GEOSTATISTICS AS A TOOL TO IMPROVE THE NATURAL BACKGROUND LEVEL DEFINITION: AN APPLICATION IN GROUNDWATER

Dalla Libera Nico [a)], Fabbri Paolo [a(*)], Mason Leonardo [b], Piccinini Leonardo [a], Pola Marco [a]

[a] Department of Geosciences, University of Padova, Padova, Via G. Gradenigo, 6 – 35131, Italy

[b] ARPAV, Department of Venice, Mestre, Via Lissa, 6 – 30171, Italy

ABSTRACT

The UE BRIDGE project suggests calculating the Natural Background Level (NBL) as the 90th percentile of the distribution of the concentration data. This method is suited for large, spatially distributed datasets providing a regional value of NBL that could be higher than the Threshold Value (TV) set by every country. Therefore, the use of a unique regional NBL value, higher than TV, in dis-homogeneous areas from the hydro-geochemical perspective, could arise problems to distinguish between natural occurrences and anthropogenic contaminant sources. Hence, the goal of this study is to improve the NBL definition using the geostatistical approach, which allows reconstructing the contaminant spatial structure keeping in account geochemical and hydrogeological relationships. We retain this integrate mapping fundamental to evaluate the contaminant's distribution impact on the NBL value, giving indications to improve it. We decided to test this method on the Drainage Basin of Venice Lagoon (DBVL), an area in northeastern Italy notoriously affected by naturally occurring arsenic contamination, where the existing NBL is seven times higher than the TV. An available geochemical dataset collected by 50 piezometers was used

to reconstruct the spatial distribution of arsenic in the densely populated area of the DBVL. A cokriging approach was applied exploiting the geochemical relationships among As, Fe and NH_4^+. The obtained spatial predictions of arsenic concentrations was divided into three different zones: i) areas with an As concentration lower than the TV, ii) areas with an As concentration between the TV and the median of the values higher than the TV, and iii) areas with an As concentration higher than the median. Subsequently, following the BRIDGE suggestions, where enough samples were available, the 90th percentile for each zone was calculated to obtain a local NBL (LNBL). Differently from the original NBL, this local value gives more detailed water quality information accounting the hydrogeological and geochemical setting, and contaminant spatial variation. Hence, the definition of a LNBL could give more indications about the distinction between natural occurrence and anthropogenic contamination.

Keywords: Local Natural Background Level (LNBL), cokriging (COK), Arsenic and Drainage Basin to the Venice Lagoon (DBVL).

1. INTRODUCTION

The assessment of the groundwater quality and the impact of the human activities are important worldwide challenges. Actually, the assessment of groundwater body qualitative status is related to the definition of Natural Background levels (NBLs) and Threshold values (TVs). The first one is mainly linked to the system's hydro-geochemical settings, while the second is associated to the public health issue. The NBL definition is a well-known problem due to its complexity, its challenging estimation and the delicate consequences on the environmental protection. The NBL of an element or compound in groundwater represents the range of concentrations resulting from the interaction of different natural atmospheric, geological, chemical and biological process during the hydrological cycle, and it could be influenced by human activities (Edmunds and Shand, 2008; Hinsby et al., 2008a; Reimann and Garrett, 2005). By defining the NBL, it is necessary to know the extension and the hydro-geochemical characteristics of the groundwater body at risk, paying

attention to use data coming from the same aquifer body and collected far from anthropogenic sources of contamination, in order to obtain pristine water samples(Coetsier et al., 2009; Hinsby et al., 2008; Wendland et al., 2008). Many studies treat the problem of the NBL definition or the definition of new approaches to improve it (Coetsier et al., 2009; Ducci et al., 2016; Molinari et al., 2012; Preziosi et al., 2010; Rotiroti et al., 2013, 2015), since the NBL assumes importance discriminating the anthropogenic contamination from the natural occurrence. A first approach to estimate the NBL value into European groundwater bodies was suggested by Edmunds and Shand (2008), and it was based on a geochemical prospective rather than a statistical one. Actually, the European Community in accordance with the European "BRIDGE" project (Background cRiteria for the Identification of Groundwater thrEsholds) (Müller et al., 2006) suggests two main statistical method to estimate the NBL. The first approach, the "Component Separation" method (CP), is based on the distinguishable data distributions produced by the natural and anthropogenic sources that can be distinguished by a statistical analysis. In this approach, the observed concentration frequency distribution is fitted by the superimposition of two different distributions that represent the natural and the anthropogenic concentrations. Once estimate the distribution (shape and statistical parameters) of the natural component, the data following it will be used to estimate the NBL. On the other hand, the second method suggests a pre-selection of the data basing on some markers' concentration that point an anthropogenic contamination out (e.g. nitrate, salinity, etc.). The basic idea of this method is that there is a correlation between the markers' concentrations and the presence of an anthropogenic pressure. Thus, where the markers show concentration values over the pre-defined limits, the groundwater samples are excluded from the NBL estimation. Subsequently, for both methods, the NBL is estimated as the 90th or 97th percentile of the modified distribution of the concentration data (Coetsiers et al., 2009; Müller et al., 2006; Wendland et al., 2008b). These approaches give a unique NBL value for the entire area, which seems to be not capable to represent the local variation of the geochemical and environmental conditions. The definition of a unique NBL value for a large area, in fact, could make a considerable uncertainty

defining the contaminant's natural occurrence against an anthropogenic source. Considering both the importance of the health problem related to the contaminant natural occurrence and the complexity to define an exhaustive NBL value, this study aims to improve the NBL concept by using geostatistical methodology such as cokriging predictor (COK). This method, like other types of kriging, allows redefining the NBL as function of the spatial distribution of contaminant concentration. However, the COK keeps also in consideration the relationships among the target contaminant and other, geochemically related, dissolved species in groundwater. This methodology is applied on the case of the Drainage Basin of the Venice lagoon (DBVL) notoriously affected by both high natural and anthropic concentrations of arsenic. Furthermore, the Regional Agency for Environmental Protection and Prevention of Veneto (ARPAV) through the "A.Li.Na" project (ARPAV, 2014) estimated an arsenic NBL of 74 μg/L that is seven times higher than the Threshold Value (TV=10 μg/L) defined by the annex III in the Commission Staff Working report of the EU's Groundwater Directive (GWD 2006/118/EC). However, the results of this project pointed out the importance of a detailed and local NBL accounting the complex and laterally variable geological and hydro-chemical settings. The results of our study could be useful to improve the definition of the NBL because it exploit the knowledge about the spatial structure and uncertainty of the naturally occurring contaminant into groundwater, giving a local detail that considers the environmental background. In addition, they could be used by the stakeholders to perform a more complete and explanatory plan for the environmental management.

2. KNOWLEDGE ABOUT ARSENIC CONTAMINATION

High arsenic concentrations in groundwater have been documented in several countries, such as Bangladesh, India (McArthur et al., 2001; Nickson et al., 1998, 2000), Vietnam (Fendorf et al., 2010), Hungary and the USA (Sorg et al., 2014; Welch et al., 2000). The natural occurrence of arsenic in these countries is often linked to the alluvial plain systems, but high arsenic concentrations can be found in volcanic environments or mining districts (e.g., Argentina, Chile,

Brazil, Ethiopia, Greece and Italy) (Nordstrom, 2003; Rango et al., 2013). In Italy, high natural concentrations of arsenic in groundwater have been documented in volcanic aquifers of the central and southern Italy related to the uprising of geothermal fluids (Baiocchi et al., 2011; Ducci et al., 2016; Preziosi et al., 2010). Furthermore, in northern Italy, a high arsenic concentration can be found in the alluvial aquifers of the Po Plain (Molinari et al., 2013; Rotiroti and Fumagalli, 2013; Rotiroti et al., 2014; Zavatti et al., 1995) and the Venetian Plain (Carraro et al., 2013; Ungaro et al., 2008). Few materials (e.g., gold or sulfide mineral deposits, volcanogenic sources, alluvial sediments and organic matter) are currently recognized as significant sources of arsenic or drivers for its mobilization in groundwater. Nevertheless, the occurrence of arsenic in groundwater could be caused by anthropic activities, such as mineral extraction, industrial processes or fertilizer use. Arsenic is a metalloid that has three oxidation states nevertheless the most common forms in groundwater are the followings: the first is oxidized with an oxidation number (ON) equal to + 5, whereas the second is reduced with an ON equal to + 3. The reduced form appears to be more movable than the oxidized form in an aqueous medium, although both species have good mobility. The arsenic toxicity is due to its affinity to Phosphorus (P) that could be replaced by arsenic (As) in the metabolic processes of RNA and DNA synthesis. Arsenic is not widespread in the Earth's continental crust but is commonly concentrated in sulfide-bearing mineral deposits associated with Pyrite or Fe-hydroxides. Furthermore, many geothermal waters naturally exceed the TV of 10 μg/L. As determined in many studies, arsenic often is adsorbed onto Fe-Mn oxides or hydroxides, as well as into the organic matter (Baviskar et al., 2015; Carraro et al., 2015; McArthur et al., 2004, 2001, Molinari et al., 2015, 2013; Nickson et al., 2000; Rowland et al., 2007). Dissolution of these minerals in an acid-reducing environment causes the desorption processes that enable the release of As in an aqueous medium, exceeding its threshold value (McArthur et al., 2016; Smedley and Kinniburgh, 2002). Therefore, the organic matter could be a source of As, in addition to its role as a redox driver (Rotiroti et al., 2014). In particular, the release of As in groundwater is influenced by the variation of physical and chemical parameters, such as the Oxidation-Reduction Potential (ORP)

and pH (Smedley and Kinniburgh, 2002; Sorg et al., 2014), as well as the presence of elements and compounds that can promote its release (e.g., NO_3^-, NH_4^+, HCO_3^-, SO_4^{2-}, PO_4^{3-}) (Biswas et al., 2014).

