



The distribution of arsenic in shallow alluvial groundwater under agricultural land in central Portugal: Insights from multivariate geostatistical modeling

A.I.A.S.S. Andrade ^{a,*}, T.Y. Stigter ^b

^a Centro de Geofísica, University of Coimbra, Av. Dias da Silva, 3000-134 Coimbra, Portugal

^b Geo-Systems Centre/CVRM, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001, Lisbon, Portugal

HIGHLIGHTS

- ▶ We studied arsenic levels in shallow alluvial groundwater under agricultural land.
- ▶ Multivariate geostatistics is used to model spatio-temporal arsenic occurrence.
- ▶ Results indicate the dominant role of reductive dissolution of Fe (hydr)oxides.
- ▶ Redox state is linked to rice crops, fertilizers can provide additional As source.
- ▶ Best regression model for As uses season, crop, well and water depth, Eh and NO₃⁻.

ARTICLE INFO

Article history:

Received 23 October 2012

Received in revised form 8 January 2013

Accepted 8 January 2013

Available online 12 February 2013

Keywords:

Arsenic
Alluvial groundwater
Redox conditions
Agriculture
Factorial regression
Indicator kriging

ABSTRACT

In this study multivariate and geostatistical methods are jointly applied to model the spatial and temporal distribution of arsenic (As) concentrations in shallow groundwater as a function of physicochemical, hydrogeological and land use parameters, as well as to assess the related uncertainty. The study site is located in the Mondego River alluvial body in Central Portugal, where maize, rice and some vegetable crops dominate. In a first analysis scatter plots are used, followed by the application of principal component analysis to two different data matrices, of 112 and 200 samples, with the aim of detecting associations between As levels and other quantitative parameters. In the following phase explanatory models of As are created through factorial regression based on correspondence analysis, integrating both quantitative and qualitative parameters. Finally, these are combined with indicator-geostatistical techniques to create maps indicating the predicted probability of As concentrations in groundwater exceeding the current global drinking water guideline of 10 µg/l. These maps further allow assessing the uncertainty and representativeness of the monitoring network. A clear effect of the redox state on the presence of As is observed, and together with significant correlations with dissolved oxygen, nitrate, sulfate, iron, manganese and alkalinity, points towards the reductive dissolution of Fe (hydr)oxides as the essential mechanism of As release. The association of high As values with rice crop, known to promote reduced environments due to ponding, further corroborates this hypothesis. An additional source of As from fertilizers cannot be excluded, as the correlation with As is higher where rice is associated with vegetables, normally associated with higher fertilization rates. The best explanatory model of As occurrence integrates the parameters season, crop type, well and water depth, nitrate and Eh, though a model without the last two parameters also gives quite satisfactory results.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Although As is considered to be essential to some organisms, and even to humans according to some authors (e.g. Reimann and Caritat, 1998; Mendes and Oliveira, 2004), it is also toxic, teratogenic (e.g. Reimann and Caritat, 1998) and carcinogenic (e.g. Manahan, 1994; IARC, 2012). In fact, As is considered a very important environmental toxicant (Moyano et al., 2009), and the most significant water pollutant within the metalloids (Manahan, 1994). There are several

studies concerning health problems associated with As toxicity (e.g. Anawar et al., 2002; Duker et al., 2005; Hopenhayn, 2006). The potential impact on surface and groundwater and on soil–plant ecosystems, together with its toxicity and carcinogenicity, even in low concentrations, has led the WHO (2003) to consider it as presenting a particular concern.

From all the sources of As in the environment the greatest threat to human health is probably drinking water (Smedley and Kinniburgh, 2002), considered the main route for As to enter the human body (Moyano et al., 2009). In addition, the high potential of As-contaminated water is considered a worldwide priority problem (García-Sánchez et al., 2005). The drinking water guideline for As is

* Corresponding author. Tel.: +351 239 802376.

E-mail address: aiandrade09@gmail.com (A.I.A.S.S. Andrade).

10 µg/l according to the WHO (2011), the European Commission (EC, 1998) and the Environmental Protection Agency of the USA (EPA, 2009). Its natural concentration in waters is usually less than 1–2 µg/l (WHO, 2011), but the presence of natural or anthropogenic contamination resulting in As concentrations beyond this low guideline value is not uncommon.

As is affected by redox reactions and can be found in the environment between the oxidation state -3 to $+5$, although in natural waters inorganic As is normally present as oxyanions of As(III) and As(V); organic forms are rarely present at important concentrations except in cases of waters strongly affected by industrial pollution (Smedley and Kinniburgh, 2002). The mobility of As in groundwater depends on the species present, the solubility of As minerals in the existent pH–Eh range and water composition in the aquifer, as well as the presence of adsorbing surfaces (Deutsch, 1997); in the environment, the mobility increases with a decrease of Eh and pH (Siegel, 2002) which, according to Smedley and Kinniburgh (2002) are the most important factors controlling As speciation. Arsenic toxicity depends on the redox speciation, with As(III) presenting a higher acute toxicity than As(V) (Deutsch, 1997).

High concentrations of As in the groundwater (> 10 µg/l) have been reported by many authors all over the world, including Bangladesh (e.g. Zahid et al., 2008; Chakraborti et al., 2010; Selim Reza et al., 2010), India (e.g. Kar et al., 2010), Pakistan (e.g. Farooqi et al., 2007), Taiwan (e.g. Liu et al., 2003) and the United States (e.g. Moore and Woessner, 2003). High values of As are found in different types of aquifers and under variable depths, under reducing or oxidizing environments (in the latter case associated with high pH) and under arid, humid or temperate climates (Smedley and Kinniburgh, 2002). The most serious and extensive contamination with As is considered to occur in alluvial aquifers (Kim et al., 2009; Smedley and Kinniburgh, 2002), as well as in inland or closed basins in arid or semi-arid areas (Smedley and Kinniburgh, 2002). The study of shallow, unconfined aquifers is especially important as, according to Hudak (2008), these aquifers present a special vulnerability to contamination from As surface sources.

In Portugal, most of the As values reported for groundwater are lower than the drinking water guideline of 10 µg/l, with higher values associated to (abandoned or active) mining areas, mainly of metallic sulfides (Cama et al., 2008). According to the National Water Institute (INAG) database (<http://snirh.pt>), concentrations over 100 µg/l occur in the North–Northeast of the country. In the report of INAG (2009) it is also mentioned that a large number of wells present As values over 10 µg/l in the Central and Tagus River Basin Districts. As far as we know, it was detected for the first time in the study area of this paper located in Central Portugal. Existing studies on As contamination in groundwater in Portugal are mostly limited to mining areas (e.g. Antunes et al., 2002; Ávila et al., 2006; Pereira et al., 2008), who report values exceeding 1000 µg/l. Morgada et al. (2008) refer that the problem of As contamination is becoming increasingly important in the country. Notwithstanding, Cama et al. (2008) consider that the population in those areas most affected by water contamination by As should not be in danger if the contamination sources are known, the As mobility in the affected areas is controlled and adequate methods for remediation and prevention are applied.

Although the main source of As in contaminated aquifers in the world is considered to be natural, e.g. resulting from volcanic and geothermal activities, clay-rich aquitards or manganese and iron oxides present in aquifers, in many areas the presence of As in groundwater can be related to human activities (Gómez et al., 2006), providing additional sources of As and/or promoting its mobility (BGS and MacDonald, 2000; Klump et al., 2006). Anthropogenic sources of As include the combustion of fossil fuels, mining activity, the use of As-pesticides/herbicides, the use of As additives in livestock feeding or the use of As products for wood preservation (Smedley and Kinniburgh, 2002). Although some of the uses of As products were considerably reduced in the last decades, the impact on the environment

will remain for some time at least in the application areas (Smedley and Kinniburgh, 2002).

Arsenic mobilization by human activities can be determinant for its presence in groundwater. According to Stüben et al. (2003) the mechanism of As solubilization can be more important than source type or size. This may lead to higher water concentrations in areas where only moderate levels of arsenic occur in the host formation (Garellick et al., 2008). Where high As concentrations occur, it is therefore important to analyze all the potential sources and causes of mobilization, and to find correlations between As concentrations and other relevant physicochemical parameters, as well as hydrological setting and land use. Doing so may enable making relevant considerations regarding the presence of As at other locations with similar settings. The present paper therefore aims to jointly apply hydrochemical, multivariate and geostatistical techniques to model the occurrence of As concentrations in shallow alluvial groundwater in Central Portugal, as well as to address the associated uncertainty. Following a basic application of scatter plots and principal component analysis to study the correlation of As with physicochemical variables, factorial regression based on correspondence analysis is performed to construct advanced explanatory models of As occurrence, integrating both quantitative and qualitative (hydrogeological and land use) parameters. Uncertainty is assessed by applying indicator-geostatistical techniques that result in probability of concentrations in groundwater exceeding the drinking water guideline.

2. Study area

2.1. Location and climate

The study area, located west of Coimbra city, is part of the Mondego River drainage basin in Central Portugal. It covers a flat, approximately E–W elongated area of almost 51 km² with a maximum elevation of nearly 13 m above mean sea level on the eastern side and a minimum elevation of 4 m on the western side, close to the upper limit of the river estuary (Fig. 1). The natural drainage network of the study area has been altered by the construction of an artificial channel for the Mondego river that provides water to an irrigation system, allowing land irrigation with river water in the main part of the study area; the water for this system is diverted from the river approximately 3 km east of the study area.

The climate in the region is temperate, with hot and dry summers and mild and rainy winters (Rebelo et al., 1990); the mean annual rainfall (1970–1998) is about 900 mm, showing little spatial variation within the study area, and the mean monthly temperature ranges from 10 °C in January to nearly 23 °C in July (Andrade, 2005).

