a non-gaussian variable can be transformed to a gaussian one, to be submitted to geostatistical analysis.

Fitting of LMC allows, among other things, the estimation of spatial cross-correlation functions between pairs of variables (i.e., cross-variograms), in particular between the target variable (i.e., arsenic) and the auxiliary variables, which enable getting a deeper insight into the hydrogeochemical processes influencing the As content in groundwater.

Accordingly, for each spatial structures $g^u(\mathbf{h})$ at each spatial scale u, codisperison coefficients (ρ_{ij}^u) can be calculated using the following equation:

$$\rho_{ij}^u = \frac{b_{ij}^u}{\sqrt{b_{ii}^u b_{jj}^u}} \quad (6)$$

For each spatial scale u, the equation considers the partial sills of the pair of direct variograms (i.e., b_{ii}^u and b_{jj}^u) and of the corresponding cross-variogram (b_{ij}^u). It appears clear that codispersion coefficients, as defined in Eq. (6), can be seen as scale-dependent correlation between two variables.

After applying CK, the maps of both the Gaussian As estimate and its 95% confidence interval limits were back-transformed through the GA into maps with the original units (i.e., μg/L).

3.3.2. Indicator and Probability Kriging

Geostatistical estimation of the probability of exceeding a certain threshold (z_{thr}) is based on a simple binary transformation (Journel, 1989). In more detail, a numerical variable ($z(\mathbf{x}_\alpha)$), describing a property of the selected study area (e.g., the concentration of an element in groundwater) at the observation point $\mathbf{x}_\alpha$, is transformed into a binary variable, called indicator variable ($i(x_\alpha; z_{thr})$), as follows:

$$i(\mathbf{x}_\alpha; z_{thr}) = \begin{cases} 1 & \text{if } z(\mathbf{x}_\alpha) \geq z_{thr} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

The binary variable is 0, if the numerical values lie below the selected threshold, and 1 if the threshold is exceeded. In the case of the As con-centration in groundwater, the threshold value is represented by the NBL, either the unique or the local aquifer-based one (see Section 4.1).

The spatial probability estimation performed with Indicator Kriging (IK) is based on the following univariate estimator ($i*(\mathbf{x}_0; z_{thr})$):

$$i^*(\mathbf{x}_0; z_{thr}) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{x}_0; z_{thr})\, i(\mathbf{x}_\alpha; z_{thr}) \quad (8)$$

where λ_α represents the weight assigned to the indicator datum ($I(\mathbf{x}_\alpha; z_{thr})$) falling within the interpolation neighbourhood. The weights are calculated through an equation system that uses the variogram model of the indicator variable ($\gamma i(\mathbf{h}; z_{thr})$):

$$\gamma_i(\mathbf{h}; z_{thr}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{x}_\alpha; z_{thr}) - i(\mathbf{x}_\alpha + \mathbf{h}; z_{thr})]^2 \quad (9)$$

where xα is the sampling location and N(**h**) is the number of indicator value pairs at sampling locations separated by a distance vector (lag)**h**.

A univariate approach, such as IK, considers only one groundwater attribute. To consider the effect of the processes involved in the As release in aquifers, Probability Kriging was applied (Journel, 1989; Carr and Mao, 1993), representing an enhancement of IK because it uses multivariate data set to improve the estimation of the exceedance prob-ability. Since the indicator variable is binary, taking the values 0 or 1 (Eq. (7)), all the numerical auxiliary variables need to be converted into variables ranging between 0 and 1 with uniform distribution (k($\mathbf{x}_\alpha$)) through a relative rank order transformation:

$$k(\mathbf{x}_\alpha) = \frac{r(\mathbf{x}_\alpha)}{N} \quad (10)$$

where r($\mathbf{x}_\alpha$) ∈ (1,N) is the rank with data in increasing order, and N is the total number of the sampling locations ($\mathbf{x}_\alpha$).

Like CK, PK requires constructing a LMC of direct and cross-variograms of all variables, including indicator and relative rank trans-form variables (Eq. (1)). In PK, the probability of exceedance is then estimated as follows:

$$i^*(\mathbf{x}_0; z_{thr}) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{x}_0; z_{thr})\, i(\mathbf{x}_\alpha; z_{thr}) + \sum_{j=1}^{n_k} \sum_{\alpha=1}^{n_j} \nu_{\alpha,j}(\mathbf{x}_0; z_{thr})\, k_j(\mathbf{x}_\alpha) \quad (11)$$

where the number of observations for indicator and relative rank transform variables within the interpolation neighbourhood of $\mathbf{x}_0$ are given by n and n_j, respectively; n_k is the number of the considered auxiliary variables in addition to As; λ_α and $\nu_{\alpha,j}$ are the weights assigned to indicator and relative rank transform variables, respectively.

3.3.3. Model performance evaluation and approach comparison

To compare univariate (i.e., Indicator Kriging) and multivariate (i.e., Probability Kriging) methods to map the exceedance probability applied to both unique and variable NBL values, cross-validation has been applied (Vessia et al., 2020b). In detail, the observations has been removed sequentially one at a time from the dataset and then each one was estimated at the sampling location, using the remaining part of the data

and the model under evaluation. The following statistics were calculated on the errors (i.e., the differences between experimental (z_i) and estimated (z_i^*) values): Mean Error (ME), Root Mean Squared Error (RMSE), Mean Standardized Error (MSE) and Root Mean Squared Standardized Error (RMSSE). In MSE and RMSSE, the CK standard deviation is used for standardization. The equations of these statistics are shown below with their optimal values:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(z_i^* - z_i) \rightarrow 0 \quad (12)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i^* - z_i)} \rightarrow 0 \quad (13)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{z_i^* - z_i}{\sigma_i}\right) \rightarrow 0 \quad (14)$$

$$RMSSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{z_i^* - z_i}{\sigma_i}\right)^2} \rightarrow 1 \quad (15)$$

Based on these model performance evaluation statistics, estimation is considered unbiased when ME and MSE are close to zero, precise when RMSE is also close to zero and accurate when RMSSE tends to one (Cressie, 2015).

All geostatistical analyses were performed with Isatis software(Geovariances, 2018).

4. Results and discussion

4.1. Hydrogeochemical characterization and NBLs assessment

The whole dataset was composed of 536 water samples characterized by different chemical compositions due to the water-rock interaction processes occurring in each investigation area. The EC values of groundwater range between 13 and 13190 µS/cm, with an average value of 550 µS/cm and a median of 281 µS/cm, while the average pH value is 7.12 (range 5.0-10.4). The waters show temperature values ranging from 4.3 to 38.4 °C and predominately oxidizing redox conditions (min. -266 mV; max 594 mV; median, 165 mV). The Total Ionic Salinity (TIS) values range between 0.59 and 366.72 meq/L, with an average of 11.73 meq/L and a median of 5.14 meq/L (Table S1, Supplementary material). Based on the average of the main cations and anions, Ca is found as the predominant cation, followed by Na > Mg > K, while HCO_3 appears as

458 the dominant anion, followed by SO_4 > Cl > NO_3 > F. As concentrations range from
459 0.01 to 435 µg/L with the highest anomalies located in restricted areas of the region,
460 possibly related to local hydrogeological features (Table 1). The 95th percentile of non
461 parametric distribution, was equal to 6.55 µg/L, and this value was set as
462 undifferentiated arsenic NBL. Due to the heterogeneous hydrogeological setting and
463 the different types of rocks affecting water-rock interaction processes, the
464 determination of differentiated NBLs was essential to correctly assess the water quality
465 depending on the type of aquifer. Based on the groundwater attributes and
466 hydrogeological considerations, differentiated preselection criteria were chosen, and
467 applied to derive appropriate NBLs.

468

469 *4.1.1. Crystalline-metamorphic waters group (CM-group)*

470 The CM-group consists of 337 water samples originating mainly in the Sila Massif,
471 Serre Massif, and Coastal Chain (see Fig. 1). The whole group of waters shows a
472 median pH value of 6.9, a median temperature value of 12.6 °C, and oxidizing redox
473 conditions (median Eh value of 177.3 mV) prevail (Table 1). Electrical conductivity
474 (EC) ranges from 13 to 1245 µS/cm with a median value of 194.2 µS/cm, pointing out
475 that this group includes samples characterized by different evolution grade due to
476 different times of water-rock interaction. Indeed, Total Ionic Salinity (TIS) of waters
477 ranges from 0.6 to 19.6 meq/L (Fig. 2c). Based on the triangular plots of major anions
478 and major cations (Fig. 2a, b), 165 of the available 337 groundwater samples from CM-
479 group can be attributed to the Ca-HCO_3 chemical type followed by Na-HCO_3 (110
480 samples), Na-Cl (33 samples) and Mg-HCO_3 (23 samples) hydrochemical facies (Fig.
481 2a, b). These compositions reflect the typical groundwater evolution (Apollaro et al.,
482 2019a) from shallow to deep crystalline-metamorphic aquifers:

$$\text{NaCl (low TIS value)} \rightarrow \text{NaHCO}_3 \text{ (low TIS value)} \rightarrow \text{CaHCO}_3 \rightarrow \text{NaHCO}_3 \text{ (high TIS value)}$$

483

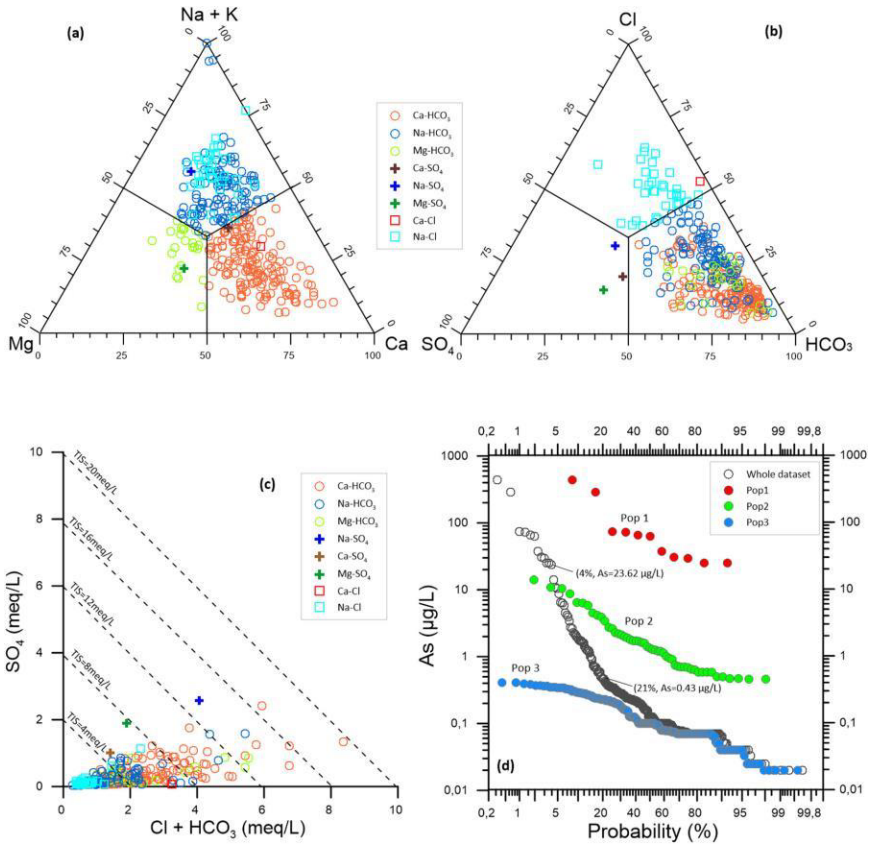


Fig. 2. Triangular plots of (a) major cations and (b) major anions obtained from concentrations in equivalent units for the crystalline-metamorphic water group (CM-group); (c) correlation diagram of SO_4 vs. HCO_3 + Cl reporting the iso-salinity lines; (d) probability plot of As concentrations (µg/L) and partition of the distributions in log-normal populations. This figure illustrates all the elaborations performed for each group of waters, to define aquifer-based NBL values (Figs. S1–S5, Supplementary material).