3. GEOLOGICAL AND HYDROGEOLOGICAL SETTINGS

The study area is located in the middle-low Venetian Plain, including the Padua, Treviso and Venice provinces. The area covers approximately 2038 km^2 with a topographic gradient ranging from 0.6% to 0.1% near the Venice lagoon (Figure 1A, B). Two primary alluvial hydrogeological units occur in the Venetian plain: a large unconfined aquifer extending 15-20 km in the upper region of the plain from the foot of Prealps, and a multi-layered confined aquifer system in the lower region of the plain towards the Adriatic Sea. These two units correspond, respectively, to the upper and middle-lower plain environments. The plain spring's belt (named the "Fontanili" belt) shows the transition from the upper plain to the lower plain, where the water table is very shallow and locally intersects the topographic surface. From a geological perspective, the study area comprises gravelly and sandy alluvial deposits in the upper region near the "Fontanili" belt and silty-clayey deposits in the distal region (upper Pleistocene – Holocene) (Bondesan et al., 2004; Fontana et al., 2004, 2008; Mozzi et al., 2003). These fine sediments often contain peat layers formed in a floodplain environment, with a thickness of decimeters and a lateral extension of kilometers. The Drainage Basin to the Venice Lagoon (DBVL) extends to the lower plain environment, alternating silty layers with low permeability and sandy permeable layers, where the primary aquifers are located (Cambruzzi et al., 2009; Dal Prà et al., 1992; Fabbri and Piccinini, 2013; Fabbri et al., 2011, 2013, 2016; Piccinini et al., 2015, 2016; Vorlicek et al., 2004) (Figure 1C). Anyway, the subsoil structure of the DBVL is heterogeneous, due to the complex alluvial sedimentation processes that characterized the Brenta's alluvial plain (Trevisani and Fabbri, 2010) and surely influences the spatial distribution of As-bearing materials. This is important because Arsenic may be concentrated in sulfide-minerals, such as Pyrite (FeS_2), or adsorbed onto oxides or hydroxides composing the fine portion of the alluvial sediments (Baviskar et al., 2015; Rowland et

al., 2007). Therefore, the peat layers play an important role in the arsenic release mechanism because the degradation of the organic matter controls the aquifer's redox conditions and the thus the dissolution of these As bearing minerals (Rotiroti et al., 2014).

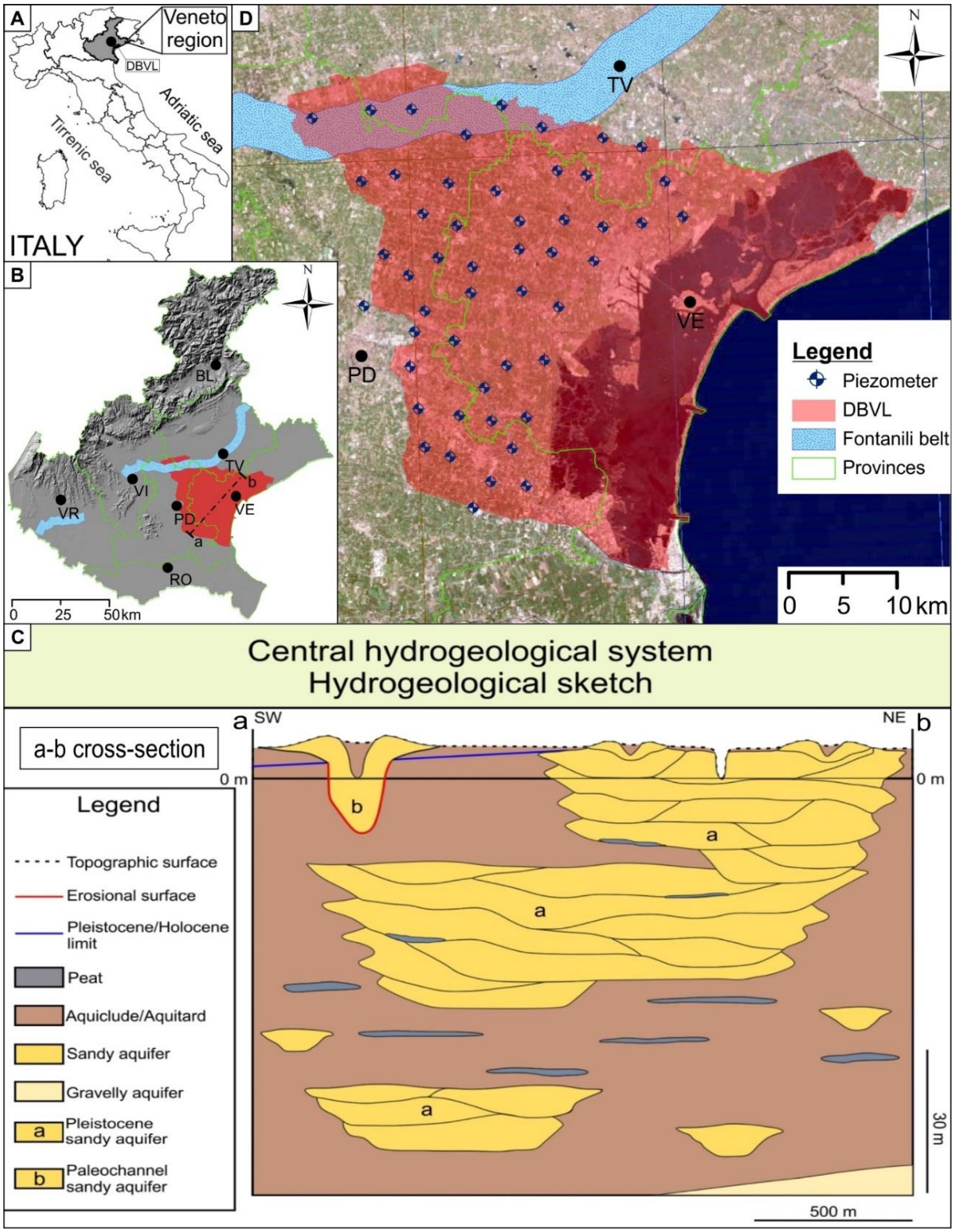

Figure 1. A) The location of the Veneto region in Italy; B) The location of the study area in the Veneto region, which covers the entire Brenta's Plain; C) A hydrogeological sketch of the low Brenta's Plain along the "a-b cross-section"; D) The piezometer locations in the DBVL. All of the piezometers are in the shallow aquifer, between 10 and 20 meter below ground level (BGL).

4. MATERIALS AND METHODS

4.1.Dataset

The data used in this work was obtained from the "A.Li.Na" project, which was developed to estimate the natural background levels (NBLs) of As, Fe, Mn and NH_4^+ in the groundwater of the DBVL, according to the pre-selection method suggested by BRIDGE project. The hydro-geochemical sampling network (50 piezometers) was established with a variable spacing grid, covering all of the investigated area (Figure 1D). The sampled piezometers, used for NBL definition, are located far from the anthropogenic pollutant sources in order to obtain some pristine water samples. The piezometers tap the shallow aquifer at a depth of 10 – 20 meters below ground level (BGL). The hydro-geochemical parameters were collected during four seasonal surveys conducted from 2013 to 2014. The groundwater sampling was performed according to the standard protocols established by the Italian Environment Protection and Technical Services Agency (APAT, 2006). In addition, the physical parameters (e.g., pH, ORP, electric conductivity, temperature and dissolved oxygen) were measured by a multi-parametric probe YSI mod. 556 MPS. The chemical parameters were estimated in the laboratory through standardized methods. Arsenic and Manganese were estimated by the UNI EN ISO 17294-2:2005 (revised by ISO 17294-2:2016) method, Iron by the APAT CNR IRSA 3160 Man 29 2003 method and Ammonium by the APAT CNR IRSA 4030 A1 Man 29 2003 method (APAT, 2003). Furthermore, each sampling point was georeferenced in the Gauss-Boaga coordinate system (Roma 1940 datum, West zone). According to BRIDGE suggestions, the data were processed to estimate the average values from each available concentration time series, in order to create an average dataset that depicts the average system

behavior. Therefore, an average dataset was determined, depicting. Afterwards, the concentration values below the limit of quantification (i.e., As < 1 μg/L) were changed to make the entire "A.Li.Na" dataset suitable for mapping purposes. In such a case, the As concentrations below this limit were set equal to 0.5 μg/L. The DBVL shows high concentrations of arsenic in reducing environments and often in association with rich organic matter reservoirs (Carraro et al., 2015; Molinari et al., 2015; Rotiroti et al., 2014), according with the geological genesis of the area. Looking at the Figure 2, the major part of the arsenic data show an ORP values from 0 mV to -150 mV and a pH between 7 and 8 according with the releasing as reduced forms.