2.2. Geology and hydrogeology

In the study site and surrounding area the Holocene infill corresponds to fluvial and estuarine deposits covering a bedrock of Mesozoic–Tertiary sandstones, mudstones and limestones, overlying a Paleozoic–Precambrian basement (Soares et al., 1985; Barbosa et al., 1988; Almeida et al., 1990). A geological map is provided in Andrade and Stigter (2011). The alluvial plain extends from Coimbra (east) to Figueira da Foz (west) and represents an area of 150 km², with a maximum width of 4 km (Almeida et al., 1990). The thickness of the alluvium generally increases downstream, reaching 40 m at Figueira da Foz but higher values were identified at several places (the maximum value found, according to Almeida et al. (2000), is nearly 50 m and is just west of Coimbra).

A progressive reduction of the volumetric content of coarser materials is verified from upstream to downstream (Cristo, 1998). In the area of Coimbra and to the east (see Fig. 1), the sediments are essentially composed of gravel and sand with occasional very high cobble content (Almeida et al., 2000); mud levels (grain size below 63 µm) are of reduced thickness and not so frequent (Cristo, 1998). West of Coimbra an evident increase in thickness and frequency of muddy

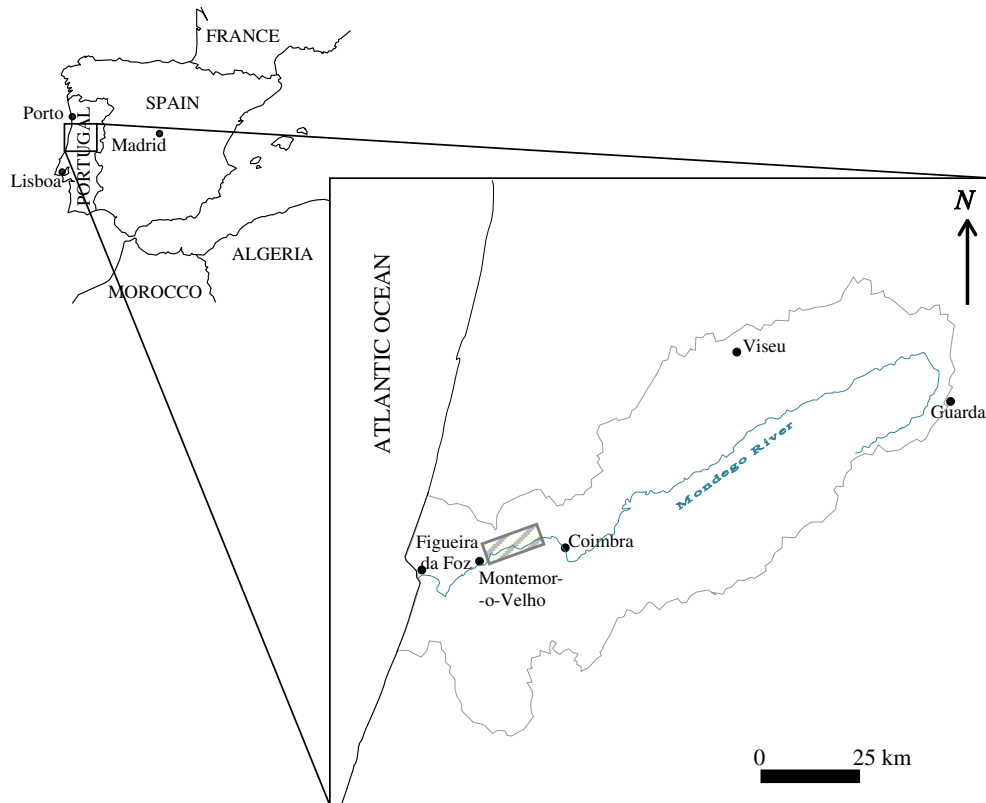


Fig. 1. Location of the study area in the Mondego drainage basin – Central Portugal.

levels occurs, and according to [Cristo \(1998\)](#) near Montemor-o-Velho, and downstream, the muddy levels are thicker and almost always conchiferous.

In the study area, the upper levels of the Holocene formation are predominantly constituted by sandy sediments, in the eastern and central zones, with an increase of muddy sediments towards the west ([Andrade, 2005](#)). This can be seen in [Fig. 2](#) providing a map with the log description of each borehole where an observation well was installed in the shallow Holocene layers of the area.

According to [Almeida et al. \(2000\)](#), in these Holocene sediments the existence of a porous multilayered aquifer system with an unconfined to semi-confined/confined hydraulic behavior is considered, as a result of the variability of textural and structural conditions of the alluvial beds; the aquifer formations correspond to sands and sands with pebbles and cobbles, with clay and mud intercalations. The hydraulic conductivities (K) in the aquifer system globally range from 40 m/day in the west, downstream of Montemor-o-Velho, to 220 m/day in the east, near and upstream of Coimbra ([Cristo, 1998](#), see [Fig. 1](#)). In the study area, K values between 4.2 and 181 m/day were obtained for the coarser grained sediments of the upper layers from grain-size distribution curves of 10 sediment samples using the Hazen method ([Hazen, 1911](#)). The poor sorting of the samples explains the somewhat lower K . The interpretation by the Hvorslev method ([Fetter, 2001](#)) of two slug tests developed in wells installed in muddy levels gave values of 1.3 and 1.6 cm/day; these K values can reveal a problem of clogging of the well screen by the finer fraction ([Andrade, 2005](#)).

Groundwater flow is locally in the direction towards the streams that cross the area longitudinally and regionally is along the alluvial body towards the ocean ([Almeida et al., 2000](#)), roughly E–W. In the areas where the bedrock is composed of permeable formations some discharge to lateral and underlying aquifers is also admitted ([Cristo, 1998](#)). Besides direct infiltration from precipitation (main recharge), other recharge components can be considered such as: a) river bank infiltration ([Cristo, 1998; Almeida et al., 2000](#)), b) recharge from lateral

and underlying aquifers ([Almeida et al., 2000](#)) and c) infiltration associated to an excess of irrigation ([Andrade, 2005](#)). The average recharge of the aquifer by precipitation is estimated to be about 300 mm/year ([Almeida et al., 2000](#)). Although there are no significant spatial variations in precipitation and evapotranspiration ([Andrade, 2005](#)) recharge is considered to be somewhat higher in the east, where the alluvial sediments are coarser at the surface.

2.3. Soil type and land use

The soils of the study area are mainly fluvisols and have predominantly silt loam texture ([DGADR, 2008](#)). The major land use in the region is agriculture supported mainly by river water irrigation. Maize (*Zea mays* L.) and rice (*Oryza sativa* L.) are the main crops; maize predominates in the east and rice in the west, where alluvial sediments are finer-grained. Other minor crops, namely vegetables and garden nurseries, are also cultivated in the area. The main crops are seasonal and the cultivation period occurs mainly between April and October. A land use map is provided in [Andrade and Stigter \(2009\)](#).

3. Methods

3.1. Sampling and analytical procedures

Groundwater samples for arsenic, nitrate, nitrite and phosphate analyses were collected from 29 observation wells ([Fig. 3](#)), ranging in depth from 2.3 to 5.1 m every two months from September 2001 to September 2002, using a bailer. The wells have a diameter of 6.3 cm and were closed at the top to prevent the direct introduction of any contaminant from the surface. The well screen was placed between 0.2 and 1 m above the bottom. Due to the finer alluvium sediments in some areas ([Fig. 2](#)), it was not possible to purge the wells on the day of water sampling as the water level would not recover in time for sampling. In addition, the collection of water samples immediately after purging would have resulted in