The remaining 1% of the dataset comprises few waters ascribable to the Na-Ca-Mg-SO_4 and Ca-Cl chemical facies.

The As concentrations of the waters belonging to the CM-group vary widely, from 0.01 to 435 µg/L (median value of 0.13 µg/L) (Table 1).

Once applied to the preselection criteria (see Section 3.2), the dataset appeared non-normally and non-lognormally distributed (Shapiro-Wilk, Kolmogorov-Smirnov tests). Multiple populations were highlighted by using the cumulative probability plot shown in Fig. 2d and applying the partitioning method (Sinclair, 1974). In this group of waters, two thresholds of concentration (inflexion points) and three sample populations (Pop1, Pop2 and Pop3) were identified (Fig. 2d). The two thresholds were set at 23.62 and

504 0.43 μg/L. The higher threshold was interpreted as the lower boundary for
505 groundwaters interacting with localized As-bearing mineralization (Pop 1). This
506 population is characterized by high As concentrations ranging from 23.62 to 435 μg/L.
507 These high arsenic concentrations are a peculiar characteristic of the shallow
508 groundwaters circulating in a restricted area of the Calabria region, representing an
509 unexplored mineralized spot. Indeed, these samples were collected in a small basin
510 located in the southern sector of the Sila Massif (Catanzaro province). Figoli et al.
511 (2020) reported some analyses of the rocks outcropping in this area, showing the
512 presence of pyrite containing variable arsenic concentrations (200-1100 mg/kg).
513 Moreover, a consistent amount of arsenic often appears associated with adsorbed
514 species onto secondary minerals.

515 The lower threshold (0.43 μg/L) was interpreted as the upper bound for groundwaters
516 whose As concentrations originate mainly from atmospheric inputs (Smedley and
517 Kinniburgh, 2002; Gallo et al., 2017). In other words, the As concentrations
518 characterizing this population (Pop3, 0.01-0.43 μg/L) are not related to the water-rock
519 interaction processes.

520 The log-normal distribution of Pop2 shows arsenic concentrations ranging from 0.43
521 to 14 μg/L, which can be related to a widespread and not predictable occurrence of As-
522 sulphides into the basement (Bonardi et al., 1982). Based on these considerations, the
523 95th percentile, 7.08 μg/L, was set as the natural background level of arsenic for the
524 crystalline-metamorphic aquifers of the Calabria region (Table 2).

525 *4.1.2. Ophiolite waters group (Oph-group*

526 The Oph-group consists of 41 water samples from limited areas of the Calabria region,
527 mainly located in the Coastal Chain and Mount Reventino area (Southern Sila Massif)
528 (see Fig. 1), where crops up the ophiolitic sequence, mainly including metabasalts,
529 serpentinites, epimetamorphic phyllites and quartzites (Alvarez, 2005).

530 The considered waters show a median pH value of 7.4, a median temperature value of
531 13.8 °C and predominantly oxidizing redox conditions (median Eh value of 178 mV).
532 The EC and TIS values vary between 91 and 1780 μS/cm (median EC value of 360
533 μS/cm) (Table 1), 1.3 and 23.7 meq/L (median TIS value of 8.9 meq/L, Fig. S1,
534 Supplementary material), respectively. The main chemical types are represented by Ca-
535 HCO_3 (24 samples) and Mg-HCO_3 (15 samples) compositions (Apollaro et al., 2019b)
536 whereas, the two samples belonging to the Na-Cl and Na-HCO_3 represent low evolved
537 waters (Fig. S1, Supplementary material). In this geological setting, Mg-HCO_3
538 composition is typical of the waters interacting with serpentinite rocks (Fuoco et al.,
539 2020; Apollaro et al., 2021). The Ca-HCO_3 chemical composition could be controlled
540 by the dissolution of Ca-rich phases forming the metabasalts (Apollaro et al., 2011;

Critelli et al., 2015). Generally, these rocks can contain a variable amount of arsenic, for instance, 0.4-45 mg/kg, 0.5-143 mg/kg, 2.2-7.6 mg/kg for greenstones, phyllite/ slate and quartzite, respectively (Smedley and Kinniburgh, 2002), which can be released during the water-rock interaction processes.

The As concentrations of considered waters range from 0.01 to 5.97 μg/L (Table 1) and appear as a population lognormally distributed (Fig. S1), whose 95th percentile, 2.45 μg/L, was set as the natural back-ground level of arsenic for the ophiolitic aquifers of the Calabria region (Table 2).

4.1.3. Calcareous-dolomitic waters group (CD-group)

The CD-group consists of 37 water samples falling mainly in the Pollino Massif (northern sector of the Calabria Region) (see Fig. 1). The considered waters discharge from limestones, dolostones, chert-bearing limestones and marly limestones (Apollaro et al., 2021), representing the main aquifers for drinking water supplies in Calabria (Allocca et al., 2007).

The whole dataset shows a median pH value of 7.7, a median temperature value of 11.4 °C and oxidizing redox conditions with the median Eh value of 56.2 mV (Table 1). The EC values range from 122 μS/cm to 660 μS/cm (median EC value of 369.5 meq/L), whereas the TIS values vary between 3.4 and 14 meq/L (median TIS value of 8.2 meq/L). These waters are ascribable to the Ca-HCO_3 hydrochemical facies (Fig. S3), representing the typical composition of the carbonate aquifers. However, some samples appear shifted towards the magnesium (Mg) vertex, pointing out the control of the dolostones' magnesian component (Apollaro et al., 2021). Generally, limestones and dolostones contain a low concentration of arsenic. Some authors quoted their As content in the range of 0.1-21.1 mg/kg (Smedley and Kinniburgh, 2002; Adriano, 2001), which can be released during the water-rock interaction processes. The waters interacting with these rock types show As concentrations between 0.01 and 5.97 μg/L (Table 1) and appear as a single population lognormally distributed (Fig. S3). The 95th percentile of this distribution, 1.11 μg/L, was set as the natural background level of arsenic for the calcareous-dolomitic aquifers of the Calabria region (Table 2).

4.1.4. Evaporite waters group (Ev-group)

The Ev-group consists of 31 water samples falling mainly to the northern part and on the eastern side of Calabria Region (see Fig. 1).

575 The pH values are in the range 6.5-8.0 and the median temperature value is 16.5 °C
576 (Table 1). The EC values range from 798 µS/cm to 4325 µS/cm (median EC value of
577 1035 meq/L) whereas, the TIS values vary between 17.7 and 103.7 meq/L (median TIS
578 value of 32 meq/L). The Ev-group shows both reducing and oxidizing conditions with
579 the median Eh value of -23 mV (min. -233 mV; max 268 mV). Based on triangular
580 plots (Fig. S3, Supplementary material) 18 of the available 31 groundwater samples
581 were classified as Ca-SO_4 chemical type, whereas the Ca-HCO_3 (9 samples), Mg-HCO_3
582 (2 samples) and Na-Cl (2 samples) hydrochemical facies represent the remaining
583 samples. These compositions derived from the interaction of these waters with the
584 carbonate-evaporite deposits (calcite/dolomite, gypsum, anhydrite and halite) of
585 Triassic or Messianian age outcropping in the Calabria Region (Gaglioti et al., 2019;
586 Vespasiano et al., 2015b). The concentrations of arsenic in sulphate minerals like
587 gypsum or anhydrite were estimated in the range < 1-6 mg/Kg whereas, the halite
588 crystals can contain until 30 mg/Kg of arsenic (Smedley and Kinniburgh, 2002). The
589 waters belonging to the Ev-group show an As content ranging from 0.05 to 2.70 µg/L
590 (Table 1) and after the application of the preselection criteria (see Section 3.2), the
591 dataset appeared non-normally and non-lognormally distributed (Fig. S3,
592 Supplementary material). The presence of two populations was highlighted using the
593 partitioning method (Sinclair, 1974), identifying a thresh-old of concentration at 0.35
594 µg/L and, consequently, two populations, both normally distributed (Pop1 and Pop2).
595 The identified threshold was interpreted as the upper bound for groundwater (Pop2).
596 As concentrations are less controlled by water-rock interaction processes, thus deriving
597 mainly from atmospheric inputs (Gallo et al., 2017; Smedley and Kinniburgh, 2002).
598 Pop 1 shows a higher median value of temperature than Pop2 (18.8 °C, Pop1; 15.5 °C,
599 Pop 2). The higher temperature could cause an increase in the leaching of arsenic and
600 its enrichment into the groundwaters (Apollaro et al., 2017). Based on these
601 considerations, the 95th percentile of the first population, 2.13 µg/L, was set as the
602 natural background level of arsenic for the waters interacting with evaporitic rocks of
603 the Calabria region (Table 2).