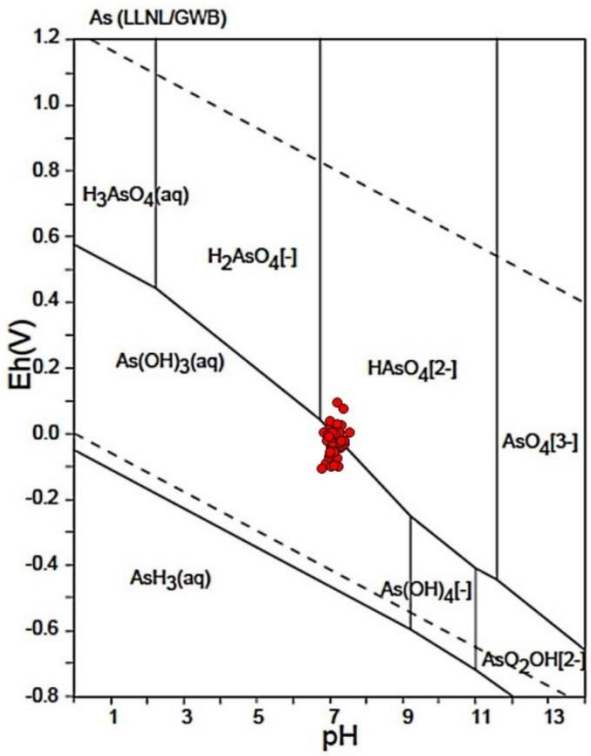

Figure 2. A Pourbaix diagram of Arsenic species built under the conditions of T = 298.15 [°K] and P = 10^5 [Pa]. The red dots represent the experimental data used for this study. The majority of the sample data falls into the arsenic reduced species stability field (Takeno, 2005).

4.2.Geostatistical methods

The arsenic concentration in groundwater is linked to complex hydro-geochemical processes not yet well understood, hence its distribution in groundwater can be more usefully approached by geostatistical methods than by deterministic ones (e.g., Inverse Distance Weighting and Natural Neighbors). The geostatistical kriging predictor is based on a mixed model [1]:

$$\hat{Z}(\boldsymbol{u_i}) = R(\boldsymbol{u_i}) + m(\boldsymbol{u_i}) \quad [1]$$

where the predicted concentration $\hat{Z}(\mathbf{u_i})$, at location $\mathbf{u_i}$, is composed of a residual component $R(\mathbf{u_i})$ plus the mean $m(\mathbf{u_i})$, representing, respectively, the stochastic component and the deterministic component of the predicted geostatistical concentration. Moreover, the kriging method needs to assume the stationary condition, considering the spatial correlation of the variable Z independent of their spatial position ($\mathbf{u_i}$) but dependent only on the separation distance (h)(Bivand et al., 2008; Isaak and Srivastava, 1989; Posa and De Iaco, 2009). Because the kriging predictor was built according to Equation [1], the problem is to define the mean m dependent on the position ($\mathbf{u_i}$). Depending on the mean characteristics, three types of predictors can be considered as follows: i) Simple (co)kriging ($m(u_i)=m$, the mean is known and constant), ii) Ordinary (co)kriging ($m(u_i)=m$, the mean is unknown but constant), and iii) Universal (co)kriging ($m(\mathbf{u_i})$, the mean is not constant but defined by a spatial function). In the context of this study, the more acceptable condition was a mean that was constant within a moving search neighborhood but unknown, and thus, the selected prediction was an ordinary kriging approximation of $\hat{Z}(\boldsymbol{u_o})$.

$$\hat{Z}(\boldsymbol{u_o}) = \sum_{i=1}^{n} \omega(\boldsymbol{u_i})\, Z(\boldsymbol{u_i}) \quad [2]$$

If auxiliary variables correlated to the primary variable are available, a cokriging approach can be considered. Theoretically, it gives a prediction enhancement respect to ordinary kriging application. Cokriging is based on the same principles of kriging but also uses correlated auxiliary variables (in our case Fe and NH_4^+) to predict the primary variable (in our case As).

$$\hat{Z}(\boldsymbol{u_o}) = \sum_{i=1}^{n} \omega(\boldsymbol{u_i})\, Z(\boldsymbol{u_i}) + \sum_{j=1}^{n} \tau\left(\boldsymbol{u_j}\right) V\left(\boldsymbol{u_j}\right) \quad [3].$$

As evidenced in Equation [3], the cokriging predictor uses a weighted linear combination of the principal variable Z($\mathbf{u_i}$) and of the auxiliary variables V($\mathbf{u_j}$) in the different position ($\mathbf{u_i}$) and ($\mathbf{u_j}$). In this manner, improving the principal variable prediction is possible. Using the matrix notation the cokriging system is written as

$$\boldsymbol{K_{CK}}\, \boldsymbol{L_{CK}}(\boldsymbol{u}) = \boldsymbol{k_{CK}} \quad [4].$$

Where $\mathbf{K_{CK}}$ is the n($\mathbf{u}$) x n($\mathbf{u}$) matrix of data covariances, $\mathbf{L_{CK}}$(u) is the vector of the weights ω($\mathbf{u_i}$) and τ($\mathbf{u_j}$), $\mathbf{k_{CK}}$ represents the vector of data-to-unknown covariances. Basing on the [4], the cokriging weights required by the COK estimator [3] are obtained by multiplying the inverse of the data covariance matrix by the vector of data-to-unknown covariances:

$$\boldsymbol{L_{CK}}(\boldsymbol{u}) = \boldsymbol{K_{CK}^{-1}}\, \boldsymbol{k_{CK}} \quad [5].$$

Furthermore, to satisfy the stationary and unbiased conditions, the estimated weights from [5] have to undergo the following constrains:

$$\begin{cases} \sum_{i=1}^{n} \omega_i(u) = 1, \\ \sum_{j=1}^{n} \tau_j(u) = 0. \end{cases} \quad [6].$$

As reported in literature, cokriging method work better when the primary variable is less sampled respect to the auxiliary one. Furthermore, one could also use it when main and auxiliary variable are sampled in the same location (collocated cokriging) even if it gives lower prediction advantage than the classical application. Nevertheless, in our study we consider the collocated cokriging method because it produces a reduction on prediction error around 10% respect to OK. The geostatistical analysis was conducted in the R environment (R Core Team, 2015), in particular with the gstat package (Pebesma, 2004). The logical process (Figure 3) to estimate the arsenic spatial distribution

was composed of four primary steps: Exploratory Data Analysis, Variographic Analysis, Cross-validation, Ordinary cokriging.

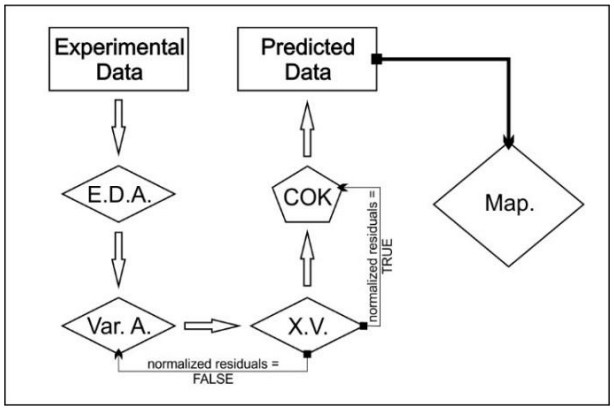

Figure 3. The flow chart of the classical logical process implemented in the geostatistical analysis. E.D.A = Exploratory Data Analysis, Var.A. = Variographic Analysis, X.V. = Cross–Validation, COK = Ordinary cokriging, Map. = Mapping the predicted principal variable.