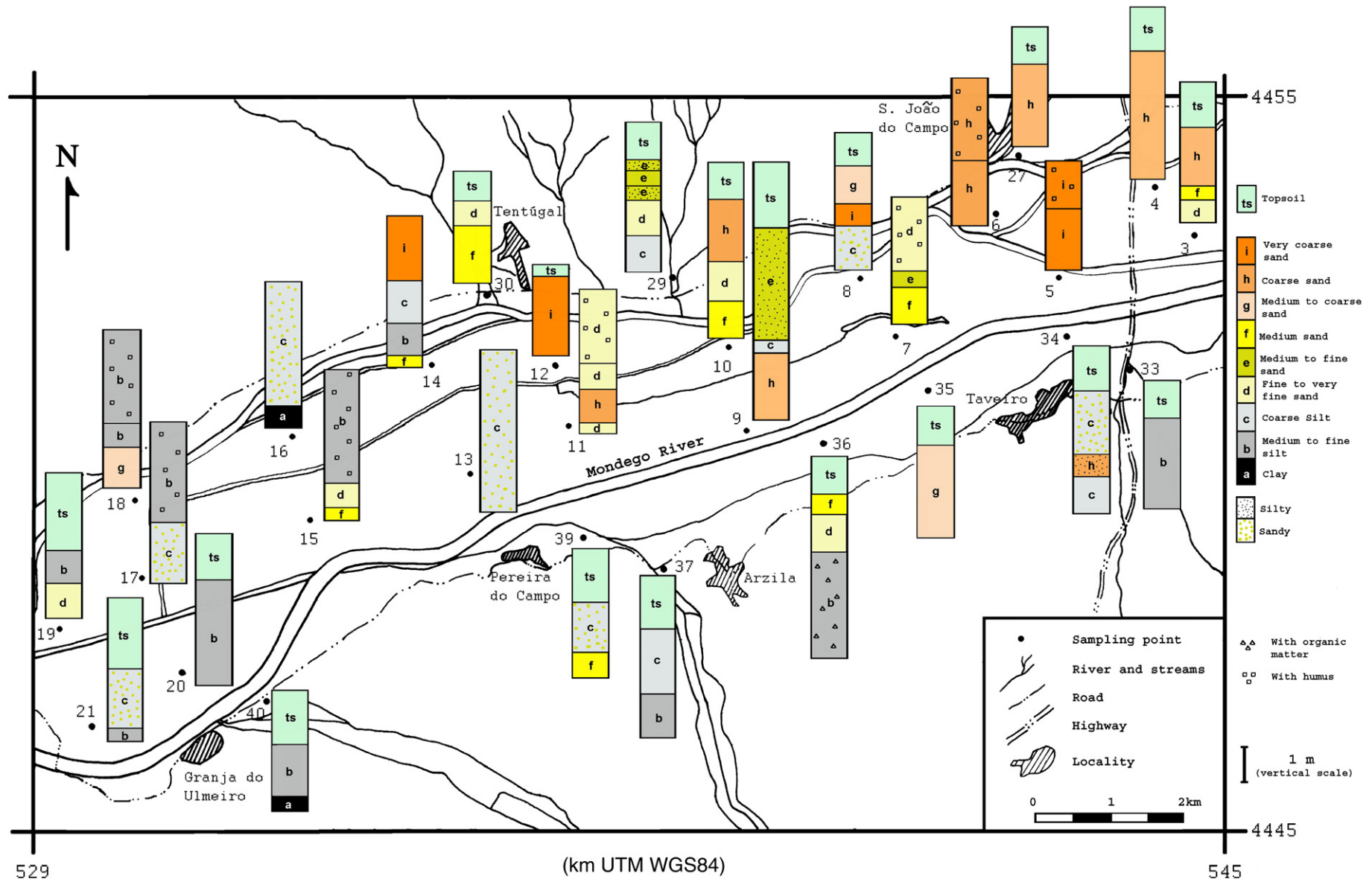


Fig. 2. Map showing the log descriptions of the observation well boreholes, all installed in the shallow layers of the Holocene alluvial deposits in the study area, revealing the high horizontal and vertical heterogeneity.

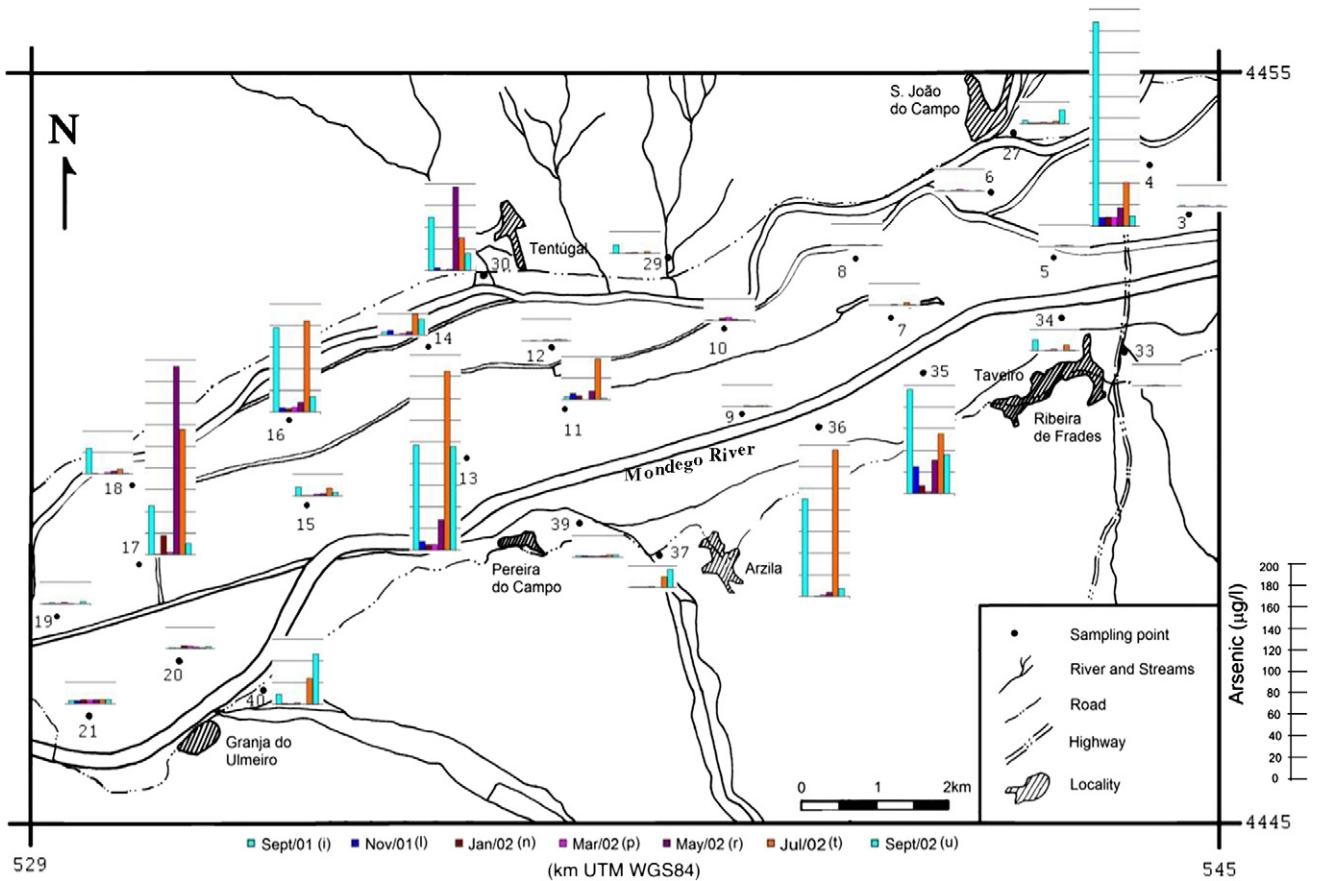


Fig. 3. Distribution of arsenic concentrations in the observation wells for each of the seven sampling campaigns (in chronological order from left to right).

abnormal results for some field parameters and additional sampling difficulties due to water turbidity. The wells were therefore purged two weeks before sampling and immediately after, ensuring that the water could not stay in the well longer than this period of time. Prior to sample collection the first volume sampled by the bailer was rejected, except where there was not enough water in the well. This was to minimize the interference from atmospheric dissolved oxygen (DO) diffusing into reduced groundwater prior to groundwater sampling, a problem that could arise from sampling two weeks after purging the well. As will be shown in the results, and also referred in Andrade and Stigter (2011), some samples indeed appear to represent more reduced conditions than shown by their DO and Eh values measured in the field. Appelo and Postma (2005) further underline the difficulties that occur during field measurements of DO and, particularly common, Eh. Notwithstanding, it will be shown that the values of Eh and DO, as well as the behavior of other chemical ions, are valid indicators of the redox processes taking place.

In January 2002 it was impossible to collect the groundwater sample from the well 33, due to a low water table. In four of the seven sampling campaigns (September 2001, January, March and September 2002) samples were also collected for the analyses of alkalinity, sulfate, chloride, calcium, magnesium, sodium, potassium, manganese and total iron (as ferrous iron).

At each sampling point two different samples were collected: (1) for the determination of arsenic and other cation concentrations and (2) for the determination of nitrate, nitrite and orthophosphate; a third sample was collected in the four sampling campaigns where alkalinity, sulfate and chloride were analyzed. All the samples were collected in tight capped polyethylene bottles and were transported to the laboratory under low temperature conditions in iceboxes. The samples for cation

analysis were acidified in the laboratory with ultra-pure nitric acid till $\text{pH} < 2.0$ after filtration with a Whatman filter number 40 (procedure completed on the day of sample collection). They were then stored in the laboratory at low temperature until analyzed. The samples for nutrient analysis were frozen until the day they were processed.

Measurements of water level, temperature (T), pH, Eh, electrical conductivity (EC), and dissolved oxygen (DO) were performed in situ at all wells during the seven sampling campaigns using portable instruments (dissolved oxygen and temperature meter—Yellow Spring Incorporated, model 58; Crison pH/mV, model 506; conductivity meter—Yellow Spring Incorporated, model 33). The electrical conductivity (EC) was compensated for a standard T of 25 °C.

Alkalinity (as HCO_3^-), was obtained by titration with sulfuric acid till a pH of 4.5; the analyses were performed on the day of sample collection. Chloride was obtained by titration with silver nitrate (Mohr's method) and the concentrations of sulfate, nitrate, nitrite and orthophosphate were measured by UV/VIS spectrophotometry. Cation concentrations were obtained by atomic absorption spectrophotometry; arsenic was obtained by hydride generation atomic absorption spectrometry. All the analyses were performed in the Earth Sciences Department of the University of Coimbra.

The precision of the measurements was checked by taking replicates. The accuracy of the analyses on the four campaigns where major cations and anions were analyzed was determined through the calculation of the ion-balance error, based on Electro Neutrality (Appelo and Postma, 2005). The samples generally presented an error within $\pm 5\%$; few samples presented an error value above 5% and were mostly excluded. An exception was made for samples with total ion concentration below 5 meq/l, as samples with low mineralization are prone to higher analytical errors.

3.2. Principal component analysis

Principal component analysis (PCA) is among the most commonly used multivariate statistical methods in natural sciences. Developed by Hotelling (1933) in the thirties, from original work of Pearson (1901), the main objective of this method is to summarize a large amount of data in terms of several so-called components or factors without losing significant information (Brown, 1998; Pereira and Sousa, 2000). The aim is to discover the underlying pattern of relationships within the data and subsequently rearrange them into a small number of uncorrelated components (or factors, or axes). These are extracted from the data by statistical transformations involving the diagonalization of the variables' correlation or variance–covariance matrix (Brown, 1998; Pereira and Sousa, 2000). Each component describes a certain amount of the statistical variance of the analyzed data and is interpreted according to the intercorrelated variables. Variable loadings are defined by the orthogonal projection of the variables on each of the factors. The correct interpretation of each factor requires the researcher's expertise and experience with the case study. The significant factors are selected, based either on their eigenvalues (generally > 1), cumulative percentage of data variance explained, eigenvalue stabilization or a combination of these criteria.