604

605 *4.1.5. Sedimentary water group (Sed-group)*

606 The Sed-group consists of 66 water samples distributed throughout the Calabria region
607 (see Fig. 1). These waters are linked to aquifers developed in terrigenous and alluvial
608 successions, originating during the Neogene-Quaternary age. These sedimentary
609 successions fill mainly the transversal and longitudinal basins, known as Crotone
610 Basin, Crati Valley, and Catanzaro Trough (Van Dijk et al., 2000) and include the
611 deposits of the coastal plain (Vespasiano et al., 2019). The petrographical-
612 mineralogical and chemical composition of sedimentary successions is highly variable

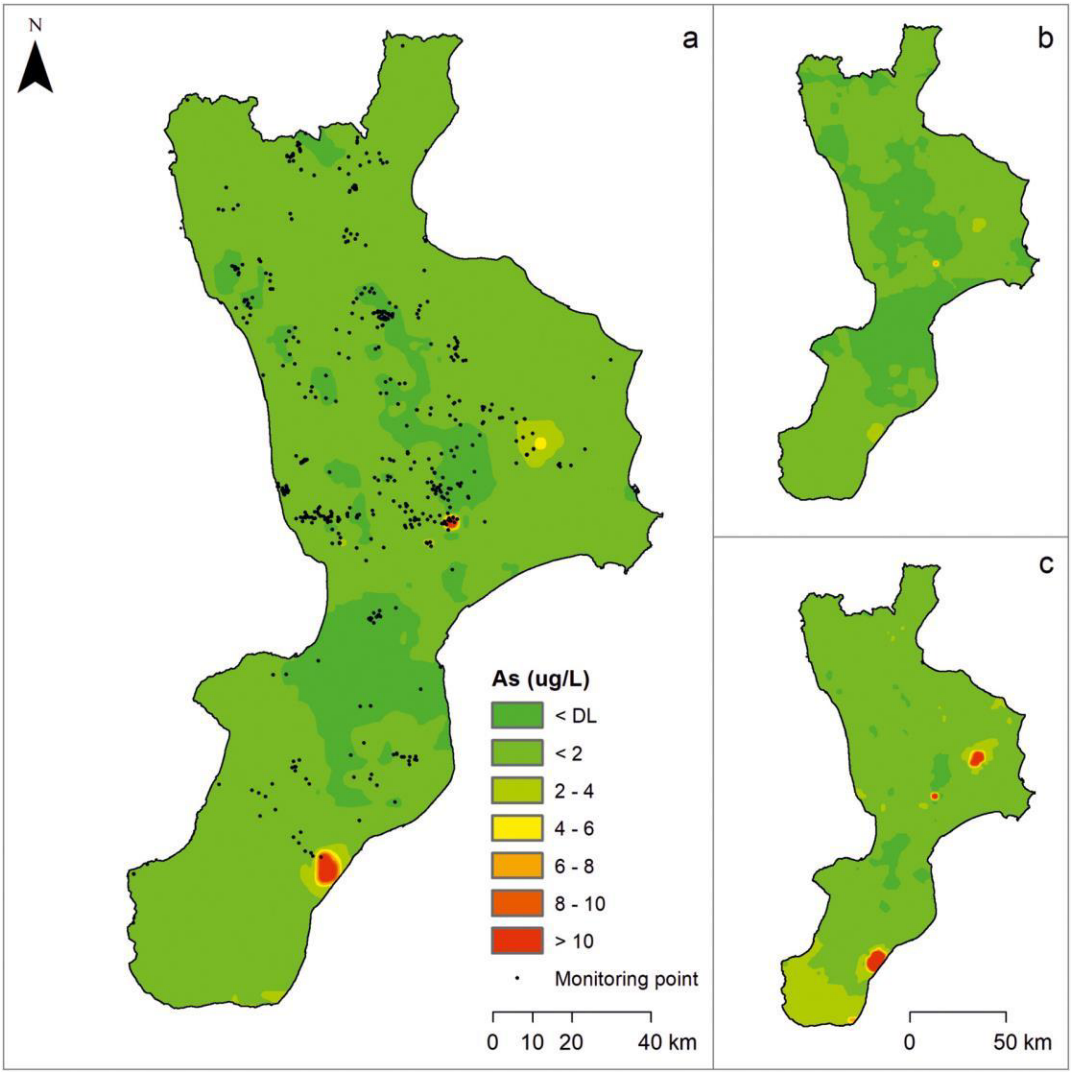
because they derived from erosion, transport and deposition of all rocks forming the CPO.

The whole dataset belonging to the Sed-group shows a median pH value of 7.5, a median temperature value of 16.2 °C and oxidizing redox conditions with the median Eh value of 180 mV (Table 1). The median EC value is 457 μS/cm (range 75-1570 μS/cm) whereas, the median TIS value is 8.2 meq/L (range 1-30.2 meq/L) (Fig. S4, Supplementary material). Considering the triangular plots (Fig. S4, Supplementary material), these waters were classified as Na-Cl (5 samples), Na-HCO_3 (4 samples) and Ca-HCO_3 (57 samples). The first group includes both low evolved waters and samples which probably interacted with soluble salts whereas, the Na-HCO_3 and Ca-HCO_3 water types represent more evolved waters.

On average, the concentrations of As in sedimentary rocks are in the range 5-10 mg/Kg (Smedley and Kinniburgh, 2002; Adriano, 2001). Sands and sandstones tend to have a lower concentration of As than argillaceous deposits, reflecting the As concentrations of their main forming minerals (Smedley and Kinniburgh, 2002). The waters belonging to the Sed-group shows As concentrations ranging from 0.07 to 3.02 μg/L (Table 1). Also in this case, after the application of the prese-lection criteria (see Section 3.2), the dataset appeared non-normally and non-lognormally distributed (Fig. S4, Supplementary material) due to the high number of samples characterized by arsenic values below the detection limit (24% of the whole dataset). The concentration threshold was identified at 0.1 μg/L by using the partitioning method (Sinclair, 1974). Based on these considerations, the 95th percentile of the first population, 2.59 μg/L, was set as the natural background level of arsenic for the sedimentary aquifers of the Calabria region (Table 2).

4.1.6. Thermal waters group (Th-group)

The Th-group comprises 24 water samples distributed throughout the Calabria region (see Fig. 1) and belonging to the main local geothermal circuits. The samples are hot and deep waters interacting with different kinds of aquifers, like crystalline-metamorphic complex, ophiolite complex, carbonate-evaporite complex and with aquifers



644

645 Fig. 3. Map of the distribution of As in groundwater (a), obtained with CK, and the
646 corresponding lower (b) and upper (c) limits of the 95% confidence interval. 11

647

648 hosted in the sedimentary successions of the Neogene-Quaternary age (e.g. Apollaro et
649 al., 2019c, 2016, 2012; Vespasiano et al., 2015b, 2015c, 2015d, 2015e, 2014, 2012a,
650 2012b; Vespasiano et al., 2021).

651 The dataset belonging to the Th-group shows pH values ranging from 6.6 to 8.6, and a
652 median temperature value of 26.2 °C (range 20.2-38.4 °C). The samples show both
653 reducing and oxidizing redox conditions with the median Eh value of -14.8 mV (min.
654 -266 mV; max 192 mV) whereas, the median EC value is 2708 µS/cm (range 1019-
655 13,190 µS/cm) (Table 1). This water group shows the highest salinity with TIS values
656 in the range of 22.8-366.7 meq/L (median TIS value 62.8 meq/L) (Fig. S5,
657 Supplementary material). Based on the triangular plots (Fig. S5, Supplementary
658 material) six hydrochemical facies were recognized: Na-Cl (9 samples), Ca-SO_4 (7

samples), Na-SO_4 (5 samples), Ca-HCO_3 (1 sample), Mg-HCO_3 (1 sample) and Na-HCO_3 (1 sample).

The considered waters show arsenic values ranging from 0.4 to 17.4 μg/L and fall within the range estimated for the geothermal waters by Smedley and Kinniburgh (2002). The data appear as a single population lognormally distributed (Fig. S5, Supplementary material). The 95th percentile, 21.90 μg/L, was set as the natural background level of arsenic for the thermal waters of the Calabria region (Table 2).

4.2. Arsenic distribution in groundwater

A nested LMC was fitted to Gaussian transformed data, including a short-range (6600 m) spherical structure and a long-range (20,000 m) spherical structure (Fig. S6, Supplementary material). Each one of these structures describes variation at the corresponding spatial scale, which is affected by particular hydrogeochemical processes as repre-sented by the considered physico-chemical variables (i.e., As, Na^+, K^+, Ca^{2+}, Mg ,Cl , HCO_3^-, and SO_4^{2-}, temperature, electrical conductivity, pH, and redox potential). Table 3 compares the Pearson correlations with the two structural codisperion matrices, to point out the scale-dependent hydrogeochemical processes affecting the arsenic content in groundwater.

The codispersion coefficients at short-range (Table 3b) show that As has significative positive correlations with Cl^-, Na^+ and K^+, a high positive correlation with the water temperature, and a negative correlation with pH. Since the mixing between low-temperature groundwater and geothermal fluids is characterized by relatively higher temperatures and a prevalent Na-K-Cl water type (Table 1), this short-range variability of As concentration can be ascribed to the Th-group, which occurs in some specific parts of the study area. Additionally, an inverse relation-ship between As and pH is typical of oxidative dissolution of sulphide minerals, such as the As-bearing mineralization which have been recognized in some limited areas of the crystalline aquifers; thus, CM-group contributes to the small-scale variability of As concentrations in ground-water. At a larger scale (Table 3c), As is positively correlated with EC, T, pH, and all the major ions, while it has a significative negative correlation with Eh. Water-rock interaction and, to a minor extent, atmospheric precipitation may account for these relations. The long-range codispersion coefficients essentially indicate that As content in groundwater increases as the residence time in the aquifers increases. However, the total contribution of the large-scale variability has been proved to be of few micrograms per litre (see Section 4.1). As a result, the As concentration in the different aquifers seems to be highly conditioned by the physico-chemical properties of groundwater.

Fig. 3 shows the distribution of As throughout the whole region (Fig. 3a), together with the 95% confidence interval limits maps (Fig. 3b and c) to evaluate the uncertainty of estimation. From the spatial distribution obtained with CK, it appears clear that the concentration of As in most of the regional territory is generally very low, with values that reach at most 2 μg/L regardless of the type of aquifer or groundwater. The atmospheric precipitation and the water-rock inter-action in aquifers characterized by a very limited amount of arsenic in their mineral phases (i.e., Oph-group, CD-group, and Sed-group) ac-count for these extremely low concentrations in groundwater. Besides, at some locations, As concentrations are intermediate to high, frequently exceeding 10 μg/L. These well-defined and limited areas correspond to sites where groundwater interacts with As-bearing mineralizations (e.g., As-rich pyrite) or groundwater mixes with geo-thermal fluids.

Comparing the three maps, it results that As concentration estimates in the eastern and southern sectors of the study area, where the density of the sampling points of the monitoring network was considerably lower than elsewhere, are more uncertain. However, the uncertainty is, on average, of the order of a few μg/L, confirming the good performance of the LMC demonstrated by the cross-validation (Table 4).

4.3. Mapping the probability of NBL exceedance

Even though the arsenic spatial distribution in the Calabria region is strongly dependent on the physico-chemical properties of groundwater, the natural hydrogeochemical processes do not occur identically and to the same extent in all the aquifer types. This implies that the natural contribution of each aquifer to the As content in groundwater has to be necessarily considered differentiated, to effectively define areas where the natural background concentration can be exceeded because of either anthropogenic inputs or natural anomalies.