5. RESULTS AND DISCUSSION

5.1.Geostatistical analysis

According to the goal of this study, the average dataset was considered. Initially, a study of the correlations between arsenic and other related chemical components (i.e., iron and ammonium) was performed. The robust correlation among As and other parameters indicated that the cokriging (COK) approach might be appropriate. To evaluate the correlation, we used a Spearman's correlation test and scatterplots. The Spearman's correlation test was selected because the experimental data do not have a normal distribution and the relation among the variables is not linear, so the Pearson's coefficient could be less exhaustive (Isaak and Srivastava, 1989). In that case, the Spearman's test, basing on ranks correlation, allows evaluating the strength of the correlation among the variables even if it is not linear. In addition, the scatterplot graphically explores linear correlations. The relationships among As, Fe and NH_4^+ are illustrated in Figure 4 in which the scatterplots show the poor linear correlation between two variables emphasized by the local regression function LOESS (red line) and the numerical coefficients represent the Spearman's ρ. The juxtaposition of these elements allow showing the poor relevance of the linear correlation on our data. The Spearman's ρ is greater than 0.5 and statistically meaningful (p-value < 5% significance level). Based on these relationships, we decided to use Fe and NH_4^+ as the auxiliary variables in the COK prediction process of As. Subsequently, histograms, scatterplots explored the statistical structure of the raw data. In this manner, estimating the statistical parameters of the data (i.e., means, standard deviation, variance, skewness, etc.) was possible. Arsenic, iron and ammonium data followed a lognormal distribution with positive skewness coefficients (3.13, 2.20, 1.87, respectively; Figure 5). The raw data were transformed to a normal score in order to perform a COK prediction with lower error as possible. Furthermore, the normal score transformation allows an easier back-transformation of the processed data, giving lower error respect to the use of log-transformation.

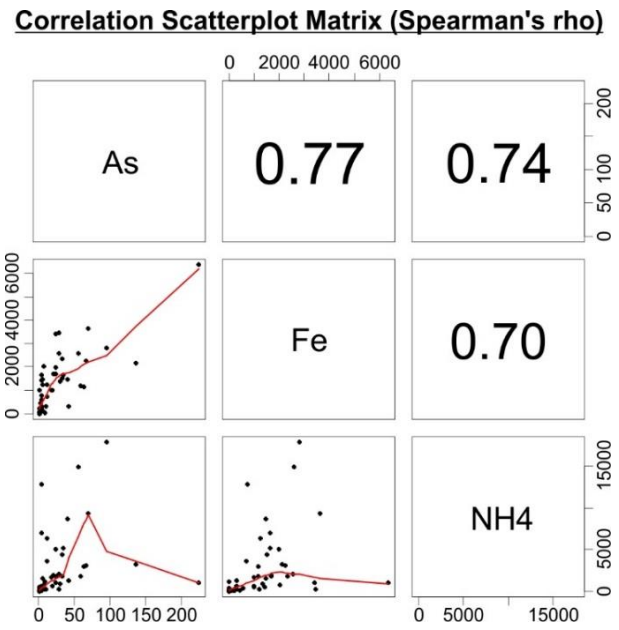

Figure 4. A correlation scatterplot matrix based on Spearman's method, providing the ρ coefficients for the pairs As-Fe, As-NH_4^+ and Fe-NH_4^+. In the lower part of the matrix, the scatterplot related to the variable pairs are shown. They represent the poor linear correlation emphasized by the red line that represents the local regression function LOESS (Cleveland, 1979). The correlation coefficients ρ are shown in the upper part of the matrix. The used dataset is the average of the four surveys and the Spearman's correlation test was conducted with the experimental concentration variables.

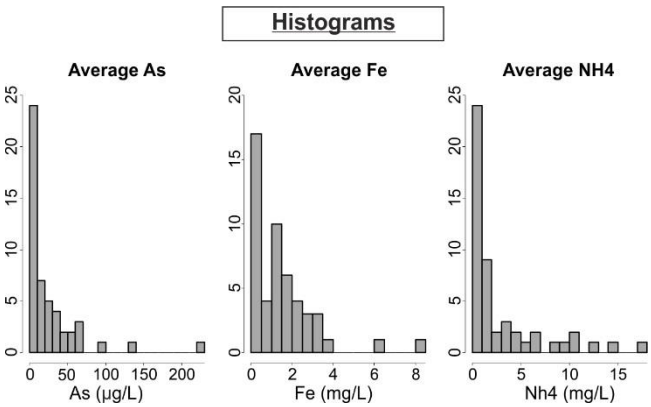

Figure 5. The three histograms represent the distributions of the average data of As, Fe and NH^{4+}, respectively.

The second step of the geostatistical process was the Variographic Analysis of the transformed data. The variogram γ(h) describes how spatial continuity changes with the distance (h) and the direction. In addition, the cokriging approach requires the variogram and the cross-variogram analyses. An analysis of the surface variograms in Figure 6A did not reveal any anisotropy in the continuity structures of the studied variables. Considering this, only omnidirectional variograms and cross-variograms were considered (Figure 6B). A Linear Model of Co-regionalization (LMC) fits a theoretical variogram model with both experimental variograms and cross-variograms. The LMC provides authorized theoretical models of variograms and cross-variograms of two or more variables so that the variance of any possible linear combination of these variables is always positive (Isaak and Srivastava, 1989). The resulting theoretical variograms (Figure 6) were used during the cokriging prediction. The variogram model used in this study was spherical both for auto- and cross-variograms:

$$\gamma(h) = \begin{cases} nu + c; & |h| > a \\ nu + c\left[\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|}{a}\right)^3\right]; & |h| \leq a \end{cases} \qquad [7].$$

where nu is the nugget effect, c represents the partial sill and a represents the range. The variogram models referred to As, Fe and NH_4^+, and the three cross-variogram models of As-Fe, As-NH_4^+ and Fe-NH_4^+ are visible in Table 1.

Table 1. The variogram and cross-variogram models used to predict the arsenic spatial distribution.

Variogram model	Parameter
$\gamma(h) = \begin{cases} 0.486 + 0.523\,; & \lvert h\rvert > 8000 \\ 0.486 + 0.523\left[\frac{3}{2}\frac{\lvert h\rvert}{8000} - \frac{1}{2}\left(\frac{\lvert h\rvert}{8000}\right)^3\right]; & \lvert h\rvert \leq 8000 \end{cases}$	[As]
$\gamma(h) = \begin{cases} 0.730 + 0.260\,; & \lvert h\rvert > 8000 \\ 0.730 + 0.260\left[\frac{3}{2}\frac{\lvert h\rvert}{8000} - \frac{1}{2}\left(\frac{\lvert h\rvert}{8000}\right)^3\right]; & \lvert h\rvert \leq 8000 \end{cases}$	[Fe]

$\gamma(h) = \begin{cases} 0.055 + 0.884\,; & \lvert h\rvert > 8000 \\ 0.055 + 0.884\left[\frac{3}{2}\frac{\lvert h\rvert}{8000} - \frac{1}{2}\left(\frac{\lvert h\rvert}{8000}\right)^3\right]; & \lvert h\rvert \le 8000 \end{cases}$	[NH_4^+]
$\gamma(h) = \begin{cases} 0.581 + 0.148\,; & \lvert h\rvert > 8000 \\ 0.581 + 0.148\left[\frac{3}{2}\frac{\lvert h\rvert}{8000} - \frac{1}{2}\left(\frac{\lvert h\rvert}{8000}\right)^3\right]; & \lvert h\rvert \le 8000 \end{cases}$	[As-Fe]
$\gamma(h) = \begin{cases} 0.034 + 0.521\,; & \lvert h\rvert > 8000 \\ 0.034 + 0.521\left[\frac{3}{2}\frac{\lvert h\rvert}{8000} - \frac{1}{2}\left(\frac{\lvert h\rvert}{8000}\right)^3\right]; & \lvert h\rvert \le 8000 \end{cases}$	[As-NH_4^+]
$\gamma(h) = \begin{cases} -0.004 + 0.429\,; & \lvert h\rvert > 8000 \\ -0.004 + 0.429\left[\frac{3}{2}\frac{\lvert h\rvert}{8000} - \frac{1}{2}\left(\frac{\lvert h\rvert}{8000}\right)^3\right]; & \lvert h\rvert \le 8000 \end{cases}$	[Fe-NH_4^+]