In this study, PCA was performed on two data matrices, one including six active variables and 200 samples collected during the seven campaigns, and the other built up of ten active variables and 112 samples that were gathered in the four campaigns for a more complete analysis of major and minor ions. The idea was to perform one analysis with fewer variables but more samples, hence providing a more complete picture for the considered variables, whereas the other analysis permitted the study of correlation with other variables included, despite a lower number of samples. A number of variables were projected as supplementary on the factorial planes, meaning they did not actively participate in the component extraction. The reason for this option is that PCA was performed with the specific goal to study the correlations with As and in the many additional runs performed, including the mentioned supplementary variables, correlations between As and other variables were weaker. By including these variables as supplementary, their correlation with the components extracted on the active variables, could still be studied.

3.3. Factorial regression analysis

Factorial regression analysis (FRA) is based on factorial correspondence analysis (FCA), developed by Benzécri in the early sixties (Benzécri, 1973). The method has certain aspects in common with PCA, such as the diagonalization of a similarity matrix for the extraction of the factors. As symmetry is conferred to the data matrix, correlations within and between variables and samples can be studied simultaneously (Benzécri, 1973; Pereira and Sousa, 2000). Moreover, both qualitative and quantitative variables may be used, by dividing them into classes (modalities). Hence, quantitative variables become ordinal and may be correlated to qualitative variables that are either ordinal or nominal. After analyzing the correlations among active and supplementary variables (the latter do not interfere with the statistical algorithms), regression of the response variable is performed by factor extraction on the so-called contingency table of the explanatory variables.

Several applications of FCA to water quality studies are known (e.g., Lachance et al., 1979; Johannesson et al., 1996; Farnham et al., 2003; Andrade and Stigter, 2009). Stigter et al. (2008) perform FRA in a case study of groundwater contamination by nitrate. They provide a detailed description of the theory behind FRA, as well as the significance of the obtained results. In brief, the initial classified and codified data matrix is transformed into a contingency table representing the number of matches of each response variable class with each of the classes of the explanatory variable, in the current case As. A binary matrix is constructed for the samples regarding the explanatory variable (1 = “belongs to the

class”, 0 = “does not belong”), called the supplementary sample matrix (Pereira and Sousa, 2000). Subsequently, the table is subjected to FCA and the first extracted factorial axis is retained. The sample scores resulting from projection on this axis are calculated as follows:

$$F_i = \frac{1}{p\sqrt{\lambda}} \sum_{j=1}^m \delta_j L_j \quad (1)$$

where F_i is sample i 's projection score on the first factor, p is the number of explanatory variables, λ is the eigenvalue of first factor, δ_j is the Boolean code, equals 1 if sample belongs to modality j and 0 if not, L_j is the factor loading of j and m is the number of modalities. When plotting the observed sample values of the response variable R against their factor scores F , a regression model for this variable can be calculated. For a linear model the equations becomes:

$$R' = a + \frac{b}{p\sqrt{\lambda}} \sum_{j=1}^m \delta_j L_j \quad (2)$$

where R' is the estimated value of response variable R and a and b are the regression coefficients. The statistical significance of the regression model is analyzed by examining the residuals, whose variance ($s_{R'}^2$) is a measure of the variation of the plotted points about the regression line:

$$s_{R'}^2 = \sum_{i=1}^n \frac{(R_i - R'_i)^2}{n-2} \quad (3)$$

where R_i is the observed and R'_i the estimated value of sample i . The square root of this formula ($s_{R'}$) is the standard deviation of the residuals, also known as the standard error of estimate.

In the current study several FRA runs were performed using different variables. A first attempt was made to build explanatory models of As based solely on non-physicochemical, and therefore more easily obtainable parameters. The most significant parameters were selected following FCA where their classes were projected as supplementary on the factorial plane constructed with classes from four quantitative variables, namely As, EC, Eh and NO_3^- . The latter two variables were subsequently added to the models and the change in model quality was assessed, based on the coefficient of determination (R^2) and the standard deviation of the residuals ($s_{R'}$, Eq. (3)). Class definition for the most relevant variables and corresponding sample frequencies are specified in Table 1. Regarding the non-physical variables, the several model variants involved land use, season, water and well depth, as well as aquifer lithology.

As no additional data were available to perform model validation, the latter was performed using the following approach: data of two of the seven campaigns were removed from the data set, namely the second summer and winter campaigns of 2002, and new models were built using only the five remaining campaigns. The resulting models were subsequently used to estimate the well As concentrations for the two campaigns left out of the model calculations and then compared to observed As concentrations.

3.4. Probability maps

Since samples were collected for As analysis at each of the 29 observation wells during seven campaigns covering the four seasons, a good way to spatially characterize the occurrence of As with regard to the drinking water guideline of 10 $\mu\text{g}/\text{l}$ is by creating probability maps. Such maps are built with a geostatistical tool called indicator kriging (IK), and illustrate the probability of As concentrations in groundwater exceeding the guideline value. The tool is of particular interest for groundwater management purposes, as it allows the detection of spatial correlation patterns, accounting for the uncertainty inherent to monitoring, as well as contributing to its optimization

Table 1
Parameter description and classification for FCA and FRA.

Parameter	Class	Description	Freq.
Total As concentration	As1	<1 µg/l	89
	As2	1–5 µg/l	60
	As3	>5 µg/l	^a 51
pH	pH1	<6.5	50
	pH2	6.5–7	98
	pH3	>7	52
Eh	Eh1	<200 mV	55
	Eh2	200–500 mV	117
	Eh3	>500 mV	28
EC	EC21	<300 µS/cm	54
	EC22	300–600 µS/cm	80
	EC23	600–1000 µS/cm	36
	EC24	>1000 µS/cm	30
Nitrate concentration	NO31	<1 mg/l	48
	NO32	1–2.5 mg/l	45
	NO33	2.5–10 mg/l	59
	NO34	>10 mg/l	48
Temperature	T1	<16.3 °C	68
	T2	16.3–19.6 °C	66
	T3	>19.6 °C	66
Dissolved oxygen content	O21	<2.98 mg/l	67
	O22	2.98–4.90 mg/l	67
	O23	>4.90 mg/l	66
Well depth	Dpt1	<4 m	158
	Dpt2	>4 m	42
Water depth	WDpt1	<0.5 m	44
	WDpt2	0.5–2 m	105
	WDpt3	>2 m	51
Season	Aut	Autumn	29
	Win	Winter	57
	Spr	Spring	29
	Sum	Summer	85
Aquifer lithology	Cr	Sand	82
	CrFn	Sand and silt	49
	Fn	Sandy silt and/or silt	69
Crop type	NoCr	No crop	80
	Mz	Maize	36
	Rc	Rice	22
	Ot	Other crop (mainly vegetable)	17
	MzOt	Maize and other crop	25
	MzRc	Rice and maize	10
	RcOt	Rice and other crop	7
	MROt	Rice, maize and other crop	3

^a 35 samples with As > 10 µg/l.

(Clark, 1979; Deutsh and Journal, 1998). A similar approach is used by Liu et al. (2004) for As, and by Stigter et al. (2005) and Andrade and Stigter (2009) for nitrate concentrations.

In the current case, prior to IK the data were separated based on sampling season. This was done because the preceding PCA and FCA methods detected seasonality as a relevant parameter associated to the presence of As in groundwater. Hence, the use of “season probability maps” reduces the uncertainty that unnecessarily would be introduced by using all the data. Following the separation among seasons, all samples were binarily codified, applying 0 when the concentration was below or equal to the cut-off level of 10 µg/l and 1 otherwise. The next step consisted of the structural (variographical) analysis of the spatial distribution of the newly created indicator variable, averaged for each observation point (within the same season), followed by the fitting of theoretical models to the experimental semi-variograms.

Table 2
Descriptive statistics for arsenic and other physicochemical parameters.

Parameter	n	Units	Min.	Max.	Median	Mean	SD	VC
As (total)	200	µg/L	<ld	188.39	1.19	12.39	31.85	2.57
T	200	°C	12.60	23.50	17.20	17.79	2.83	0.16
pH	200	–log[H ⁺]	5.21	7.79	6.80	6.72	0.45	0.07
EC	200	µS/cm	94.14	1696.43	476.27	583.06	375.67	0.64
Eh	200	mV	36.00	562.00	416.00	352.77	157.35	0.45
DO	200	mg/L	0.67	9.60	3.67	4.24	2.09	0.49
NO ₃ ⁻	200	mg/L	<dl	466.00	2.79	14.59	49.20	3.37
NO ₂ ⁻	200	mg/L	<dl	0.50	0.01	0.03	0.06	2.29
PO ₄ ³⁻	200	mg/L	<dl	3.44	0.03	0.19	0.54	2.86
HCO ₃ ⁻	112	mg/L	17.60	1190.17	236.11	347.48	299.73	0.86
Cl ⁻	112	mg/L	8.58	148.63	27.37	39.51	31.79	0.80
SO ₄ ²⁻	112	mg/L	<dl	122.08	18.02	27.81	30.53	1.10
Ca ²⁺	112	mg/L	10.36	264.44	62.27	81.84	61.01	0.75
Mg ²⁺	112	mg/L	1.21	108.55	9.65	21.57	23.72	1.10
Na ⁺	112	mg/L	4.08	170.63	17.00	30.56	36.77	1.20
K ⁺	112	mg/L	0.66	11.50	2.45	3.70	2.89	0.78
Fe (total)	112	mg/L	<dl	28.19	0.11	2.92	6.54	2.24
Mn ²⁺	112	mg/L	<dl	9.00	0.61	1.53	2.13	1.40

SD = standard deviation, n = number of samples, EC = Electrical Conductivity compensated at 25 °C, DO = dissolved oxygen, VC = variation coefficient, and dl = detection limit (0.003 mg/L for total Fe, 0.001 mg/L for Mn²⁺, 0.5 mg/L for SO₄²⁻ and NO₃⁻, 0.01 mg/L for NO₂⁻ and PO₄³⁻, 1 µg/L for As).