Different values of NBL of arsenic (i.e., differentiated NBL), one for each group of waters and aquifer type (Table 5), have then been considered in the NBL probability of exceedance mapping. To stress the improvement of the approach compared to the ones used in literature so far, the maps of NBL probability of exceedance have also been obtained by IK using both regional and aquifer-based NBL values, and by applying PK technique to the unique regional NBL value (i.e., 10.41 μg/L).

A spherical variogram model has been fitted in IK (Figs. S7 and S8, Supplementary material). In contrast, the LMC in PK includes two structures: a short-range (6600 m) spherical model and a long-range (20,000 m) spherical structure (Figs. S9 and S10,

731 Supplementary material). The nested LMCs related to the unique or differentiated
732 NBLs have the same ranges as the ones used in CK.

733 The difference between the models used in IK and PK points out that including other
734 hydrogeochemical covariates, related to the same processes, allowed improving the
735 interpretation of the experimental variogram of the indicator variable and identifying
736 two scale-dependent variability structures. Accordingly, a robust and reliable nested
737 LMC was fitted. The results of cross-validation (Table 6) confirm this evidence by
738 proving that the PK estimates are more precise (i.e., lower RMSE values) and accurate
739 (i.e., RMSSE closer to 1) than the IK ones. Besides, both ME and MSE values
740 demonstrate that all the models are unbiased because their values are always close to
741 zero.

742 Fig. 4 compares the maps of NBL exceedance probability obtained with IK and the
743 unique and differentiated NBL (Fig. 4a and b) and PK considering the unique and
744 differentiated NBL (Fig. 4c and d).Since the multivariate approach (i.e., Probability
745 Kriging) has proved to be more reliable and representative of the dynamics affecting
746 As content in groundwater, the implications of using an aquifer-based NBL will be
747 highlighted considering only the PK results (Fig. 4c and b).

748

749

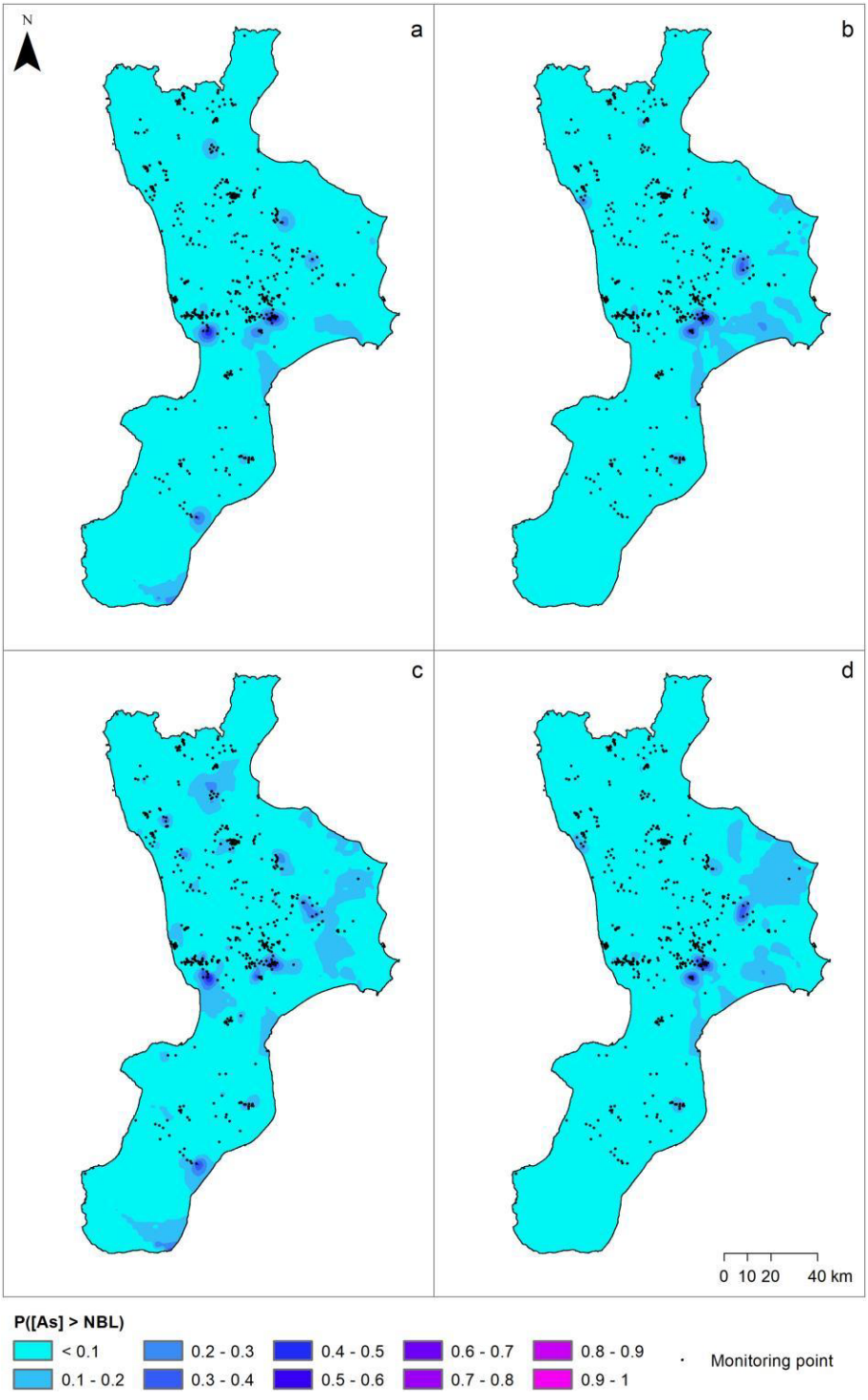


Fig. 4. Maps of the probability of NBL exceedance for As, obtained with the four combinations of NBL values and methods: a) regional NBL and IK, b) aquifer-based NBL and IK, c) regional NBL and PK, and d) aquifer-based NBL and PK.

For both regional and aquifer-based NBLs, it appears clear that within most of the regional territory the probability that As concentration exceeds the level attributable to the natural contribution is below 0.1. This evidence is consistent with the spatial distribution of As esti-mated with CK (Fig. 3), which basically points out very low concentrations in most of the study area. The highest probabilities (i.e., up to 0.6-0.7) are instead related to the areas, where local mineralization or geothermal fluids uprising occurs. In addition, probability values in the order of 0.2-0.3 can be observed in the eastern areas of the Calabria region, which might be ascribed to anthropogenic sources. However, the map of NBL probability exceedance that considers the regional value (Fig. 4c) attributes probabilities greater than 0.1-0.2 to all the hydrogeochemical anomalies, although arsenic concentrations are in the order of few micrograms per litre and do not represent real environmental problems to be managed. This results in "noisy" maps because all these anomalous areas are supposed to have As inputs additional to the natural contribution. In addition, As concentrations that at a regional level would not represent a problem, because below the regional NBL, might instead be "not natural" in some hydrogeological contexts, such as in sedimentary of calcareous-dolomitic aquifers (Table 2). The map obtained considering aquifer-based NBL values (Fig. 4d) overcomes this problem by filtering the natural hydrogeochemical features of each group of waters and pointing out critical situations that need to be investigated more in detail, due to anthropogenic inputs or very anomalous natural contributions. In this way, maps of NBL exceedance probability can definitely become actual instruments for contamination risk management.

As a result, the map obtained by the application of PK to the aquifer-based NBL, besides spotting the highly hydrogeochemical anomalies due to local mineralizations in the crystalline aquifer and to mixing be-tween fresh groundwater and geothermal fluids in the Sila Massif, points out areas at south east where probability is in the order of 0.2. This area is characterized by the presence of sedimentary and evaporitic aquifers (Fig. 1). Since these kinds of aquifers have very low NBL values, here the As content may also be due to anthropogenic sources and/or hydrogeological interaction with hydrogeochemical anomalies within the nearby aquifers. Although a certain level of uncertainty is clearly present, these are in any case areas that should be investigated with further detail by site-specific surveys to discriminate the actual sources of arsenic concentrations in groundwater.

5. Conclusions

An integrated probabilistic approach to the natural background level assessment of arsenic that combines aquifer-based preselection criteria and multivariate non-parametric geostatistics has been proposed and applied to the Calabria region.

In detail, groundwater samples collected throughout the entire territory were split into groups of waters, each one related to a specific type of aquifer or hydrogeological feature, as in the case of mixing between fresh groundwater and uprising geothermal fluids. The aquifer-based NBL values for As, assigned to the groups of waters using the probability distribution approach, were used in the Probability Kriging method to map the probability of exceedance. In this multivariate approach, the same covariates used to estimate the spatial distribution of arsenic in groundwater with the Co-Kriging method were considered, even though rank-order transformed. These physico-chemical auxiliary variables proved to describe most of the scale-dependent hydrogeochemical processes affecting As concentrations in the aquifers. For this reason, they allowed estimating more accurately the spatial distribution of the probability of NBL exceedance, as demonstrated by comparing the Probability Kriging and the univariate Indicator Kriging results. In addition, using aquifer-based NBL values in the Probability Kriging allowed obtaining a less "noisy" map, because the natural hydrogeochemical features of each group of waters were filtered. In this way, the resulting maps of NBL exceedance probability can definitely become actual instruments for contamination risk management. They can point out critical situations (i.e., anthropogenic inputs or very anomalous natural contributions) that need to be investigated more in detail and properly managed.

The proposed multivariate probabilistic approach, which considers aquifer-based NBL values and the main physico-chemical variables affecting the behaviour of a certain element in groundwater, is very robust and flexible because it is physically based and can be adapted to different hydrogeological and hydrogeochemical contexts.

CRediT authorship contribution statement

C. Apollaro: Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition. D. Di Curzio: Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. I. Fuoco: Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. A. Buccianti: Conceptualization, Methodology, Data curation. E. Dinelli: Conceptualization, Methodology, Data curation. G. Vespasiano: Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing

– original draft, Writing – review & editing, Visualization. A. Castrignanò: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. S. Rusi: Conceptualization, Writing – original draft, Writing – review & editing. D. Barca: Conceptualization, Methodology, Formal analysis. A. Figoli: Conceptualization, Data curation, Funding acquisition. B. Gabriele: Conceptualization, Data curation, Funding acquisition. R. De Rosa: Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.

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.