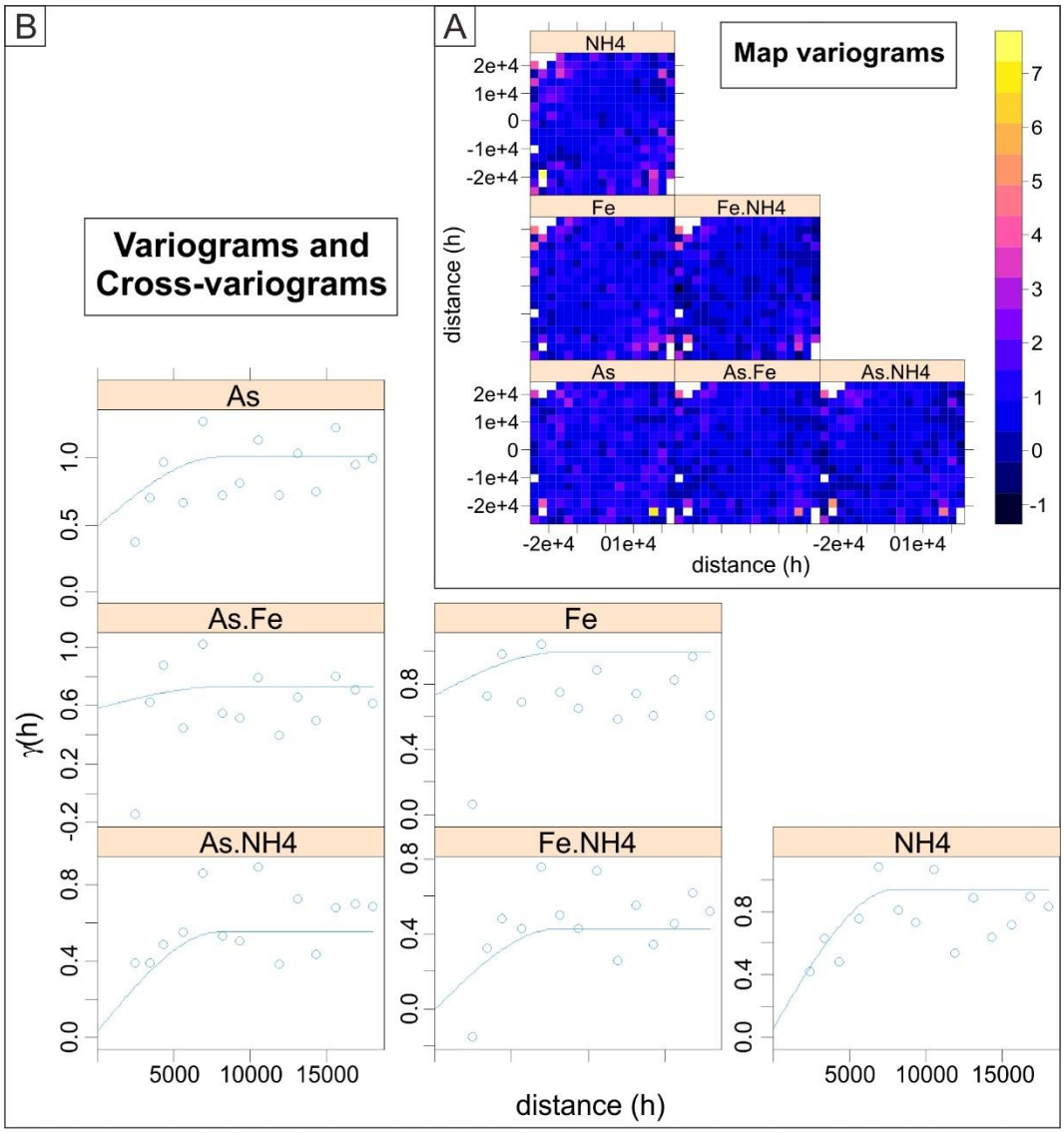

Figure 6. A) The surface variograms and cross-variograms. The surface variograms are shown in the first column. B) The omnidirectional variogram and cross-variograms. The variograms are on the diagonal, and the associated cross-variograms (referred to the average values of surveys) are below the diagonal.

Following the above variogram analysis, a cross-validation procedure was applied. The cross-validation tests the ability of the model to reproduce the spatial continuity structure by the prediction of the used dataset. The cross-validation, applied on our data, gives a mean estimation error equal to 0.001 µg/L, confirming the good prediction given by the chosen variogram model (Isaak and Srivastava, 1989). Looking the Figure 7, it is evident the normality of the estimation errors and the good reproduction of the experimental data, highlighting the model's goodness. Furthermore, in our case, the cross-validation resulted in an acceptable normalized root mean square error (nRMSE) equal to 13.3%, which is around 10% lower than the nRMSE obtained by the ordinary kriging prediction on the same data. A similar nRMSE is likely linked to the presence of punctual high values of arsenic (hot spots), influencing the prediction process. Generally, the problems linked to the hot spots are common in environmental datasets.

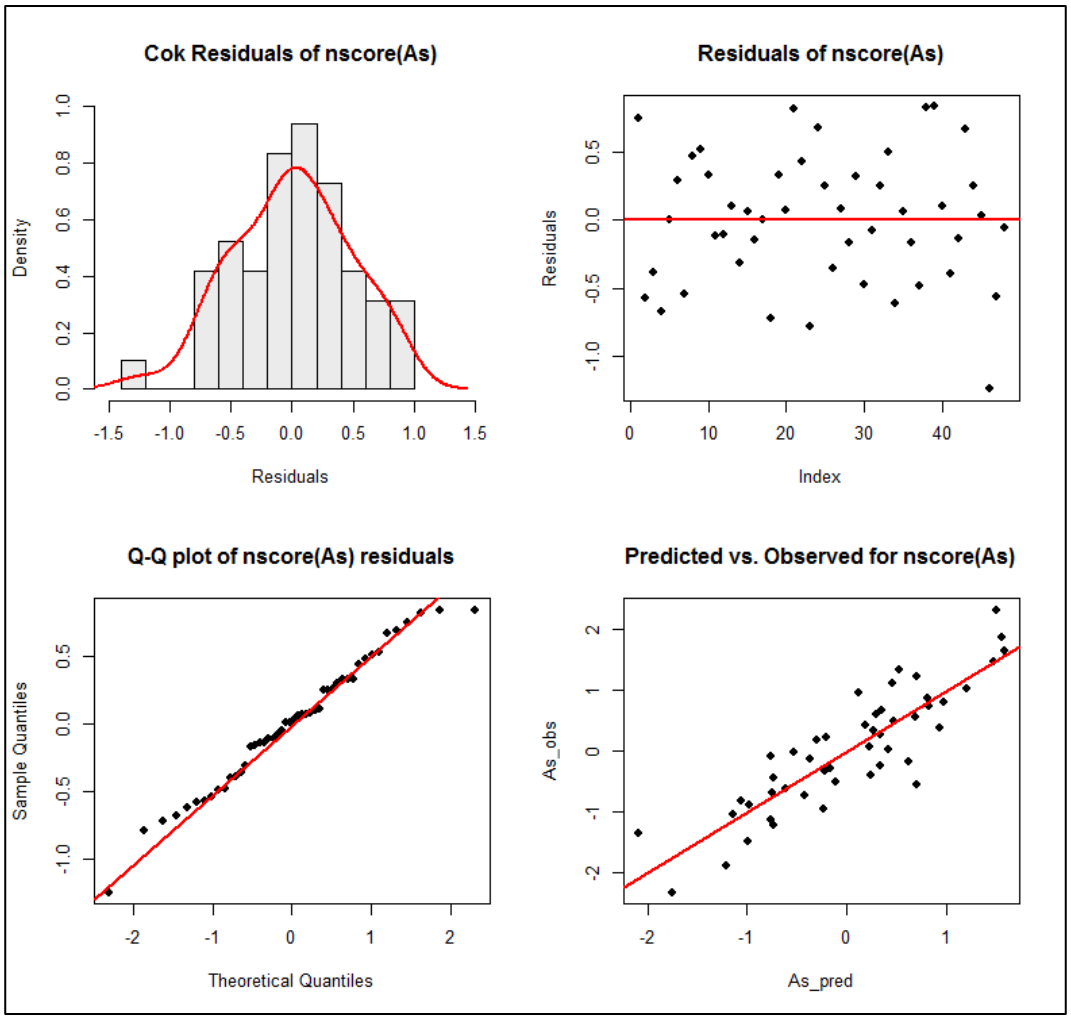

Figure 7: Graphical representation of the arsenic residuals coming from the cross-validation process by assessing the goodness of the chosen spatial model. Moving from the upper left corner toward the lower right one, there are the probability density function of residuals, the distribution around the zero value, the QQ-norm plot and the scatterplot observed As vs. predicted As. In the last diagram, the red line represents the linear trend line (correlation coeff. = 0.67)

5.2.Arsenic distribution

Following the previous geostatistical analysis, an ordinary cokriging predictor was applied. The prediction process allowed mapping the arsenic distribution on the shallow groundwater of the DBVL (Figure 8). By improving the NBL concept and getting a more detailed value, we decided to consider the TV and the median (M) of the original dataset as thresholds to reclassify the distribution map. The reclassification process is necessary to identify the data that have similar arsenic values and then estimate some Local NBL following the 90th percentile method. In our case, the median was chosen because the "A.Li.Na" NBL value is several times higher than the TV and it seems adequate as intermediate threshold, separating the 50% of the data. However, the choice of the intermediate threshold is linked to both the dataset structure and the gap between the existing NBL and TV. In this way, we are able to define some sub-area in which estimate, effectively, some local NBL (LNBL) values (Table 2). A unique value of NBL does not seem to be realistic in regional environments with a large variation of concentrations.

Table 2. Arsenic concentration classes used to evaluate the LNBLs.