Model parameter values (range, sill, nugget effect, anisotropy) were obtained and subsequently introduced into an ordinary kriging algorithm to construct the probability maps. A major advantage of kriging is that every estimate is accompanied by a corresponding standard deviation of the estimation error, also known as the standard error (SE). In the specific case of indicator variables and their binary-coded data, indicator kriging results in values between 0 and 1, representing probabilities between 0 and 100% of exceeding the drinking water guideline. The probability maps were also produced for two of the regression models built for As, resulting in “predicted probability maps”, one of which is based exclusively on non-physicochemical parameters.

4. Results

4.1. Descriptive statistics and scatter plots

The descriptive statistics for As, and all the other physicochemical parameters measured in situ or in the laboratory are presented in Table 2. The distribution of As concentrations in each of the 29 observation wells for the seven campaigns is presented in Fig. 3.

Overall salinity (indicated by EC) and ion concentrations show large variations, and increase generally from east to west. The pH values vary between 5.2 and 7.8, but predominantly reveal a slightly acid, near-neutral environment. Eh and DO values indicate the existence of both aerobic and anaerobic environments, the latter is also revealed by the high iron and manganese concentrations. Nonetheless, median values show that there is a prevalence of aerobic environments. A detailed description and explanation of temporal and spatial variations of major ion concentrations, as well as NO₃⁻ and Fe are provided in Andrade and Stigter (2011). It can be noted that As concentrations present a large range of values, confirmed by the high variation coefficient. The large difference between mean and median concentrations also reveals the presence of anomalies. Fig. 3 shows that As concentrations have a heterogeneous distribution both in space and in time, with water samples of many wells revealing seasonal fluctuations; the higher As values were found mainly in the summer campaigns (July and September).

Scatter plots of As versus several physicochemical parameters are presented in Fig. 4. It can be observed that all samples with high As concentrations (i.e. >10 µg/l) present low NO₃⁻ concentrations (<10 mg/l). The same phenomenon is observed for SO₄, though less clear, as some samples with high As also contain SO₄. High As values

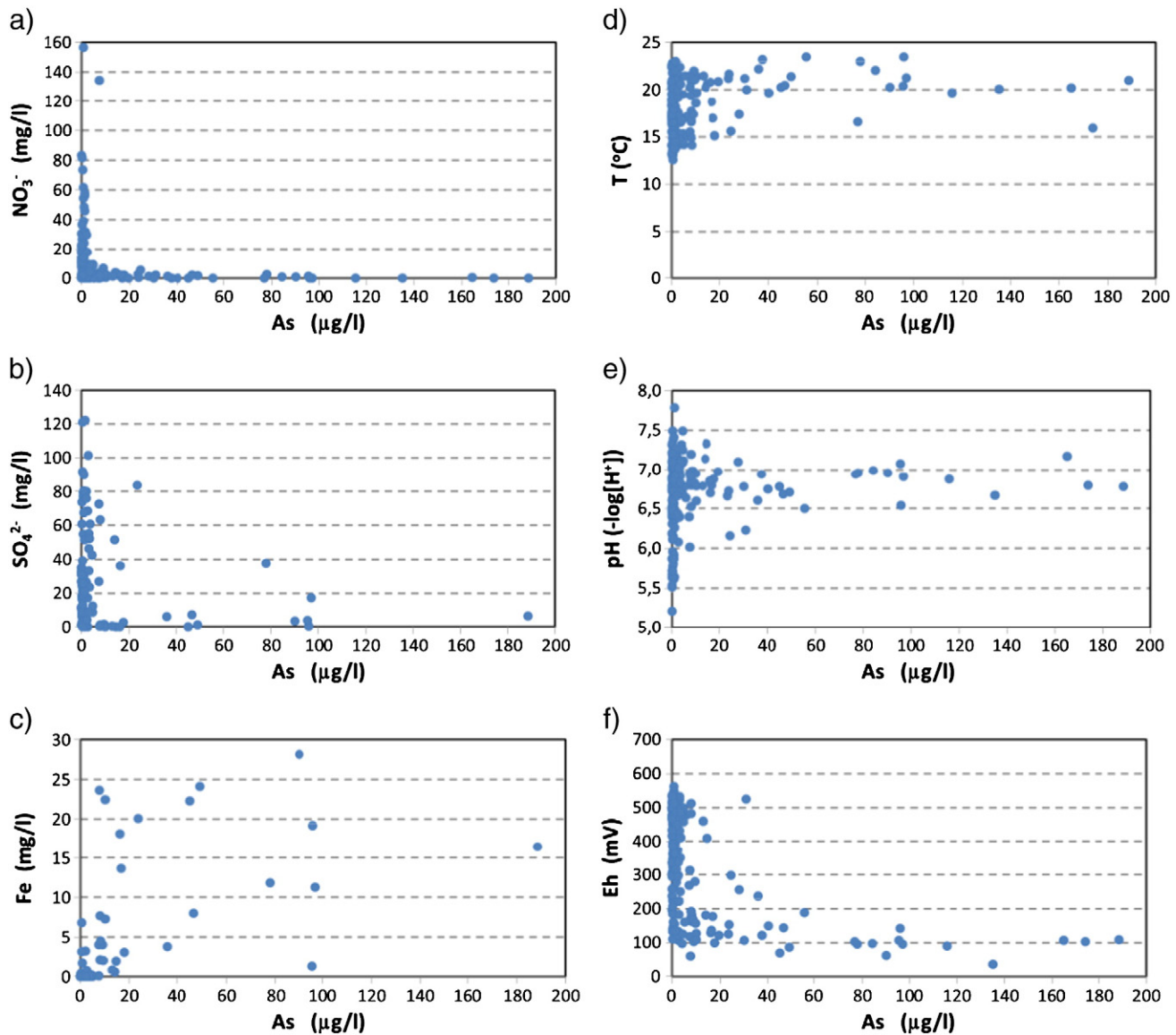


Fig. 4. Scatter plots of arsenic versus NO_3^- (a), SO_4^{2-} (b), $\text{Fe}_{(\text{total})}$ (c), temperature (d), pH (e) and Eh (f); outlier (As 1.57 $\mu\text{g/l}$, NO_3^- 466.0 mg/l) on plot (a) not shown.

are clearly associated with a low redox potential. The correlation with Fe is not that evident, although most samples with $\text{As} > 10 \mu\text{g/l}$ also present relevant iron content (near or above 10 mg/l). Furthermore, most of the samples with $\text{As} > 40 \mu\text{g/l}$ have high temperature ($> 20 \text{ }^\circ\text{C}$) and near neutral pH (6.7–7.2).

From the projection of all the groundwater samples on the Eh–pH diagram (not shown) it was observed that only one sample presents As in the form of As^{3+} , all others presenting the form of As^{5+} . Andrade and Stigter (2011) refer the difficulties that occurred during field measurements and that could have resulted in Eh values higher than the actual Eh values; this phenomenon may have caused an upward shift of the samples in the diagram, concealing the presence of arsenic as As^{3+} in more samples.

4.2. Principal component analysis (PCA)

The variables that participated actively in PCA, on the two data matrices, as well as their obtained loadings are shown in Table 3 and Fig. 5. Among the variables projected as supplementary on the factorial planes (Fig. 5) are most of the major ions, as well as temperature (T), electrical conductivity (EC) and pH. In the many additional runs performed, including the supplementary variables, correlations

between As and other variables were weaker. When including the EC, mineralization dominated the first factorial plane. Including temperature enhanced the seasonality in the first extracted factor due to the shallow sampling depths.

In Fig. 5, both the first factorial planes, explaining 59% and 47% of the total variance of PCA on respectively the seven and four campaign data matrices, show As opposing Eh and dissolved O_2 , and to a lesser extent NO_3^- (and SO_4^{2-} on the four campaign data matrix) on the first axis, reflecting the presence of aerobic vs. reduced environments. PCA on the seven campaign data matrix (Fig. 5a) also shows negative correlation of water depth (WDpt) to the first axis, i.e. an increased occurrence of aerobic conditions with larger water depths, less visible in the reduced data matrix (fewer samples, more variance due to added variables). Well depth (Dpt) shows (weak) inverse correlation with water depth on the first axis, but direct correlation on the second axis of the left PCA plane. When applied to the four campaign data matrix (Fig. 5b), including more parameters, PCA shows positive association of Fe to As, Mn and HCO_3^- on the first axis. The first axis hence shows high dissolved concentrations of these four ions under anaerobic conditions, as well as higher EC values (positive correlation of this supplementary variable). The distribution of the sample factor scores on the factorial planes reveals a certain seasonal separation of

Table 3
Loadings of first four factors extracted by PCA.

	First run: data of seven campaigns				Second run: data of four campaigns			
	F1	F2	F3	F4	F1	F2	F3	F4
Eigenvalue	2.3	1.3	0.9	0.7	3.3	1.4	1.3	1.1
Expl. var. (%)	37.6	21.4	14.9	11.3	32.7	14.5	13.3	11.2
Cum. expl. var.	37.6	59.0	73.9	85.3	32.7	47.2	60.5	71.7
As ($\mu\text{g/l}$)	0.68	-0.03	0.52	0.42	0.61	0.39	-0.24	0.35
Eh (mV)	-0.83	-0.03	-0.15	-0.08	-0.79	-0.25	0.12	-0.10
O ₂ (mg/l)	-0.75	-0.05	0.13	0.54	-0.66	0.24	0.05	0.33
NO ₃ ⁻ (mg/l)	-0.44	0.36	0.73	-0.37	-0.40	-0.19	-0.50	0.50
Dpt (m)	0.28	-0.83	0.13	-0.26	0.37	-0.13	0.68	0.28
WDpt (m)	-0.52	-0.68	0.20	0.06	-0.32	0.17	0.63	0.51
HCO ₃ ⁻ (mg/l)					0.67	-0.53	0.06	0.08
SO ₄ ²⁻ (mg/l)					-0.29	-0.61	-0.28	0.41
Fe (mg/l)					0.71	0.38	-0.23	0.32
Mn (mg/l)					0.64	-0.54	0.10	0.04

Dpt = well depth and WDpt = water depth. Significant loadings in bold.

the samples, with clearly more summer than winter samples showing positive correlation to As. Notwithstanding, many summer samples also show negative correlation to As, as illustrated in Fig. 5. In a few wells (4, 17, 30 and 36) reduced conditions persist almost all year long, though not always enough to maintain high As concentrations. Spring and autumn samples have intermediate positions.