Acknowledgements

The work has been supported by the project “AsSe” n. CUP: J28I17000030006, cofounded by Fondo FESR POR Calabria FESR FSE 2014-2020-Azione 1.2.2,

References

Adhikary, P.P., Dash, C.J., Bej, R., Chandrasekharan, H., 2011. Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in of najafgarh block, Delhi, India. Environ. Monit. Assess. 176 (1), 663–676. https://doi.org/10. 1007/s10661-010-1611-4.

Adriano, D.C., 2001. Trace Elements in Terrestrial Environments. 2nd ed. Springer, New York, New York, NY https://doi.org/10.1007/978-0-387-21510-5.

Aiuppa, A., Bellomo, S., Brusca, L., d'Alessandro, W., Federico, C., 2003. Natural and anthropogenic factors affecting quality of an active volcano (Mt. Etna, Italy). Appl. Geochem. 18 (6), 863–882.

Allocca, V., Celico, F., Celico, P., De Vita, P., Fabbrocino, S., Mattia, S., Monacelli, G., Musilli, I., Piscopo, V., Scalise, A.R., Summa, G., Tranfaglia, G., 2007. Note illustrative della Carta idrogeologica dell’Italia meridionale. Istituto Poligrafico e Zecca dello Stato. ISBN: 88-448-0215-5, p. 211.

Alvarez, W., 2005. Structure of the Monte reventino greenschist folds: a contribution to untangling the tectonic-transport history of Calabria, a key element in italian tectonics. J. Struct. Geol. 27, 1355–1378.

Apollaro, C., Fuoco, I., Bloise, L., Calabrese, E., Marini, L., Vespasiano, G., Muto, F., 2021. Geochemical Modeling of Water-Rock Interaction Processes in the Pollino National Park. Geofluids.

Apollaro, C., Caracausi, A., Paternoster, M., Randazzo, P., Aiuppa, A., De Rosa, R., Fuoco, I., Mogelli, G., Muto, F., Vanno, E., Vespasiano, G., 2020. Fluid geochemistry in a low-enthalpy geothermal field along a sector of southern Apennines chain (Italy). J. Geochem. Explor. 219, 106618.

Apollaro, C., Buccianti, A., Vespasiano, G., Vardè, M., Fuoco, I., Barca, D., Bloise, A., Miriello, D., Cofone, F., Servidio, A., De Rosa, R., 2019a. Comparative geochemical study between the tap waters and the bottled mineral waters in Calabria (Southern Italy) by compositional data analysis (CoDA) developments. Appl. Geochem. 107, 19–33.

Apollaro, C., Fuoco, I., Brozzo, G., De Rosa, R., 2019b. Release and fate of cr (VI) in the ophiolitic aquifers of Italy: the role of Fe (III) as a potential oxidant of Cr (III) sup-ported by reaction path modelling. Sci. Total Environ. 660, 1459–1471.

Apollaro, C., Tripodi, V., Vespasiano, G., De Rosa, R., Dotsika, E., Fuoco, I., Critelli, S., Muto, F., 2019c. Chemical, isotopic and geotectonic relations of the warm and cold waters of the galatro and antonimina thermal areas, southern Calabria, Italy. Mar. Pet. Geol. 109, 469–483.

Apollaro, C., Vespasiano, G., Barca, D., De Rosa, R., Fuoco, I., Miriello, D., Bloise, A., 2017. Preliminary Study of Geogenic Arsenic in Thermal Waters of Northern Calabria (Italy). Goldschmidt 2017 Conference. Paris, FR. 13-18 August 2017.

Apollaro, C., Vespasiano, G., Muto, F., De Rosa, R., Barca, D., Marini, L., 2016. Use of mean residence time of water, flowrate, and equilibrium temperature indicated by water geothermometers to rank geothermal resources. application to the thermal water circuits of northern Calabria. J. Volcanol. Geotherm. Res. 328, 147–158.

Apollaro, C., Dotsika, E., Marini, L., Barca, D., Bloise, A., De Rosa, R., Doveri, M., Lelli, M., Muto, F., 2012. Chemical and isotopic characterization of the thermo mineral water of Terme sibarite springs (Northern Calabria, Italy). Geochem. J. 46, 117–129.

Apollaro, C., Marini, L., Critelli, T., Barca, D., Bloise, A., De Rosa, R., Liberi, F., Miriello, D., 2011. Investigation of rock-to-water release and fate of major, minor, and trace elements in the metabasalt–serpentinite shallow aquifer of mt. reventino (CZ, Italy) by reaction path modelling. Appl. Geochem. 26 (9–10), 1722–1740.

Apollaro, C., De Rosa, R., Ferruzza, G., 2003. The Cosenza and Messina sheets of the geochemical map of Italy: explanatory notes. Geochemical Baselines of Italy. Pacini Editore.

Appelo, C., Postma, D., 2005. Carbonates and carbon dioxide. Geochemistry, Groundwater and Pollution, Second edition.

Avila-Sandoval, C., Júnez-Ferreira, H., González-Trinidad, J., Bautista-Capetillo, C., Pacheco-Guerrero, A., Olmos-Trujillo, E., 2018. Spatio-temporal analysis of natural and anthropogenic arsenic sources in groundwater flow systems. Int. J. Environ. Res. Public Health 15 (11), 2374. https://doi.org/10.3390/ijerph15112374.

Bloise, A., Barrese, E., Apollaro, C., Miriello, D., 2009. Flux growth and of Ti- and Ni-doped forsterite single crystals. Cryst. Res. Technol. 44 (5), 463–468.

Bloise, A., Catalano, M., Critelli, T., Apollaro, C., Miriello, D., 2017. Naturally occurring asbestos: potential for human exposure, san Severino lucano (Basilicata, southern Italy). Environ. Earth Sci. 76 (19), 1–13.

Boccaletti, M., Nicolich, R., Tortorici, L., 1984. The calabrian arc and the Ionian Sea in the dynamic evolution of the Central Mediterranean. Mar. Geol. 55, 219–245.

Bonardi, G., De Vivo, B., Giunta, G., Lima, A., Perrone, V., Zuppetta, A., 1982. Mineralizzazioni dell'Arco calabro peloritano Ipotesi Genetiche e Quadro Evolutivo. Boll. Soc. Geol. It. 101 (2), 141–155.

Carr, J.R., Mao, N.H., 1993. A general form of probability kriging for estimation of the indicator and uniform transforms. Math. Geol. 25 (4), 425–438. https://doi.org/10.1007/ BF00894777.

Castrignanò, A., Buttafuoco, G., Giasi, C., 2008. Assessment of groundwater salinization risk using multivariate geostatistics. geoENV VI–Geostatistics for Environmental Ap-plications. Springer, Dordrecht, pp. 191–202 https://doi.org/10.1007/978-1-4020-6448-7_16.

Castrignanò, A., Landrum, C., De Benedetto, D., 2015. Delineation of management zones in precision agriculture by integration of proximal sensing with multivariate geostatistics. examples of sensor data fusion. Agric. Conspec. Sci. 80 (1), 39–45.

Chilès, J.-P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. 2nd edition. John Wiley & Sons.

Cirrincione, R., Fazio, E., Fiannacca, P., Ortolano, G., Pezzino, A., Punturo, R., 2015. The Calabria-Peloritani Orogen, a composite terrane in Central Mediterranean; its overall architecture and geodynamic significance for a pre-Alpine scenario around the Tethyan basin. Progresses in Deciphering Structures and Compositions of Basement Rocks. 84. Periodico di Mineralogia, pp. 701–749. https://doi.org/10.2451/ 2015PM0446.3B.

Cirrincione, R., Ortolano, G., Pezzino, A., Punturo, R., 2008. Poly-orogenic multi-stage metamorphic evolution inferred via P-T pseudosections: an example from aspromonte massif basement rocks (Southern Calabria, Italy). Lithos 103, 466–502.

Cressie, N., 2015. Statistics for Spatial Data. John Wiley & Sons.

Critelli, T., Vespasiano, G., Apollaro, C., Muto, F., Marini, L., De Rosa, R., 2015. Hydrogeochemical study of an ophiolitic aquifer: a case study of lago (Southern Italy, Calabria). Environ. Earth Sci. 74 (1), 533–543.

Critelli, T., Marini, L., Schott, J., Mavromatis, V., Apollaro, C., Rinder, T., De Rosa, R., Oelkers, E.H., 2014. Dissolution rates of actinolite and chlorite from a whole-rock experimental study of metabasalt dissolution from 2≤ pH≤ 12 at 25 C. Chem. Geol. 390, 100–108.

Dalla Libera, N., Fabbri, P., Mason, L., Piccinini, L., Pola, M., 2018. A local natural background level concept to improve the natural background level: a case study on the drainage basin of the Venetian Lagoon in Northeastern Italy. Environ. Earth Sci. 77 (13), 1–15. https://doi.org/10.1007/s12665-018-7672-3.

938 De Caro, M., Crosta, G.B., Frattini, P., 2017. Hydrogeochemical characterization and natural
939 background levels in urbanized areas: Milan metropolitan area (Northern Italy). J. Hydrol. 547,
940 455–473.

941 Di Curzio, D., Rusi, S., Signanini, P., 2019. Advanced redox zonation of the San Pedro Sula alluvial
942 aquifer (Honduras) using data fusion and multivariate geostatistics. Sci. Total Environ. 695,
943 133796. https://doi.org/10.1016/j.scitotenv.2019.133796.

944 Di Curzio, D., 2019. Hydrogeochemical and hydrodynamic features affecting redox processes in
945 groundwater. Acque Sotterranee 8 (3), 7–19. https://doi.org/10.7343/as-2019-401.

946 Di Curzio, D., Palmucci, W., Rusi, S., Signanini, P., 2016. Evaluation of processes controlling Fe and
947 Mn contamination in the San Pedro Sula porous aquifer (North Western Honduras). Rend. Online
948 Soc. Geol. Ital. 41, 42–45. https://doi.org/10.3301/ROL. 2016.88.

949 Dowd, P.A., Pardo-Igúzquiza, E., 2002. The incorporation of model uncertainty in geostatistical
950 simulation. Geogr. Environ. Model. 6 (2), 147–169. https://doi.org/10.
951 1080/1361593022000029476.

952 Ducci, D., Sellerino, M., 2012. Natural background levels for some ions in groundwater of the
953 Campania region (southern Italy). Environ. Earth Sci. 67 (3), 683–693.

954 Ducci, D., de Melo, M.T.C., Preziosi, E., Sellerino, M., Parrone, D., Ribeiro, L., 2016. Combining
955 natural background levels (NBLs) assessment with indicator kriging analysis to improve
956 groundwater quality data interpretation and management. Sci. Total Envi-ron. 569, 569–584.
957 https://doi.org/10.1016/j.scitotenv.2016.06.184.

958 Edmunds, W.M., Shand, P., 2008. Natural Groundwater Chemistry. Blackwell Publishing Ltd, Oxford
959 https://doi.org/10.1002/9781444300345.