1st	As ≤ 10 µg/L (TV)	Lower ($LNBL_L$)
2nd	10 < As ≤31.4 µg/L (M)	Intermediate ($LNBL_I$)
3rd	As > 31.4 µg/L	Higher ($LNBL_H$)

According to the above-described thresholds, the map in Figure 8 (related to the average values of the four surveys) shows that many zones (about 49% of the study area) exceeded the threshold value (TV) set in the Annex III of the Commission staff working document of the GWD 2006/118/EC. In some of these zones (9% of the domain), the predicted values also exceeded the natural background level (NBL) of 74 µg/L suggested by the "A.Li.Na" project. In particular, in the central-western region of the area between the Padua and Venice provinces (including Piezometers 9, 10, 20, 37, 40 and 44), the predicted arsenic concentrations were greater than the TV, ranging from 33.75 to 224 µg/L. Similarly, arsenic concentrations reached the maximum value of 95 µg/L in

the eastern region of the study area, near the Venice airport (Piezometer 15). Examining the ORP overlaid post-map, the ORP values denoted a reducing environment in these areas, in agreement with the results of several studies on the role of ORP in As release (Carraro et al., 2015, 2013; Dalla Libera et al., In Press; McArthur et al., 2001; Molinari et al., 2015; Nickson et al., 2000; Rotiroti et al., 2014). These reducing conditions occur also in some "green" and "yellow" zones but in this case arsenic is low cause the As-bearing material are missing in the subsoil. Otherwise, analyzing the available stratigraphic logs (Figure 9), the presence of peat layers is evident in the areas with high concentrations, confirming the role of organic matter as a redox driver of As release (Dalla Libera et al., In Press; Molinari et al., 2013; Nickson et al., 2000; Rotiroti and Fumagalli, 2013; Rotiroti et al., 2014; Ungaro et al., 2008). From a hydrogeological perspective, a map analysis suggested that the arsenic contamination disagreed with the groundwater natural drainage and did not spread in accordance with the groundwater flow directions (NW-SE) (Fabbri, 2013).

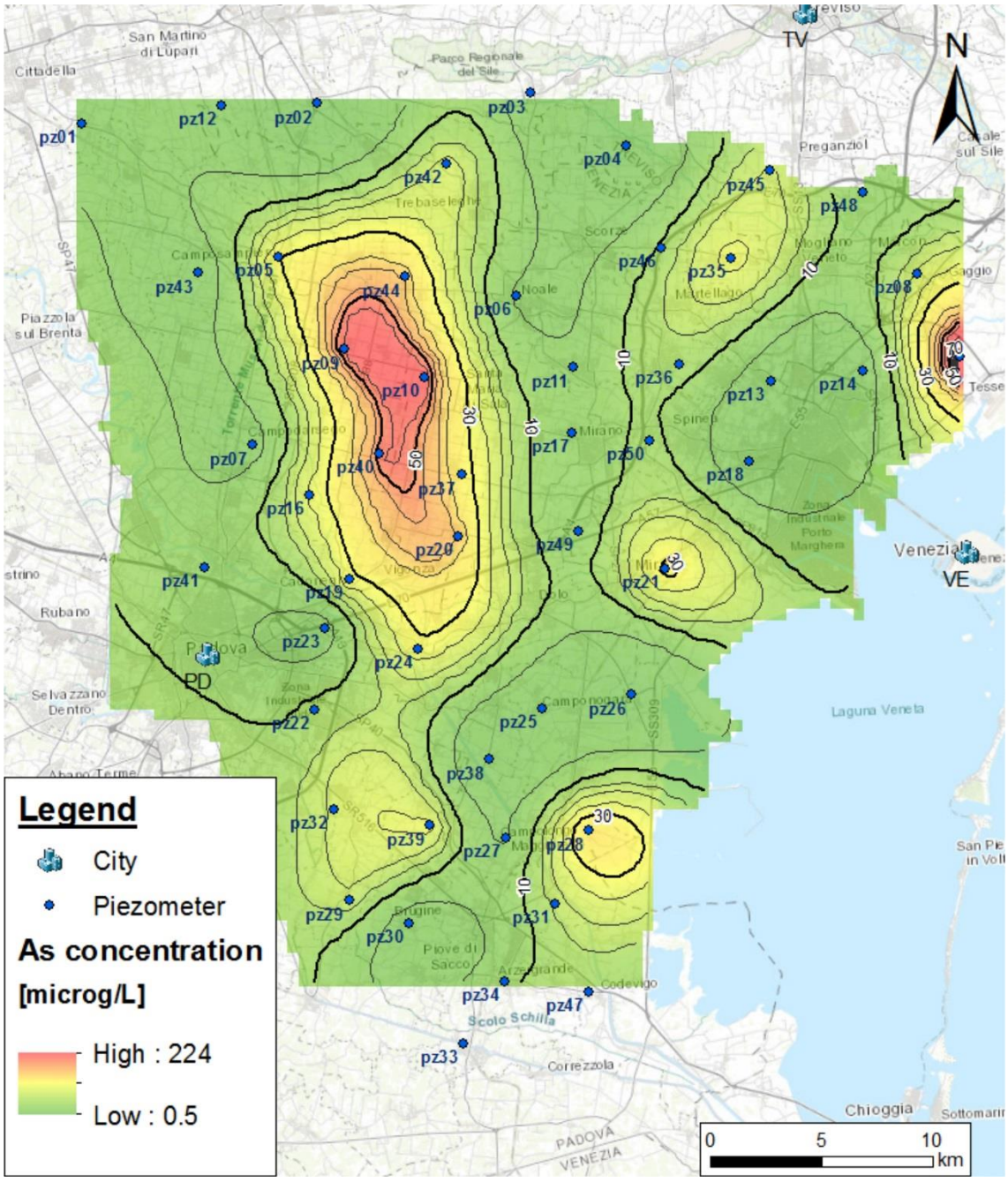

Figure 8. Original prediction of arsenic concentration within the DBVL. This map shows the raw prediction results obtained by the COK method. The contour lines represent the arsenic concentration values basing on the COK prediction with an interval of 5 µg/L. The bold lines are the main contours.

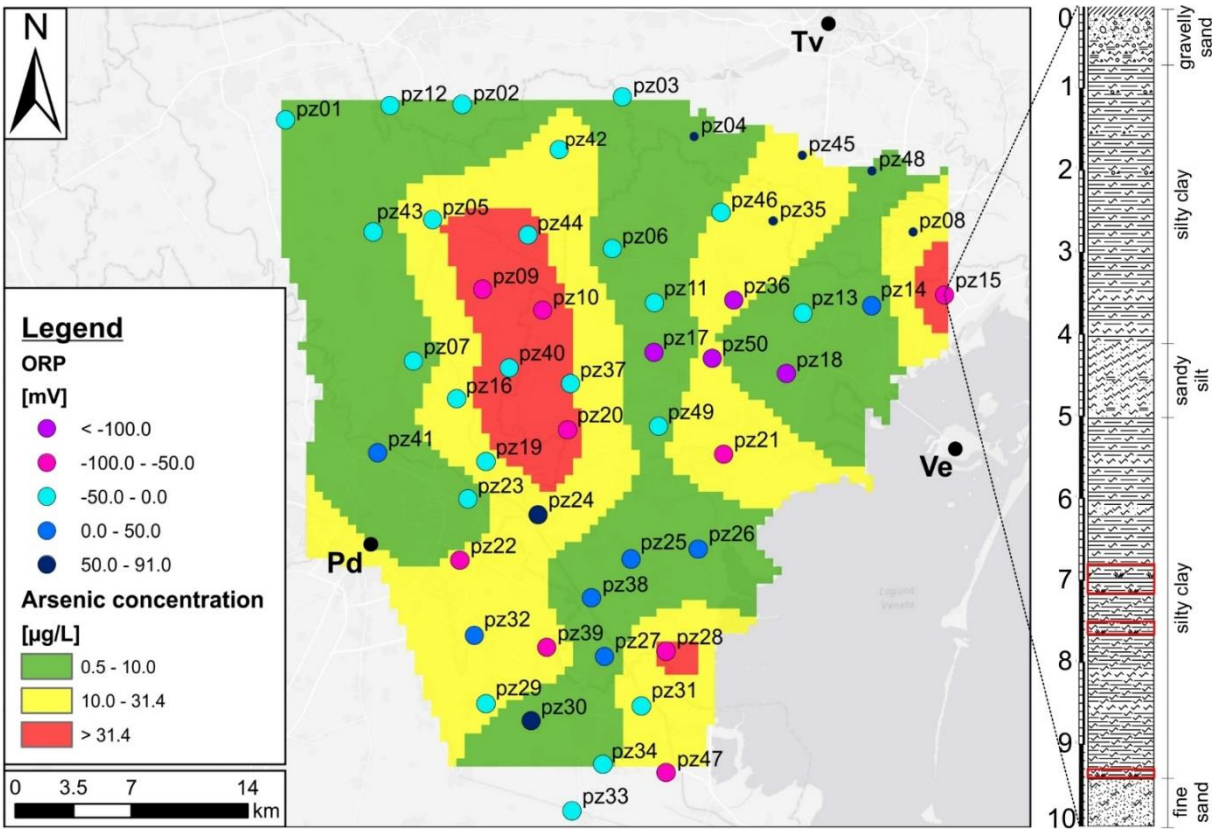

Figure 9. The reclassified arsenic prediction , obtained from the raw prediction in figure 8, and the ORP post-map. The stratigraphic log shows some peat layers corresponding to the zones with high arsenic values. The red boxes highlight the peat layers.