4.3. Factorial regression analysis (FRA)

For the FRA explanatory models, the factor loadings of the classes of the used variables are presented in Table 4 (see Table 1 for class definition). For the first three models, which only include non-physicochemical variables, the classes with the highest negative loadings belong to the variable land use, where rice crop coexists with other crops (mainly vegetables) with or without maize. These classes are directly correlated to the highest As concentration class (As3). The same is true for the other crop classes where rice is included, though somewhat less significant. Other positive associations to As3 (i.e. negative factor loadings) occur for the summer season, and a well depth > 4 m (Dpt2) for models 2 and 3, where this parameter was included. Inverse correlations exist with: other crops (mainly vegetables), without the presence of rice and/or maize (Oth), a water depth

Table 4
Factor loadings of variable classes calculated during FRA.

	Model 1	Model 2	Model 3	Model 4	Model 5
As1	0.23	0.20	0.19	-0.26	-0.29
As2	0.23	0.24	0.20	-0.25	-0.20
As3	-0.66	-0.63	-0.57	0.75	0.74
WDpt1	-0.20	-0.18	-0.19	0.20	0.23
WDpt2	-0.18	-0.18	-0.18	0.18	0.17
WDpt3	0.54	0.53	0.54	-0.54	-0.55
Aut	0.27	0.26	0.26	-0.27	-0.29
Win	0.42	0.43	0.43	-0.42	-0.41
Spr	0.03	0.03	0.03	-0.03	-0.03
Sum	-0.39	-0.39	-0.39	0.39	0.38
NoCr	0.33	0.33	0.33	-0.33	-0.31
Mz	-0.31	-0.32	-0.32	0.30	0.27
Rc	-0.56	-0.56	-0.56	0.56	0.56
Ot	0.45	0.44	0.45	-0.45	-0.47
MzOt	0.31	0.31	0.31	-0.31	-0.31
MzRc	-0.56	-0.55	-0.56	0.56	0.58
RcOt	-1.05	-1.04	-1.04	1.06	1.08
MRot	-1.71	-1.71	-1.71	1.71	1.70
Dpt1		0.17	0.16	-0.16	-0.16
Dpt2		-0.62	-0.62	0.62	0.60
Cr			-0.01		
CrFn			0.11		
Fn			-0.07		
Eh1				1.04	1.04
Eh2				-0.39	-0.39
Eh3				-0.42	-0.42
NO31					0.62
NO32					0.13
NO33					-0.14
NO34					-0.57

Significant loadings in bold.

> 2 m (WDp3) and the winter season (Win), which all have positive loadings. No significant factor loadings are attributed to any of the aquifer lithology classes in model 3. Moreover, adding this variable to the model results in a decreasing (absolute) factor loading for As3. Adding the parameter Eh to model 4 reveals its inverse correlation to As: the lowest Eh class (Eh1) has a high correlation to the extracted factor, just as As3, whose factor loading also increases significantly. The same effect is observed when adding NO₃⁻ to the regression model, along with Eh (model 5), i.e. inverse correlation to As. Moreover, the lowest As class (As1) appears more significantly correlated to aerobic conditions. The signs of all factor loadings in

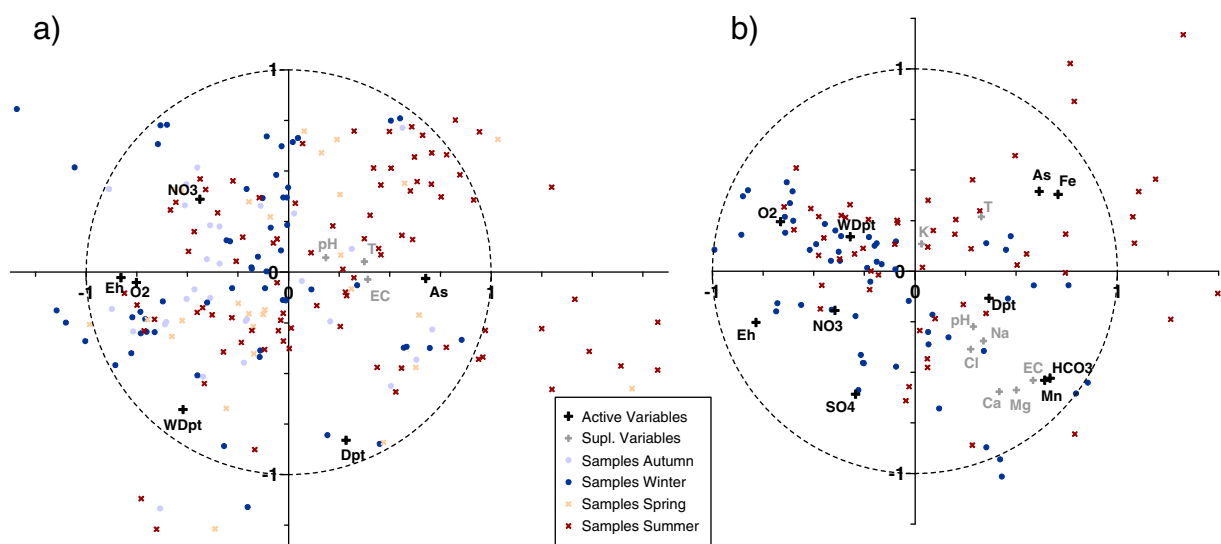


Fig. 5. Primary factorial plane resulting from PCA of a data matrix of (a) 6 active and 3 supplementary variables and 200 samples from 7 campaigns; and (b) 10 active and 8 supplementary variables and 112 samples from 4 campaigns (see Table 3); both variable loadings and sample scores are shown, the latter distributed per season of sampling.

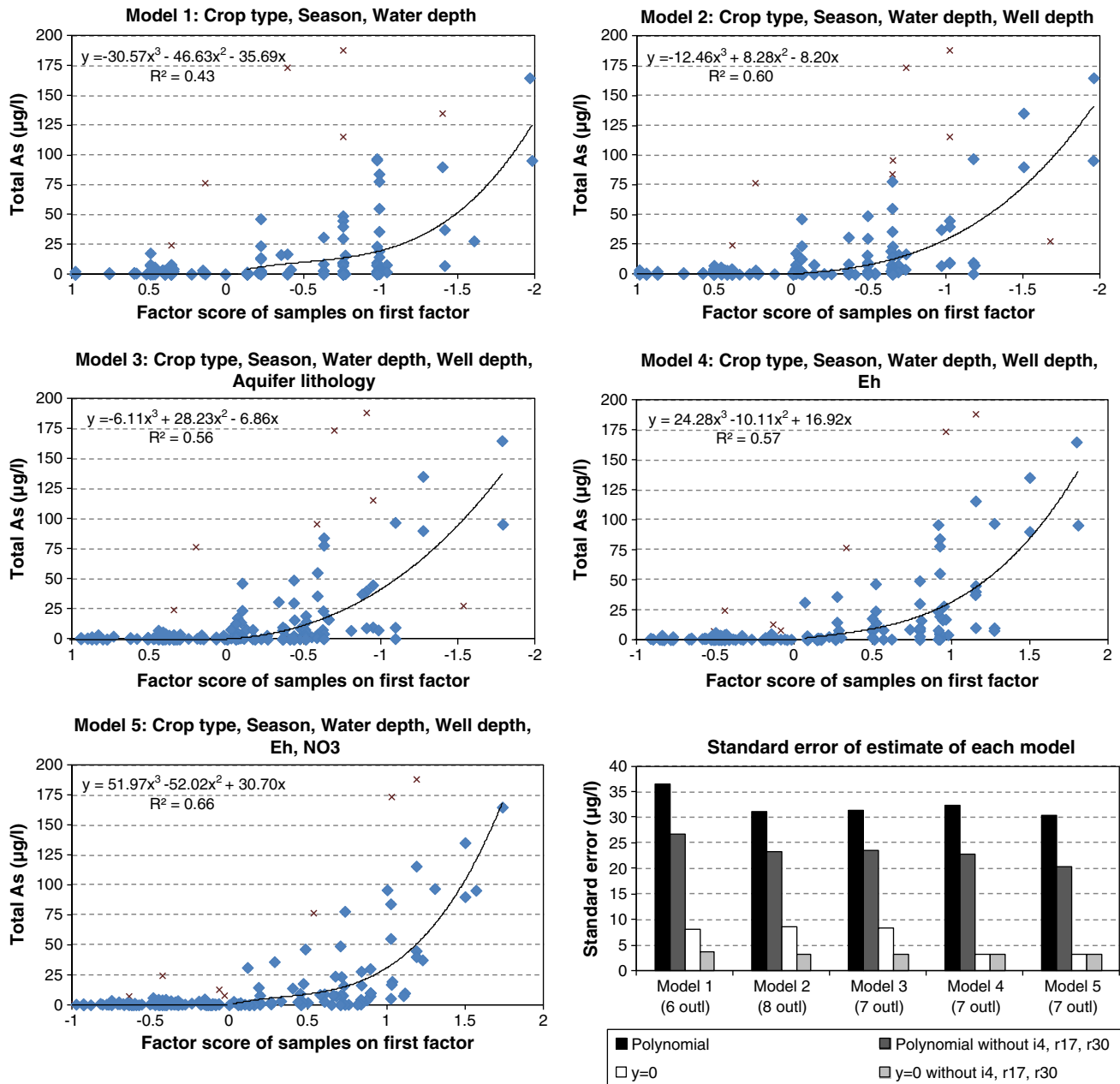


Fig. 6. Results of factorial regression model construction based on the 7 data campaigns, showing the evolution of the regression model, coefficient of determination (R^2) and standard error of estimate (s_R) with the increase of the number of variables considered. Outliers are those that fall outside the confidence interval and are indicated by \times ; the three outliers with the highest, and most underestimated, arsenic concentrations are samples i4, r17 and r30 (letter representing the campaign and number representing the well – see Fig. 3).

models 4 and 5 are inverted in relation with the signs in models 1 to 3, a result of mathematical computations, without any further consequences.