960 Edmunds, W.M., Shand, P., Hart, P., Ward, R.S., 2003. The natural (baseline) quality of groundwater:
961 a UK pilot study. Sci. Total Environ. 310 (1–3), 25–35.

962 European Commission (EC), 2006. Directive 2006/118/EC of the European Parliament and of the
963 Council of 12 December 2006 on the protection of groundwater against pollution and
964 deterioration. Off. J. Eur. Union L 372, 19–31.

965 European Commission (EC), 2000. Directive 2000/60/EC of the European Parliament and of the
966 Council of 23 October 2000 establishing a framework for community action in the field of water
967 policy. Off. J. Eur. Communities L327 (43), 1–72.

968 Figoli, A., Fuoco, I., Apollaro, C., Chabane, M., Mancuso, R., Gabriele, B., De Rosa, R., Vespasiano,
969 G., Barca, D., Criscuoli, A., 2020. Arsenic-contaminated groundwaters remediation by
970 nanofiltration. Sep. Purif. Technol. 238, 116461.

971 Fuoco, I., Apollaro, C., Criscuoli, A., De Rosa, R., Velizarov, A., Figoli, A., 2021. Fluoride polluted
972 groundwaters in Calabria region (Southern Italy): natural source and remediation. Water 2021
973 (13), 1626. https://doi.org/10.3390/w13121626.

974 Fuoco, I., Figoli, A., Criscuoli, A., Brozzo, G., De Rosa, R., Gabriele, B., Apollaro, C., 2020.
975 Geochemical modeling of chromium release in natural waters and treatment by RO/NF
976 membrane processes. Chemosphere 254, 126696.

977 Gaglioti, S., Infusino, E., Caloiero, T., Callegari, G., Guagliardi, I., 2019. Geochemical
978 characterization of spring waters in the Crati River basin. Geofluids, Calabria (Southern Italy)
979 https://doi.org/10.1155/2019/3850148.

980 Gallo, L., Corapi, A., Apollaro, C., Vespasiano, G., Lucadamo, L., 2017. Effect of the interaction
981 between transplants of the epiphytic lichen Pseudevernia furfuracea L. (Zopf) and rainfall on the
982 variation of element concentrations associated with the water-soluble part of atmospheric
983 depositions. Atmos. Pollut. Res. 8 (5), 912–920.

984 Gao, Y., Qian, H., Huo, C., Chen, J., Wang, H., 2020. Assessing natural background levels in shallow
985 groundwater in a large semiarid drainage basin. J. Hydrol. 584, 124638.

986 Geovariances, 2018. ISATIS Software: Technical References Release 2018.1. Geovariances and
987 Ecole des Mines de Paris.

988 Guadagnini, L., Menafoglio, A., Sanchez-Vila, X., Guadagnini, A., 2020. Probabilistic assessment of
989 spatial heterogeneity of natural background concentrations in large-scale groundwater bodies
990 through functional geostatistics. Sci. Total Environ. 740, 140139.
991 https://doi.org/10.1016/j.scitotenv.2020.140139.

992 Hawkes, H.E., Webb, J.S., 1962. Geochemistry in Mineral Exploration. Harper, New York. Holloway,
993 J.M., Dahlgren, R.A., 2002. Nitrogen in rock: occurrences and biogeochemical implications.
994 Glob. Biogeochem. Cycles 16 (4), 65-1-17.

995 Huang, G., Sun, J., Zhang, Y., Chen, Z., Liu, F., 2013. Impact of anthropogenic and natural processes
996 on the evolution of groundwater chemistry in a rapidly urbanized coastal area, South China. Sci.
997 Total Environ. 463, 209–221.

998 Iannace, A., Vitale, S., D'Errico, M., Mazzoli, S., Di Staso, A., Macaione, E., Messina, A., Reddy,
999 S.M., Somma, R., Zamparelli, V., Zattin, M., Bonardi, G., 2007. The carbonate tectonic units of
1000 northern Calabria (Italy): a record of Apulian palaeomargin evolution and Miocene convergence,
1001 continental crustal subduction, and exhumation of HP-LT rocks. J. Geol. Soc. Lond. 164, 1165–
1002 1186.

1003 IARC, 2012. A Review of Human Carcinogens. Part C: Arsenic, Metals, Fibers, and Dusts.
1004 International Agency for Research on Cancer, Lyon, France.

1005 ISPRA, 2009. Protocollo per la Definizione dei Valori di Fondo per le Sostanze Inorganiche nella
1006 Acque Sotterranee.

1007 Jiang, Y., Wu, Y., Groves, C., Yuan, D., Kambesis, P., 2009. Natural and anthropogenic fac-tors
1008 affecting the groundwater quality in the Nandong karst underground river system in Yunan,
1009 China. J. Contam. Hydrol. 109 (1–4), 49–61.

1010 Journel, A.G., 1989. Fundamentals of Geostatistics in Five Lessons. Vol. 8. American Geophysical
1011 Union.

1012 Juang, K.W., Lee, D.Y., 2000. Comparison of three nonparametric kriging methods for delineating
1013 heavy-metal contaminated soils. Am. Soc. Agron. Crop Sci. Soc. Am. Soil Sci. Soc. Am. 29 (1),
1014 197–205. https://doi.org/10.2134/jeq2000. 00472425002900010025x.

1015 Kumar, P.S., 2014. Evolution of groundwater chemistry in and around Vaniyambadi in-dustrial area:
1016 differentiating the natural and anthropogenic sources of contamination. Geochemistry 74 (4),
1017 641–651.

1018 Liberi, F., Piluso, E., 2009. Tectonometamorphic evolution of the ophiolite sequences from Northern
1019 Calabrian Arc. Ital. J. Geosci. (Boll. Soc. Geol. Ital.) 128, 483–493.

1020 Li, Y., Bi, Y., Mi, W., Xie, S., Ji, L., 2021. Land-use change caused by anthropogenic activities
1021 increase fluoride and arsenic pollution in groundwater and human health risk. J. Hazard. Mater.
1022 406, 124337.

1023 Manzione, R.L., Silva, C.D.O.F., Castrignanò, A., 2021. A combined geostatistical approach of data
1024 fusion and stochastic simulation for probabilistic assessment of shallow water table depth risk.
1025 Sci. Total Environ. 765, 142743. https://doi.org/10.1016/j.scitotenv. 2020.142743.

1026 Marandi, A., Karro, E., 2008. Natural background levels and threshold values of monitored
1027 parameters in the Cambrian-Vendian groundwater body, Estonia. Environ. Geol. 54 (6), 1217–
1028 1225.

1029 Marini, L., 2006. Geological Sequestration of Carbon Dioxide: Thermodynamics, Kinetics, and
1030 Reaction Path Modeling. Elsevier.

1031 Matheron, G., 1971. Theory of regionalized variables and its applications. Ecole Nationale Superieure
1032 des Mines de Paris. 5, p. 211.

1033 Matschullat, J., Ottenstein, R., Reimann, C., 2000. Geochemical background–can we calculate it?
1034 Environ. Geol. 39 (9), 990–1000.

1035 Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2019. Geostatistical multimodel
1036 approach for the assessment of the spatial distribution of natural back-ground concentrations in
1037 large-scale groundwater bodies. Water Res. 149, 522–532.
1038 https://doi.org/10.1016/j.watres.2018.09.049.

1039 Müller, D., Blum, A., Hart, A., Hookey, J., Kunkel, R., Scheidleder, A., Tomlin, C., Wendland, F.,
1040 2006. D18: final proposal for a methodology to set up groundwater threshold values in Europe
1041 BRIDGE project. Background Criteria for the Identification of Grounwater Thresholds, 6th
1042 Framework Programme Contract, p. 6538.

1043 Palmucci, W., Rusi, S., Di Curzio, D., 2016a. Mobilization processes responsible for iron and
1044 manganese contamination of grounwater in central adriatic Italy. Environ. Sci. Pollut. Res. 23
1045 (12), 11790–11805. https://doi.org/10.1007/s11356-016-6371-4.

1046 Palmucci, W., Rusi, S., Pennisi, M., Di Curzio, D., 2016b. Contribution of B and Sr isotopes to assess
1047 boron contamination of groundwater: case studies in Central Italy. Rend. On-line Soc. Geol. Ital.
1048 41, 65–68. https://doi.org/10.3301/ROL.2016.94.

1049 Panno, S.V., Kelly, W.R., Martinsek, A.T., Hackley, K.C., 2006. Estimating background and threshold
1050 nitrate concentrations using probability graphs. Groundwater 44 (5), 697–709.

1051 Parrone, D., Ghergo, S., Preziosi, E., 2019. A multi-method approach for the assessment of natural
1052 background levels in groundwater. Sci. Total Environ. 659, 884–894.

Parrone, D., Ghergo, S., Frollini, E., Rossi, D., Preziosi, E., 2020. Arsenic-fluoride co-contamination in groundwater: background and anomalies in a volcanic-sedimentary aquifer in Central Italy. J. Geochem. Explor. 217, 106590. https://doi. org/10.1016/j.gexplo.2020.106590.

Passarella, G., Masciale, R., Maggi, S., Vurro, M., Castrignanò, A., 2020. A probabilistic ap-proach to assess the risk of groundwater quality degradation. In: Faruque, F. (Ed.), Geospatial Technology for Human Well-being and Health. Springer.

Pfeifer, H.R., Häussermann, A., Lavanchy, J.C., Halter, W., 2007. Distribution and behavior of arsenic in soils and waters in the vicinity of the former gold-arsenic mine of Salanfe, Western Switzerland. J. Geochem. Explor. 93 (3), 121–134.

Pezzino, A., Angì, G., Fazio, E., Fiannacca, P., Lo Giudice, A., Ortolano, G., Punturo, R., Cirrincione, R., De Vuono, E., 2008. Alpine metamorphism in the Aspromonte Massif: implications for a new framework for the southern sector of the Calabria – Peloritani Orogen (Italy). Int. Geol. Rev. 50, 423–441.

Preziosi, E., Giuliano, G., Vivona, R., 2010. Natural background levels and threshold values derivation for naturally as, V and F rich groundwater bodies: a methodological case study in Central Italy. Environ. Earth Sci. 61 (5), 885–897.

Preziosi, E., Parrone, D., Del Bon, A., Ghergo, S., 2014. Natural background level assessment in groundwaters: probability plot versus preselection method. J. Geochem. Explor. 143, 43–53.

Quevauviller, P., 2005. Groundwater monitoring in the context of E legislation: reality and integration needs. J. Environ. Monit. 7, 89–102.

Rahman, A., Mondal, N.C., Fauzia, F., 2021. Arsenic enrichment and its natural background in groundwater at the proximity of active floodplains of Ganga River, northern India. Chemosphere 265, 129096.