5.3.Implications for water and environmental management

Groundwater arsenic pollution in alluvial systems can result from natural processes, such as weathering, the interaction between groundwater and mineral deposits or the degradation of organic matter deposits (e.g., peat layers). The distinction between anthropogenic and natural sources of contamination is important for approaching the contamination issues. Therefore, the evaluation of an NBL value plays an important role in highlighting anthropogenic pollution. The methods, suggested by the EU BRIDGE project, provide a statistically representative value for the entire study area but does not take into consideration the spatial distribution of the data. Our approach suggests a Local Natural Background Levels (LNBLs), combining the 90th percentile concept and the spatial data distribution and defining two or three classes (depending to the difference between TV and NBL), in which we can estimate the LNBLs. In this way, the errors between the natural

occurrence and the man-made pollution are minimized. By considering a unique NBL value for the entire area, an incorrect distinction between natural occurrence and pollution could occur. For example, in zones with a concentration lower than the NBL, we risk accepting an external pollution as a natural occurrence. However, in an area belonging to a higher class, we risk misinterpreting a natural occurrence as pollution. Therefore, we suggest calculating a local NBL using only the data inside the areas, defined by concentration classes established using a geostatistical prediction (e.g., the green areas in Figure 6 with predicted values between 0.5 and 10 μg/L). This step could be easily conducted if the number of available data is significantly adequate in each class ($n \geq 30$) (Müller et al., 2006). If the number of data is lower than 30, the sampling network should be enhanced. Referring to the DBVL area, all three identified zones have a number of piezometers fewer than 30 in the "A.Li.Na" sampling network. Therefore, a significant calculation of the LNBL was not performed. However, in the green areas (about 51% of the study area, Figure 9) with a predicted arsenic concentration between 0.5 to 10 μg/L (TV), 24 piezometers are available. For example, in these areas, the LNBL was estimated to be equal to 6.68 μg/L, according to the BRIDGE project method. Such a result shows that the LNBL value is not only much lower than the "A.Li.Na" NBL value (74 μg/L) but is also lower than the TV (10 μg/L).

6. CONCLUSIONS

The results presented in this paper show how a geostatistical spatial approach allows for the application of the 90th percentile method (NBL) in a more critical way. The map of the natural distribution of arsenic allows for the identification of critical areas, highlighting where the As values exceed the TV. In particular, this result illustrates the usefulness of calculating a local NBL (LNBL) related to the different defined areas. Using the cokriging approach, the definition of LNBLs takes advantage of the arsenic spatial distribution and the geochemical relationships among arsenic and other related parameters. In fact, the correlations between arsenic and auxiliary variables, such as Fe and NH_4^+, are taken into consideration for cokriging prediction. The results of

this study highlight the critical issues related to the application of a single NBL value at the regional scale. Through the BRIDGE suggestions, the European GWD proposes a NBL value without considering the spatial distribution of the parameter concentration or the geochemical relations among the correlated parameters. The LNBL concept aims to reduce the error associated with the distinction between the natural occurrence of arsenic and anthropogenic pollution. By defining a LNBL for every class of concentration, monitoring the anthropogenic pollution events is easy, particularly for the lower and intermediate classes (see Table 2). In addition, the proposed approach might be applied to a global view of the study area, informing decisions about water and environmental management and improving the characterization plan necessary to evaluate potentially contaminated sites. Finally, the suggested approach is easily applicable to estimate the LNBL of any kind of elements or compound deriving from natural processes, both in groundwater and soil.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was founded by the Venice Province, project “IDRO”, grant to P. Fabbri. We thank the ARPAV agency, in particular the internal water observatory, to share their data concerning the arsenic groundwater contamination into the DBVL.

REFERENCES

APAT, 2003. Analytical methods for chemical water analysis, vol. 1. Handbooks and guidelines, 29, I.G.E.R. srl, Roma.

APAT, 2006. Handbook for environmental investigation upon contaminated sites. Handbooks and guidelines, I.G.E.R. srl, Roma.

ARPA, 2014. ALiNa – Analysis of the Natural Background Level for some compounds within the groundwater of the shallow aquifer in the Drainage Basin to the Venice Lagoon (Brenta river alluvial system) – Report and data presentation.

Baiocchi, A., Lotti, F., Piscopo, V., 2011. Influence of hydrogeological setting on the arsenic occurrence in groundwater of the volcanic areas of central and southern Italy. AQUA mundi 2, 131–142. doi:10.4409/Am-035-11-0035

Baviskar, S., Choudhury, R., Mahanta, C., 2015. Dissolved and solid-phase arsenic fate in an arsenic-enriched aquifer in the river Brahmaputra alluvial plain. Environ. Monit. Assess. 187, 93. doi:10.1007/s10661-015-4277-0

Biswas, A., Gustafsson, J.P., Neidhardt, H., Halder, D., Kundu, A.K., Chatterjee, D., Berner, Z., Bhattacharya, P., 2014. Role of competing ions in the mobilization of arsenic in groundwater of Bengal Basin: Insight from surface complexation modeling. Water Res. 55, 30–39. doi:10.1016/j.watres.2014.02.002

Bivand R.S., Pebesma E.J., Gomez-Rubio V., 2009. Applied spatial data analysis wit R; Springer, ISBN 978-0-387-78170-9, DOI: 10.1007/978-0-387-78171-6

Bondesan, A., Meneghel, M., 2004. Geomorphology of the Venice Province (in Italian). Esedra, Padova.

Cambruzzi T., Conchetto E., Fabbri P., Marcolongo E., Rosignoli A., Zangheri P., 2009. Groundwater resources in the middle Venetian plain. Rend. Online Soc. Geol. It., 6, 127-128.

Carraro, A., Fabbri, P., Giaretta, A., Peruzzo, L., Tateo, F., Tellini, F., 2015. Effects of redox conditions on the control of arsenic mobility in shallow alluvial aquifers on the Venetian Plain (Italy). Sci. Total Environ. 532, 581–594. doi:10.1016/j.scitotenv.2015.06.003

Carraro, A., Fabbri, P., Giaretta, A., Peruzzo, L., Tateo, F., Tellini, F., 2013. Arsenic anomalies in shallow Venetian Plain (Northeast Italy) groundwater. Environ. Earth Sci. 70, 3067–3084. doi:10.1007/s12665-013-2367-2

Cleveland W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American statistical association, Vol. 74, No. 368, pp. 829 – 836.

Coetsiers, M., Blaser, P., Martens, K., Walraevens, K., 2009. Natural background levels and threshold values for groundwater in fluvial Pleistocene and Tertiary marine aquifers in Flanders, Belgium. Environ. Geol. 57, 1155–1168. doi:10.1007/s00254-008-1412-z

Dalla Libera N., Fabbri P., Mason L., Piccinini L., Pola M., (In press). Natural Arsenic in groundwater in the drainage basin to the Venice lagoon (Brenta Plain, NE Italy): the organic matter's role. Rend. Online Soc. Geol. It.

Dal Prà, A., Fabbri, P., Bortoletto, C., 1992. The artesian hydrogeological system and its exploitation in the area between the Treviso City and the Piave River (in Italian). Memorie di Scienze Geologiche, 44, 151-170.

D. Lgs. 152/06, 2006. Legislative Decree on environmental regulations.

D. Lgs. 30/09, 2009. Legislative Decree on the implementation of Directive 2006/118/EC on the protection of groundwater from contamination and degradation.

Edmunds, W.M., Shand, P., 2008. Natural Groundwater Quality.Wiley-Blackwell (488 pp., ISBN 978-14051- 5675-2).

Fabbri, P., Piccinini, L., Marcolongo, E., Pola, M., Conchetto, E., Zangheri, P., 2016. Does a change of irrigation technique impact on groundwater resources? A case study in Northeastern Italy. Environ. Sci & Policy. 63, 63.75. doi: 10.1016/j.envsci.2016.05.009

Fabbri, P., Gaetan, C., Zangheri, P., 2011. Transfer function-noise modelling of an aquifer system in NE Italy. Hydrol. Process. 25, 194–206. doi:10.1002/hyp.7832

Fabbri, P., Piccinini, L., 2013. Assessing transmissivity from specific capacity in an alluvial aquifer in the middle Venetian plain (NE Italy). Water Sci. Technol. 67, 2000–2008. doi:10.2166/wst.2013.074

Fabbri, P., Zangheri, P., Bassan, V., Fagarazzi, E., Mazzucato, A., Primon, S., Zogno, C., 2013. Hydrogeology of the Venice Province – The shallow aquifers (in Italian). Venice Province – University of Padova.

Fendorf, S., Michael, H.A., van Geen, A., 2010. Spatial and Temporal Variations of Groundwater Arsenic in South and Southeast Asia. Science (80-.). 328, 1123–1127. doi:10.1126/science.1172974

Fontana, A., Mozzi, P., Bondesan, A., 2004. The geomorphological evolution of the Venetian-Friulian plain (in Italian). Whitin the: Bondesan, A., Meneghel, M., 2004. Geomorphology of the Venice Province. Esedra, Padova. 113-138.