The five regression models of As based on the sample factor scores are presented in Fig. 6. Each model is actually a composite of two models: A) a straight line “ $y=0$ ” for the lowest As concentrations (positive sample scores in models 1 to 3, negative scores in models 4 and 5), which obviously predicts a concentration of 0 $\mu\text{g/l}$; and B) a 3rd order polynomial curve for predicted As concentrations $>0 \mu\text{g/l}$. Confidence bands were calculated for both regression functions in each model, and samples falling outside these bands were considered outliers (indicated in the plots by \times). Of models 1 to 3 using only the non-physicochemical variables, model 2 has the highest R^2 and the lowest s_R (Eq. (3)) as can be seen in the last plot of Fig. 6. Thus, including well depth in the initial model leads to a large increase in R^2 and decrease in s_R . The same does not occur

when including aquifer lithology (model 3). Model 4, which includes Eh, despite the slightly lower R^2 , has much fewer outliers (only three for the polynomial model) and a lower s_R for the “As = 0” regression

Table 5
Average observed and predicted As concentrations for each of the seven campaigns.

Campaign date	Season	Avg obs	Avg model 2	Avg model 4	Avg model 5
09/2001	Summer	19.6	13.1	17.2	16.8
11/2001	Autumn	2.3	0.2	1.2	1.7
01/2002	Winter	2.1	0.0	0.8	1.0
03/2002	Winter	1.8	0.0	0.8	1.1
05/2002	Spring	4.4	7.0	4.4	4.1
07/2002	Summer	26.8	17.3	21.7	24.9
09/2002	Summer	10.3	12.9	13.5	10.3

Outliers i4, r17 and r30 are not included.

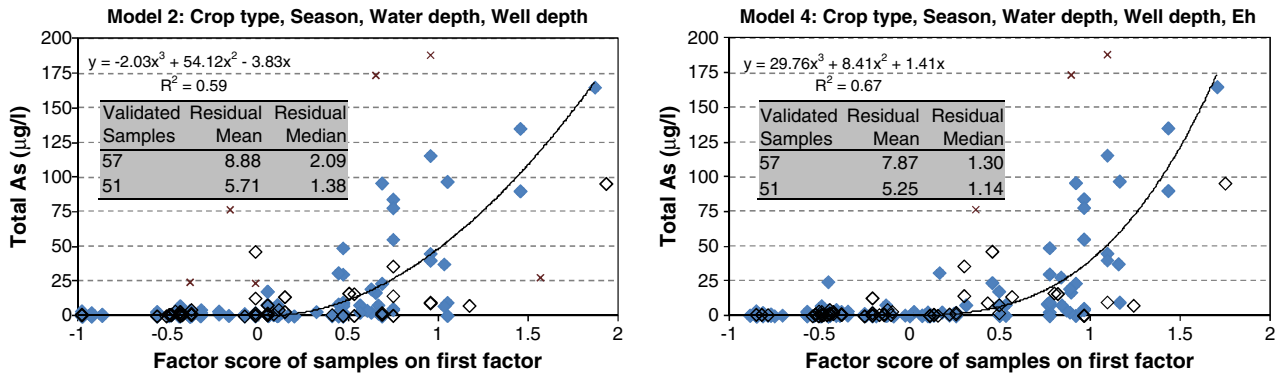


Fig. 7. Results of model construction based on 5 campaigns, for models 2 and 4, leaving out the second winter and summer campaigns of 2002. The latter are represented by the open symbols and were only used for model validation. The plot legends are identical to those of Fig. 6, showing the coefficient of determination (R^2) and standard error of estimate (s_R), as well as the outliers indicated by \times ; the three outliers with the highest, and most underestimated, arsenic concentrations continue to be samples i4, r17 and r30.

model as well as for the polynomial model, when disregarding the 3 major outliers (i4, r17, r30). Finally, by including NO_3^- , model 5 manages to obtain the highest R^2 (0.66; $R = 0.81$) and lowest s_R : 30.5 $\mu\text{g/l}$ for the polynomial model (20.4 $\mu\text{g/l}$ without outliers) and 3.2 $\mu\text{g/l}$ for the “As = 0” model.

Table 5 presents average observed and predicted As concentrations by models 2, 4 and 5 for each of the seven campaigns, excluding the three referred outliers. Model 2 generally has a tendency to underestimate the average As concentration, which happens in five of the seven campaigns. The most accurate model is model 5. However, as Eh is measured in the field and NO_3^- in the laboratory, in practice the application of model 4 would be simpler.

Fig. 7 presents the results of validation of models 2 and 4, where new versions of the models were constructed after removing the second summer and winter campaign data sets of 2002, i.e. the campaigns of

March and September 2002, which were subsequently used to validate the models, comparing predicted to observed concentrations. For each of the two models, 51 of the 57 samples used for validation (near 90%) fall within the confidence band. Mean residuals (absolute errors) for models 2 and 4 are respectively 8.9 and 7.9 mg/l, whereas the median residual is much lower: 2.1 and 1.3 mg/l. Of the 12 samples with observed concentration $> 5 \mu\text{g/l}$, both models manage to predict concentrations $> 5 \mu\text{g/l}$ in eight. Of the samples with observed concentration $< 10 \mu\text{g/l}$, six and three samples have predicted concentrations above 10 $\mu\text{g/l}$ for models 2 and 4, respectively.

Concentrations in the summer are clearly higher and correctly predicted by the regression models. For this season two types of maps were produced using models 2 and 4: a) predicted concentration maps, based on the average of predicted values in the three summer campaigns and b) predicted probability maps of As concentration

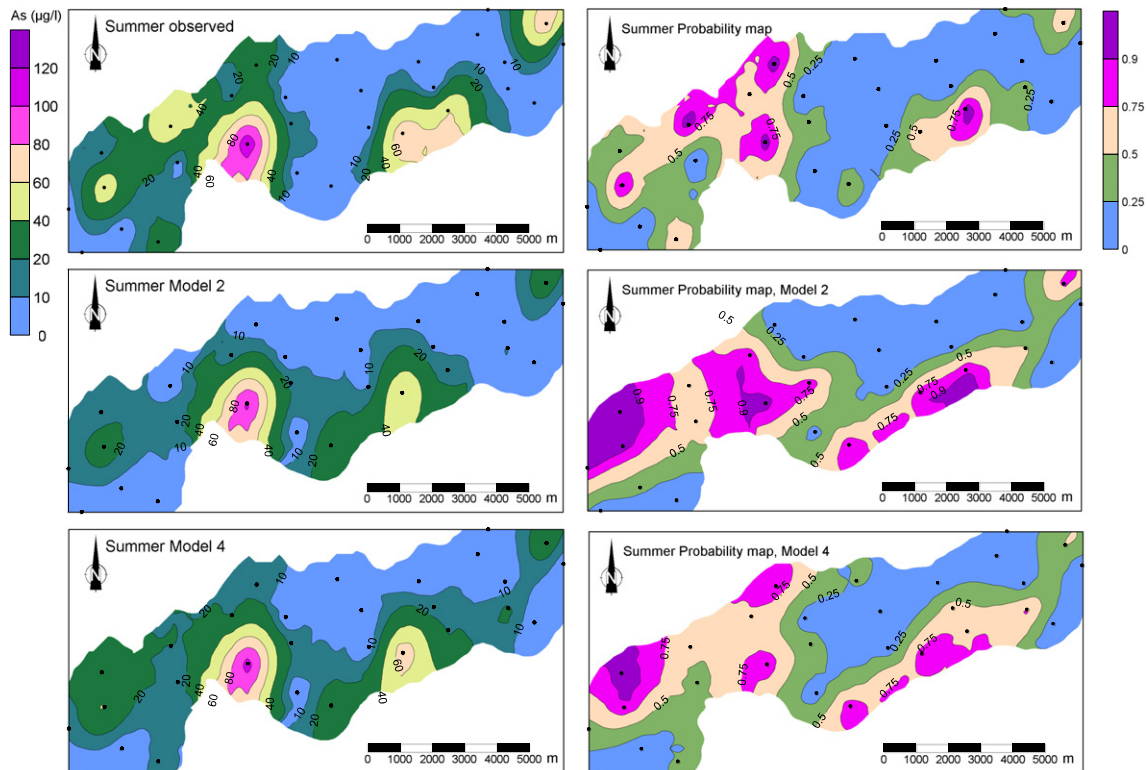


Fig. 8. Predicted As concentration maps (left) and probability maps of As concentrations exceeding the drinking water guideline of 10 $\mu\text{g/l}$ (right), in the summer season, based on FRA model 2 (middle) and model 4 (bottom), compared to observed (top).

exceeding the drinking water guideline of 10 µg/l. These figures are presented in Fig. 8 and compared to the maps constructed on the observed As values (top maps). When compared to the observed spatial distribution of As concentrations, the predicted maps (left side of Fig. 8) reveal some areas of underestimation by the models, particularly model 2, at the southeastern, northeastern and northwestern limits of the study area. On the other hand, the probability maps on the right side of Fig. 8 integrate the uncertainty that exists both within the measurements and in the model predictions. The two modeled maps in this case predict similar or higher probability of occurrence (i.e. exceeding the drinking water guideline), but apparently rarely seem to predict lower probabilities. In this sense these maps can be considered as more conservative.