Ravenscroft, P., Brammer, H., Richards, K., 2009. Arsenic Pollution: A Global Synthesis. John Wiley & Sons.

Rusi, S., Di Curzio, D., Palmucci, W., Petaccia, R., 2018. Detection of the natural origin hy-drocarbon contamination in carbonate aquifers (central Apennine, Italy). Environ. Sci.

Pollut. Res. 25 (16), 15577–15596. https://doi.org/10.1007/s11356-018-1769-9. Schabenberger, O., Gotway, C.A., 2005. Statistical Methods for Spatial Data Analysis. Taylor & Francis.

Sellerino, M., Forte, G., Ducci, D., 2019. Identification of the natural background levels in the phlaegrean fields groundwater body (Southern Italy). J. Geochem. Explor. 200, 181–192. https://doi.org/10.1016/j.gexplo.2019.02.007.

Shaddad, S.M., Buttafuoco, G., Castrignanò, A., 2020. Assessment and mapping of soil salinization risk in an egyptian field using a probabilistic approach. Agronomy 10 (1), 85. https://doi.org/10.3390/agronomy10010085.

Sinclair, A.J., 1974. Selection of threshold values in geochemical data using probability graphs. J. Geochem. Explor. 3 (2), 129–149.

1090 Sinclair, A.J., 1991. A fundamental approach to threshold estimation in exploration geochemistry:
1091 probability plots revisited. J. Geochem. Explor. 41 (1), 1–22.

1092 Singh, A., Maichle, R., 2015. ProUCL Version 5.1 User Guide Statistical Software for Envi-ronmental
1093 Applications for Data Sets With and Without Nondetect Observations. US EPA, Office of
1094 Research and Development.

1095 Smedley, P.L., Kinniburgh, D.G., 2002. A review of the source, behaviour and distribution of arsenic
1096 in natural waters. Appl. Geochem. 17 (5), 517–568.

1097 Tansi, C., Muto, F., Critelli, S., Iovine, G., 2007. Neogene-quaternary strike-slip tectonics in the
1098 central calabrian arc (southern Italy). J. Geodyn. 43, 393–414.

1099 Tisserand, D., Pili, E., Hellmann, R., Boullier, A.M., Charlet, L., 2014. Geogenic arsenic in
1100 groundwaters in the western Alps. J. Hydrol. 518, 317–325.

1101 Tortorici, L., 1982. Lineamenti geologico-strutturali dell'arco calabro-peloritano. Rendiconti SIMP.
1102 38, pp. 927–940.

1103 Tripodi, V., Muto, F., Brutto, F., Perri, F., Critelli, S., 2018. Neogene-quaternary evolution of the
1104 forearc and backarc regions between the serre and aspromonte massifs, Calabria (southern Italy).
1105 Mar. Pet. Geol. 95, 328–343. https://doi.org/10.1016/j.marpetgeo.2018.03.028.

1106 US EPA, 2002. Calculating Upper Confidence Limits for Exposure Point Concentrations at
1107 Hazardous Waste Sites.

1108 Van Dijk, J.P., Bello, M., Brancaleoni, G.P., Cantarella, G., Costa, V., Frixia, A., Golfetto, F., Merlini,
1109 S., Riva, M., Torricelli, S., Toscano, C., Zerilli, A., 2000. A regional structural model for the
1110 northern sector of the Calabrian Arc (southern Italy). Tectonophysics 324, 267–320.

1111 Vespasiano, G., Muto, F., Apollaro, C., 2021. Geochemical, geological and groundwater quality
1112 characterization of a complex geological framework: the case study of the Coreca Area (Calabria,
1113 South Italy). Geosciences 11 (3), 121. https://doi.org/10. 3390/geosciences11030121.

1114 Vespasiano, G., Cianflone, G., Romanazzi, A., Apollaro, C., Dominici, R., Polemio, M., De Rosa, R.,
1115 2019. A multidisciplinary approach for sustainable management of a com-plex coastal plain: the
1116 case of Sibari Plain (Southern Italy). Mar. Pet. Geol. 109, 740–759.

1117 Vespasiano, G., Cianflone, G., Cannata, C.B., Apollaro, C., Dominici, R., De Rosa, R., 2016. Analysis
1118 of groundwater pollution in the Sant'Eufemia plain (Calabria – south Italy). Ital. J. Eng. Geol.
1119 Environ. https://doi.org/10.4408/IJEGE.2016-02.O-01.

1120 Vespasiano, G., Apollaro, C., Marini, L., Dominici, R., Cianflone, G., Romanazzi, A., Polemio, M.,
1121 De Rosa, R., 2015a. Hydrogeological and isotopic study of the multiaquifer system of the Sibari
1122 Plain (Calabria, southern Italy). Rend. Online Soc. Geol. Ital. 39, 134–137.

1123 Vespasiano, G., Apollaro, C., De Rosa, R., Muto, F., Larosa, S., Fiebig, J., Mulch, A., Marini, L.,
1124 2015b. The small spring method (SSM) for the definition of stable isotope – elevation
1125 relationships in northern Calabria (Southern Italy). Appl. Geochem. 63, 333–346.

Vespasiano, G., Apollaro, C., Muto, F., De Rosa, R., Dotsika, E., Marini, L., 2015c. Preliminary geochemical characterization of the warm waters of the Grotta delle Ninfe near Cerchiara di Calabria (South Italy). Rend. Online Soc. Geol. Ital. 39, 130–133.

Vespasiano, G., Marini, L., Apollaro, C., De Rosa, R., 2015d. Preliminary geochemical characterization of the thermal waters of Caronte Spa springs (Calabria, South Italy). Rend. Online Soc. Geol. Ital. 39, 138–141.

Vespasiano, G., Apollaro, C., Muto, F., De Rosa, R., Critelli, T., 2015e. Preliminary geochemical and geological characterization of the thermal site of spezzano albanese (Cala-bria, South Italy). Rend. Online Soc. Geol. Ital. 33, 108–110.

Vespasiano, G., Apollaro, C., Muto, F., Dotsika, E., De Rosa, R., Marini, L., 2014. Chemical and isotopic characteristics of the warm and cold waters of the luigiane Spa near guardia piemontese (Calabria, Italy) in a complex faulted geological framework. Appl. Geochem. 41, 73–88.

Vespasiano, G., Muto, F., Apollaro, C., De Rosa, R., 2012a. Preliminary hydrogeochemical and geological characterization of the thermal aquifer in the Guardia Piemontese area (Calabria, South Italy). Rend. Online Soc. Geol. Ital. 21, 841–842.

Vespasiano, G., Apollaro, C., Muto, F., De Rosa, R., 2012b. Geochemical and hydrogeological characterization of the metamorphic-serpentinitic multiaquifer of the Scala catchment, Amantea (Calabria, South Italy). Rend. Online Soc. Geol. Ital. 21, 879–880.

Vessia, G., Di Curzio, D., Chiaudani, A., Rusi, S., 2020a. Regional rainfall threshold maps drawn through multivariate geostatistical techniques for shallow landslide early warning systems. Sci. Total Environ. 135815. https://doi.org/10.1016/j.scitotenv. 2019.135815.

Vessia, G., Di Curzio, D., Castrignanò, A., 2020b. 3D subsoil litho-technical characterization through data fusion of CPT parameters. Sci. Total Environ. 698, 134340. https://doi.org/10.1016/j.scitotenv.2019.134340.

Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction With Applications. Springer-Verlag https://doi.org/10.1007/978-3-662-05294-5.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons.

Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H.J., Wolter, R., 2005. Aproce-dure to define natural groundwater conditions of groundwater bodies in Germany. Water Sci. Technol. 51 (3–4), 249–257.

WHO, 2017. World Health Organization. Guidelines for drinking-water quality: fourth edition incorporating first addendum, 4th ed 1st add. https://apps.who.int/iris/ handle/10665/254637.

1163 Table 1

1164 Descriptive statistics of physico-chemical parameters of each water group.

Crystalline-metamorphic waters group (CM-group)

No. samples: 337

		Min	Max	Mean	I quartile	Median	III quartile	Skewness	Kurtosis
As	µg/L	0.01	435.01	4.69	0.07	0.13	0.35	10.24	116.60
Ca	mg/L	1.20	136.00	19.25	6.20	13.00	23.40	2.29	9.65
Cl	mg/L	2.70	313.90	14.49	8.00	11.60	16.10	12.33	192.10
HCO_3	mg/L	3.80	378.30	85.99	33.60	67.10	117.80	1.47	5.31
K	mg/L	0.20	52.20	1.71	0.90	1.20	1.80	12.77	194.00
Mg	mg/L	0.50	39.90	6.60	2.10	4.95	9.50	1.78	7.83
Na	mg/L	3.10	194.10	14.39	7.65	10.40	16.50	6.94	64.02
SO_4	mg/L	1.70	123.60	14.75	4.55	8.00	18.10	2.83	13.88
T	°C	4.30	24.10	12.73	10.60	12.60	14.90	0.35	3.23
pH	–	5.00	10.40	6.92	6.40	6.90	7.40	0.40	3.99
EC	µS/cm	13.00	1245.00	242.54	120.00	194.20	305.00	2.06	9.27
Eh	mV	-188.00	594.00	191.99	50.50	177.35	300.50	0.48	2.25

Ophiolite waters group (Oph-group)

No. samples: 41

Empty Cell		Min	Max	Mean	I quartile	Median	III quartile	Skewness	Kurtosis
As	µg/L	0.01	5.97	0.60	0.11	0.36	0.71	4.18	22.45
Ca	mg/L	2.10	124.70	49.18	32.20	49.90	64.70	0.42	3.67
Cl	mg/L	7.50	104.00	26.10	10.20	13.30	36.30	1.56	4.97
HCO_3	mg/L	16.80	538.30	235.52	149.50	210.50	348.70	0.46	2.48
K	mg/L	0.20	19.10	1.96	0.60	1.00	1.50	3.94	17.52
Mg	mg/L	1.30	64.20	21.77	4.20	14.10	40.00	0.52	1.89
Na	mg/L	5.90	98.40	23.52	8.20	12.00	32.60	1.70	5.27
SO_4	mg/L	2.90	212.10	33.83	6.10	14.60	44.80	2.55	9.76
T	°C	10.00	24.60	15.44	12.50	13.80	18.40	0.75	2.54
pH	–	6.10	8.20	7.44	7.20	7.40	7.80	-0.47	3.85
EC	µS/cm	91.20	1780.00	528.32	268.00	360.00	767.00	1.35	4.73
Eh	mV	-152.00	449.00	200.11	14.00	178.00	395.00	-0.10	1.69

Calcareous-dolomitic waters group (CD-group)