Fontana, A., Mozzi, P., Bondesan, A., 2008. Alluvial megafans in the Venetian–Friulian Plain (northeastern Italy): Evidence of sedimentary and erosive phases during Late Pleistocene and Holocene. Quaternary International, 189, 71-90.

GWD, 2006. Groundwater Directive 2006/118/CE, Directive of the European Parliament and of the Council on the protection of groundwater against pollution and deterioration, OJ L372, 27/12/ 2006, pp 19–31.

Hinsby, K., Condesso de Melo, M.T., Dahl, M., 2008. European case studies supporting the derivation of natural background levels and groundwater threshold values for the protection of dependent ecosystems and human health☆. Sci. Total Environ. 401, 1–20. doi:10.1016/j.scitotenv.2008.03.018

Isaaks E.H., Srivastava R.M. An introduction to Applied Geostattistics; Oxford University Press: New York, 1989.

McArthur, J.M., Banerjee, D.M., Hudson-Edwards, K.A., Mishra, R., Purohit, R., Ravenscroft, P., Cronin, A., Howarth, R.J., Chatterjee, A., Talukder, T., Lowry, D., Houghton, S., Chadha, D.K., 2004. Natural organic matter in sedimentary basins and its relation to arsenic in anoxic ground water: The example of West Bengal and its worldwide implications. Appl. Geochemistry 19, 1255–1293. doi:10.1016/j.apgeochem.2004.02.001

McArthur, J.M., Ghosal, U., Sikdar, P.K., Ball, J.D., 2016. Arsenic in Groundwater: The Deep Late Pleistocene Aquifers of the Western Bengal Basin. Environ. Sci. Technol. 50, 3469–3476. doi:10.1021/acs.est.5b02477

McArthur, J.M., Ravenscroft, P., Safiulla, S., Thirlwall, M.F., 2001. Arsenic in groundwater: Testing pollution mechanisms for sedimentary aquifers in Bangladesh. Water Resour. Res. 37, 109–117. doi:10.1029/2000WR900270

Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2015. Arsenic fractioning in natural solid matrices sampled in a deep groundwater body. Geoderma 247-248, 88–96. doi:10.1016/j.geoderma.2015.02.011

Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2012. Natural background levels and threshold values of chemical species in three large-scale groundwater bodies in Northern Italy. Sci. Total Environ. 425, 9–19. doi: 10.1016/j.scitotenv.2012.03.015

Molinari, A., Guadagnini, L., Marcaccio, M., Straface, S., Sanchez-Vila, X., Guadagnini, A., 2013. Arsenic release from deep natural solid matrices under experimentally controlled redox conditions. Sci. Total Environ. 444, 231–240. doi: 10.1016/j.scitotenv.2012.11.093

Mozzi P., Bini C., Zilocchi L., Beccatini R., Mariotti Lippi M., 2003. Stratigraphy, palaeopedology and palynology of late Pleistocene and Holocene deposits in the landward sector of the lagoon of Venice (Italy), in relation to Caranto level. Il Quaternario, 16 (1 bis), pag. 193 – 210.

Müller D., Blum A., Hart A., Hookey J., Kunkel R., Scheidleder A., Tomlin C., Wendland F., 2006. BRIDGE Project: Background criteria for identification of groundwater thresholds, D18: Final proposal for a methodology to set up groundwater threshold values in Europe.

Nickson, R., McArthur, J., Burgess, W., Ahmed, K.M., Ravenscroft, P., Rahman, M., 1998. Arsenic poisoning of Bangladesh groundwater. Nature 395, 338. doi:10.1038/26387

Nickson, R.T., Mcarthur, J.M., Ravenscroft, P., Burgess, W.G., Ahmed, K.M., 2000. Mechanism of arsenic release to groundwater, Bangladesh and West Bengal. Appl. Geochemistry 15, 403–413. doi:10.1016/S0883-2927(99)00086-4

Nordstrom, D.K., 2003. Arsenic in Ground Water, Science Mag. Springer US, Boston, MA. doi:10.1007/b101867

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691. doi: 10.1016/j.cageo.2004.03.012

Piccinini, L., Fabbri, P., Pola, M., 2016. Point dilution tests to calculate groundwater velocity, an example in a porous aquifer in NE Italy. Hydrol. Sci. J. 61(8), 1512-1523. doi:10.1080/02626667.2015.1036756

Piccinini, L., Fabbri, P., Pola, M., Marcolongo, E., Rosignoli, A., 2015. Numerical modeling to well-head protection area delineation, an example in Veneto Region (NE Italy). Rend. Online Soc. Geol. It. 35, 232-235. doi: 10.3301/ROL.2015.108

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, 885–897. doi:10.1007/s12665-009-0404-y

Posa D., De Iaco S., 2009. Geostatistics, theory and applications. G. Giappichelli Editore, Turin, Italy, ISBN 978-88-348-9744-7

R Core Team, 2015. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Rango, T., Vengosh, A., Dwyer, G., Bianchini, G., 2013. Mobilization of arsenic and other naturally occurring contaminants in groundwater of the main ethiopian rift aquifers. Water Res. 47, 5801–5818. doi:10.1016/j.watres.2013.07.002

Rotiroti, M., Fumagalli, L., 2013. Derivation of preliminary natural background levels for naturally Mn, Fe, As and NH4+ rich groundwater: The case study of Cremona area (Northern Italy). Rend. Online Soc. Geol. Ital. 24, 284–286.

Rotiroti, M., Sacchi, E., Fumagalli, L., Bonomi, T., 2014. Origin of Arsenic in Groundwater from the Multilayer Aquifer in Cremona (Northern Italy). Environ. Sci. Technol. 48, 5395–5403. doi:10.1021/es405805v

Rowland, H.A.L., Pederick, R.L., Polya, D.A., Pancost, R.D., Van Dongen, B.E., Gault, A.G., Vaughan, D.J., Bryant, C., Anderson, B., Lloyd, J.R., 2007. The control of organic matter on microbially mediated iron reduction and arsenic release in shallow alluvial aquifers, Cambodia. Geobiology 5, 281–292. doi:10.1111/j.1472-4669.2007.00100.x

Smedley, P.L., Kinniburgh, D.G., 2002. A review of the source, behaviour and distribution of arsenic in natural waters. Appl. Geochemistry 17, 517–568. doi:10.1016/S0883-2927(02)00018-5

Sorg, T.J., Chen, A.S.C., Wang, L., 2014. Arsenic species in drinking water wells in the USA with high arsenic concentrations. Water Res. 48, 156–169. doi:10.1016/j.watres.2013.09.016

Takeno, N., 2005. Atlas of Eh-pH diagrams Intercomparison of thermodynamic databases. Natl. Inst. Adv. Ind. Sci. Technol. Tokyo 285.

Trevisani, S., Fabbri, P., 2010. Geostatistical modeling of a heterogeneous site bordering the Venice lagoon, Italy. Ground Water 48, 614–623. doi:10.1111/j.1745-6584.2009.00632.x

Ungaro, F., Ragazzi, F., Cappellin, R., Giandon, P., 2008. Arsenic concentration in the soils of the Brenta Plain (Northern Italy): Mapping the probability of exceeding contamination thresholds. J. Geochemical Explor. 96, 117–131. doi:10.1016/j.gexplo.2007.03.006

Vorlicek, P.A., Antonelli, R., Fabbri, P., Rausch, R., 2004. Quantitative hydrogeological studies of the Treviso alluvial plain, NE Italy. Q. J. Eng. Geol. Hydrogeol. 37, 23–29. doi:10.1144/0036-9276/02-006

Welch, A.H., Westjohn, D.B., Helsel, D.R., Wanty, R.B., 2000. Arsenic in Ground Water of the United States: Occurrence and Geochemistry. Ground Water 38, 589–604. doi:10.1111/j.1745-6584.2000.tb00251.x

Wendland, F., Berthold, G., Blum, A., Elsass, P., Fritsche, J.-G., Kunkel, R., Wolter, R., 2008a. Derivation of natural background levels and threshold values for groundwater bodies in the Upper Rhine Valley (France, Switzerland and Germany). Desalination 226, 160–168. doi:10.1016/j.desal.2007.01.240

Wendland, F., Blum, A., Coetsiers, M., Gorova, R., Griffioen, J., Grima, J., Hinsby, K., Kunkel, R., Marandi, A., Melo, T., Panagopoulos, A., Pauwels, H., Ruisi, M., Traversa, P., Vermooten, J.S.A., Walraevens, K., 2008b. European aquifer typology: A practical framework for an overview of major groundwater composition at European scale. Environ. Geol. 55, 77–85. doi:10.1007/s00254-007-0966-5

Zavatti, A.; Attramini, D.; Bonazzi, A.; Boraldi, V.; Malagò, R.; Martinelli, G.; Naldi, S.; Patrizi, G.; Pezzera, G.; Vandini, W.; Venturini, L.; Zuppi, G. M., 1995. Occurrence of groundwater arsenic in the Po Plain: Environmental evidences and geochemical hypothesis. Quad. Geol. Appl., S2, 2.301−2.326.