No. samples: 37

Empty Cell		Min	Max	Mean	I quartile	Median	III quartile	Skewness	Kurtosis
As	µg/L	0.03	0.99	0.41	0.19	0.36	0.58	0.73	2.53
Ca	mg/L	17.70	74.30	51.15	45.90	48.20	55.40	-0.10	4.14
Cl	mg/L	2.90	34.60	7.71	4.90	6.20	8.00	3.44	15.83
HCO_3	mg/L	51.90	339.30	230.26	209.00	230.30	250.20	-0.68	6.11
K	mg/L	0.40	3.30	0.71	0.50	0.50	0.70	3.67	18.06
Mg	mg/L	1.10	29.90	14.70	11.50	13.40	17.30	0.32	3.65
Na	mg/L	2.60	29.30	5.79	3.30	3.90	4.80	2.96	11.35
SO_4	mg/L	2.10	46.90	10.10	3.40	3.80	11.70	1.87	5.36
T	°C	8.00	18.60	12.29	10.40	11.40	14.00	0.78	2.95
pH	–	7.10	8.90	7.84	7.50	7.70	8.10	0.59	2.58
EC	µS/cm	122.00	660.00	382.67	335.00	369.50	429.00	0.10	4.11
Eh	mV	0.70	265.00	93.34	48.20	56.20	133.70	1.06	2.94

Evaporite waters group (Ev-group)

No. samples: 31

Empty Cell		Min	Max	Mean	I quartile	Median	III quartile	Skewness	Kurtosis
As	µg/L	0.05	2.70	0.65	0.14	0.36	1.11	1.25	4.08
Ca	mg/L	92.30	778.30	290.98	109.50	144.70	566.80	0.75	1.83
Cl	mg/L	6.60	965.80	78.65	14.40	33.00	56.20	4.34	21.69
HCO_3	mg/L	137.30	692.00	339.55	259.30	344.80	385.90	1.04	4.72
K	mg/L	0.40	24.70	4.19	1.40	2.50	3.80	2.65	9.54
Mg	mg/L	10.90	123.00	46.75	35.90	42.30	52.70	1.72	6.66
Na	mg/L	4.20	392.20	68.46	19.50	38.70	63.20	2.58	8.59
SO_4	mg/L	77.50	2345.80	767.83	190.50	343.80	1531.80	0.69	1.90
T	°C	11.40	21.50	16.65	14.80	16.45	19.10	0.02	2.14
pH	–	6.50	8.00	7.05	6.90	7.00	7.20	1.00	5.15
EC	µS/cm	798.00	4325.00	1692.39	1035.50	1155.50	2520.50	1.17	3.66
Eh	mV	-233.00	268.00	-23.03	-174.00	-28.00	77.30	0.36	2.03

Sedimentary waters group (Sed-group)

No. samples: 66

Empty Cell	Empty Cell	Min	Max	Mean	I quartile	Median	III quartile	Skewness	Kurtosis
As	µg/L	0.07	3.02	0.66	0.16	0.40	0.82	1.54	4.72
Ca	mg/L	1.60	164.10	54.53	23.30	48.15	74.30	0.79	3.06
Cl	mg/L	4.10	398.30	29.19	8.50	16.05	31.10	6.10	44.73
HCO_3	mg/L	11.80	517.10	204.44	105.30	187.65	292.90	0.55	2.60
K	mg/L	0.10	19.10	2.49	1.00	1.80	2.60	3.79	19.36
Mg	mg/L	0.90	43.70	14.03	6.50	12.90	18.50	1.08	3.88
Na	mg/L	2.50	212.50	22.75	9.80	15.15	27.30	4.78	31.12
SO_4	mg/L	3.80	169.60	39.16	13.60	29.00	44.30	1.53	4.72
T	°C	11.30	24.30	16.80	14.25	16.20	18.75	0.52	2.47
pH	–	5.90	8.70	7.39	7.10	7.45	7.80	-0.49	3.69
EC	µS/cm	75.00	1570.00	532.31	315.00	457.00	742.00	0.87	3.79
Eh	mV	1.00	284.00	148.84	77.00	180.00	187.90	-0.40	2.15

Thermal waters group (Th-group)

No. samples: 24

Empty Cell		Min	Max	Mean	I quartile	Median	III quartile	Skewness	Kurtosis
As	µg/L	0.40	17.40	5.70	1.50	3.29	8.02	1.02	2.56
Ca	mg/L	60.40	579.40	225.62	94.85	136.05	329.35	0.93	2.57
Cl	mg/L	34.70	3932.10	821.31	108.55	201.35	1226.95	1.54	4.38
HCO_3	mg/L	19.80	535.40	205.03	68.60	191.85	270.75	0.72	2.31
K	mg/L	0.30	38.70	10.79	4.25	9.35	13.20	1.51	5.47
Mg	mg/L	0.40	107.10	37.98	6.30	28.25	59.40	0.74	2.46
Na	mg/L	37.60	3401.50	586.37	131.40	265.15	648.00	2.17	7.49
SO_4	mg/L	8.90	4747.80	893.96	142.05	399.50	1425.95	2.02	7.56
T	°C	20.20	38.40	27.98	21.90	26.15	33.80	0.30	1.55
pH	–	6.60	8.60	7.53	7.15	7.50	7.95	0.06	2.18
EC	µS/cm	1019.00	13190.00	3785.25	1303.00	2708.00	3951.00	1.43	3.96
Eh	mV	-266.10	192.00	-12.51	-97.00	-14.80	60.85	-0.10	2.25

Table 2

1171 Aquifer-based natural background levels of arsenic-related to the six groups of waters
1172 assessed as the 95th percentile of the most representative population.

Group of waters	No. samples for NBL derivation	NBL As (µg/L)
Crystalline-metamorphic (CM-group)	52	7.08
Ophiolitic (Oph-group)	37	2.45
Calcareous-dolomitic (CD-group)	36	1.11
Evaporitic (Ev-group)	15	2.13
Sedimentary (Sed-group)	39	2.59
Thermal waters (Th-group)	24	21.90

1173

1174

1175

1176

1177 Table 3. The Pearson correlation matrix is compared to the whole dataset, and the two
1178 structural codispersion matrices, calculated by the Linear Model of Co-regionalization
1179 used in CK to estimate the As distribution in Fig. 3.

l	As	Ca	Cl	EC	Eh	HCO_3	K	Mg	Na	SO_4	T	pH
(a) Whole dataset												
As	1											
Ca	0.359	1										
Cl	0.315	0.467	1									
EC	0.433	**0.900**	0.626	1								
Eh	-0.292	-0.277	-0.102	-0.264	1							
HCO_3	0.293	**0.852**	0.363	**0.803**	-0.203	1						
K	0.338	0.322	0.531	0.408	-0.229	0.226	1					
Mg	0.305	**0.770**	0.477	**0.770**	-0.232	**0.833**	0.335	1				
Na	0.299	0.437	0.889	0.624	-0.142	0.358	0.618	0.430	1			
SO_4	0.361	**0.724**	**0.706**	**0.782**	-0.189	0.566	0.504	0.644	**0.720**	1		
T	0.361	0.489	0.630	0.608	-0.237	0.427	0.381	0.454	0.662	0.656	1	
pH	0.212	0.494	0.054	0.463	-0.125	0.483	0.049	0.357	0.032	0.241	0.178	1
(b) Short-range spherical structure (6600 m)												
As	1											
Ca	-0.194	1										
Cl	0.527	-0.607	1									
EC	-0.009	**0.812**	-0.569	1								
Eh	0.264	**-0.852**	0.443	-0.468	1							
HCO_3	-0.126	**0.957**	**-0.706**	**0.809**	**-0.737**	1						
K	0.591	0.173	0.155	0.082	-0.386	0.172	1					
Mg	0.018	**0.808**	-0.392	0.553	**-0.909**	**0.729**	0.578	1				
Na	0.572	-0.649	0.489	-0.537	0.626	-0.515	0.131	-0.456	1			
SO_4	0.183	0.439	-0.225	0.288	-0.449	0.408	0.200	0.564	-0.392	1		
T	**0.732**	-0.352	0.184	-0.146	0.382	-0.176	0.526	-0.019	0.656	0.114	1	
pH	-0.400	0.275	-0.393	0.089	-0.404	0.278	0.209	0.144	-0.330	-0.370	-0.400	1
(c) Long-range spherical structure (20,000 m)												
As	1											
Ca	**0.742**	1										

l	As	Ca	Cl	EC	Eh	HCO_3	K	Mg	Na	SO_4	T	pH
Cl	0.683	**0.792**	1									
EC	**0.787**	**0.939**	**0.870**	1								
Eh	-0.426	0.090	-0.240	-0.090	1							
HCO_3	0.662	**0.866**	**0.841**	**0.910**	0.027	1						
K	0.696	0.639	**0.828**	**0.722**	-0.449	0.632	1					
Mg	0.666	**0.880**	**0.886**	**0.971**	-0.076	**0.938**	**0.747**	1				
Na	0.648	**0.766**	**0.940**	**0.850**	-0.294	**0.821**	**0.922**	**0.903**	1			
SO_4	**0.790**	**0.915**	**0.794**	**0.937**	-0.070	**0.875**	**0.803**	**0.915**	**0.836**	1		
T	0.593	0.683	**0.756**	**0.779**	-0.442	**0.758**	0.643	**0.811**	**0.768**	**0.713**	1	
pH	0.488	**0.726**	0.308	0.612	0.536	0.653	0.096	0.521	0.224	0.635	0.209	1

Bold underlined values represent high significant correlation; underlined values represent significant correlations.

Table 4. Cross-validation results related to the LMC used in the CK to estimate the spatial distribution of arsenic in the study area.

Variable	ME	RMSE	MSE	RMSSE
gAs	0.0052	0.7512	0.0014	0.9772

Table 5. Variogram model and LMC related to the four combinations of NBL values and methodological approaches.

NBL reference for the indicator variable	Method	Structure	Range (m)
Regional NBL	Indicator Kriging	Spherical	8800
Aquifer-based NBL	Indicator Kriging	Spherical	7500
Regional NBL	Probability Kriging	Short-range spherical	6600
		Long-range spherical	20000
Differentiated NBL	Probability Kriging	Short-range spherical	6600
		Long-range spherical	20000

Table 6. Comparison among the cross-validation results related to the four combinations of NBL values and methodological approaches.

Variable	NBL value	Method	ME	RMSE	MSE	RMSSE
Indicator NBL of As	Regional	IK	0.0005	0.1933	-0.0009	0.9227
	Aquifer-based	IK	-0.0011	0.1871	-0.0070	0.9556
	Regional	PK	0.0024	0.1937	0.0085	0.9367
	Aquifer-based	PK	-0.0009	0.1871	-0.0053	0.9690