In each of the maps, the interpolated area is limited by an estimation error standard deviation (SD) contour line which is smaller than the SD of the original data, a value beyond which the interpolation becomes unreliable (Clark, 1979). A smaller value was chosen, as the original data SD contour line resulted in a relative large area beyond the extent of the observation points, where the map is a result of extrapolation rather than interpolation. It can be verified that the monitoring network has an adequate spatial representativity. Applying a cross-validation technique, where measured data are eliminated one by one from the data set and predicted by means of the kriging algorithm, the mean absolute estimation error for the probability map based on the measured data is calculated as 37%. This is higher than that reported by Andrade and Stigter (2011) for the same monitoring network, but based on the probability maps for the variable NO_3^- .

5. Discussion

Both the scatter plots and the multivariate statistical analyses (PCA and FCA) reveal a clear effect of redox state on the presence of As as well as other parameters such as dissolved oxygen (DO), nitrate, sulfate, iron and manganese. The association of high As groundwaters with reduced conditions in alluvial environments is mentioned by Smedley and Kinniburgh (2002). Similar relations are found in other studies (e.g. Phuong et al., 2012; Selim Reza et al., 2010; Kim et al., 2009). According to these authors, this points towards the reductive dissolution of Fe (hydr)oxides as the essential mechanism of As release, a natural origin of As. According to Saha et al. (2010) the process causes an increase of HCO_3^- resulting in the association of this ion with As, while Wang et al. (2012) mention that a significant correlation between As and HCO_3^- suggests the occurrence of solid organic matter as electron donor to the reduction of iron oxyhydroxides. The common origin of total Fe and As and the related increase in alkalinity are supported by the association of these three parameters on the factorial plane of PCA (Fig. 5b).

A higher alkalinity of the As-rich samples also explains the narrow pH range in which they occur, between 6.5 and 7.2, but mostly below 7 (Fig. 4). Slightly higher pH values were observed by Wang et al. (2012) for As-rich groundwater under similar, strongly reducing conditions. Smedley and Kinniburgh (2002) refer to near-neutral environments as most common for As mobilization under strongly reducing conditions. An additional factor explaining the observed pH values below 7 could be the existence of higher initial CO_2 concentrations from root and microbial respiration in the unsaturated zone (Appelo and Postma, 2005; Andrade and Stigter, 2011).

The generally low concentrations of SO_4^{2-} and negative correlation to As and Fe (factorial plane of Fig. 5b) suggest that dissolved As does not result from the oxidation of As-rich sulfide minerals. Phuong et al. (2012) reached similar conclusions in their case study based on low SO_4^{2-} concentrations and high correlations between As, Fe and Eh. Natural sources (atmospheric deposition, mineral dissolution) and the application of inorganic fertilizers can explain the higher SO_4^{2-} concentrations observed at some locations, whereas

incomplete reduction can account for the co-occurrence of As and SO_4^{2-} in a few samples (Fig. 4).

Although a natural origin of arsenic is plausible, anthropogenic sources, particularly through the input of fertilizers, cannot be excluded, for a number of reasons. First, the fact that NO_3^- correlates negatively to As on the first factor in PCA and FCA does not automatically exclude a common source. The mobility and fate of NO_3^- and As are driven by very different conditions. According to the redox sequence the reduction of NO_3^- occurs prior to that of Fe-oxides. In fact NO_3^- reduction itself can be performed through oxidation by Fe(II) bearing minerals (Appelo and Postma, 2005). Under aerobic conditions NO_3^- is a very conservative ion, whereas As is highly immobile, at least in a near-neutral pH range. Kim et al. (2009) mention that in areas with good permeability of surface sediments, the introduction of NO_3^- and SO_4^{2-} from the land surface can limit As concentrations in shallow groundwaters. Fertilizers used in rice culture, a crop shown in FRA to be highly correlated to As concentrations in the study area, often contain reduced forms of nitrogen, and nitrate is rarely found below these crops (Andrade and Stigter, 2009).

Second, where rice is cultivated in the same area as other crops (mainly vegetables, with or without maize), correlation with As is higher (see Table 4), which could be an indicator of an additional source supplied by higher fertilization rates normally occurring in vegetable crops (Goulding, 2000; Ramos et al., 2002). The latter could also account for the apparently significant association of NO_3^- with As on the third factor of PCA (7 campaign data matrix, see Table 3), which could occur locally where As-bearing minerals have consumed oxygen but not yet NO_3^- .

The use of chemical fertilizers in agricultural activity is considered as one of the important factors influencing the mobility of As from soil to groundwater (e.g. see the study of Zahid et al., 2008). High As contents in fertilizers have been reported by several authors (e.g. Campos, 2002; Achryya, 2005; Otero et al., 2005; Farooqi et al., 2007), particularly in phosphorous (P) fertilizers. Arsenic was detected in NPK fertilizers used in basal and top dressing in extensive farming and horticulture in an area in Spain (Otero et al., 2005). Campos (2002) argues that As present in excessively used fertilizers was sufficient to increase aqueous As concentrations to more than 100 µg/l in Brazil. The author further states that the use of P fertilizers may increase As solubility due to the competitive $\text{PO}_4\text{-AsO}_4$ exchange in the soil; the latter observation is corroborated by Smedley and Kinniburgh (2002). In addition to site competition for adsorption, the lack of correlation between As and PO_4^{3-} supporting a common source may further be explained by the overall low mobility of PO_4^{3-} , which can also precipitate with dissolved cations such as calcium (Griffioen, 2007).

Table 4 reveals an association of silt to the highest As concentration class that is positive but too weak to be considered significant. The fact that aquifer lithology does not seem to correlate significantly to As concentrations in groundwater could be an indication of the absence of a strong affinity between the occurrence of As and the presence of finer sediments. Several authors refer to fine sediments as the most likely and abundant source of As in the form of oxide materials (Smedley and Kinniburgh, 2002; Selim Reza et al., 2010; Phuong et al., 2012). However, the lack of observed association in the study area could also be the result of the performed level of well classification not having a high enough resolution. In order to clarify this question, analyses of the sediment must be performed.

Whether of natural or anthropogenic (agricultural) origin, the association of high As concentration in shallow groundwater to the summer season as well as to rice culture, points towards the influence of ponding, which takes place during the growing season of rice (mainly summer), on the mobility of As. Fig. 4 revealing the occurrence of As values over 40 µg/l at temperatures (T) higher than 20 °C, is quite illustrative in this sense. The fact that in PCA, where T was projected as supplementary, the loadings are not that high on either of the PCA planes indicates that aerobic conditions also exist in

the summer, which is true for land not cultivated with rice. This further explains why the factor loading of the summer season in FCA ("Sum" in Table 4) does not reach a very high value (applying PCA and FCA to the rice cultivated area alone would result in more significant factor loadings, but lose its intended purpose).

The winter season appears correlated to lower As concentrations, which can be explained by the lack of an As source from fertilizers or the transport of As to deeper zones. Besides, winter recharge will also lead to dilution, which can play an important role in the shallow aquifer (Andrade and Stigter, 2009). Finally, the immobilization of As due to the aeration of the soil and recharge from rainfall is most likely a relevant process taking place. Diffusion of oxygen into the aquifer can be relatively fast in most of the area, due to the absence of thick layers of very fine-grained material. An exception is formed at four locations (wells 4, 17, 30 and 36), where a low Eh is maintained, which could be related to a higher organic matter content: two of the referred wells are surrounded by trees, one presented animal burrows at the time of sampling and in the remaining one a layer with abundant organic matter was found during well installation. Only two of these wells reveal significant As concentrations in the winter, possibly linked to the persisting reduced environment; slow groundwater flow, a process mentioned by Achryya (2005) and Guo et al. (2008) as occurring in As affected areas under similar conditions, could play an additional role.

Regarding the factorial regression models, it can be concluded that their results are quite promising, particularly considering the fact that the presence of As in groundwater is a complex phenomenon with high spatial variability. Model validation using five data campaigns to construct new model versions and the remaining two to validate the obtained results, reveals that the model accuracy is adequate, with 90% of the predicted concentrations falling within the confidence band of the model. For about 7% of the samples the models do not manage to estimate the observed concentrations above the drinking water guideline. Some authors mention that the complex behavior of As and its high spatial variability make it difficult to model its spatial occurrence, even when the concentrations in adjacent wells are known (e.g. Smedley and Kinniburgh, 2002; Zahid et al., 2008). Meliker et al. (2008) find that their geostatistical explanatory model of As in private well groundwater in southeastern Michigan, accounting for bedrock and unconsolidated geologic boundaries, does not behave better than the nearest neighbor method, which the authors attribute to the high spatial variation of As, further revealed by the existence of a large nugget effect in the sample variogram. The latter was not observed in the current case study, possibly due integration of seasonality into the models. Nevertheless, cross-validation of the kriging interpolation of observed As concentrations resulted in very large residuals (absolute estimation errors). In other words, when modeling the spatial occurrence of As without taking into account existing correlations, interpolated values of As at a well based solely on neighboring wells, can be quite different than the observed value itself. A denser monitoring network would thus be needed to obtain a more representative picture of the spatial distribution.

The latter fact also supports the use of probability maps for management purposes, rather than absolute concentration maps. Besides allowing a rapid identification of the contamination risk at the drinking water guideline level, these maps incorporate the uncertainty inherent to monitoring and spatial variations of As. The uncertainty linked to temporal oscillations is also reduced by integrating seasonality into the maps.

It should be noted that the current study focuses on the upper levels of the alluvium body. Monitoring of deeper layers is almost inexistent, but, given the presence of As in relevant concentrations in the upper sediments, it is mandatory before considering the possible exploitation of deeper groundwater resources for human supply. Under reduced conditions transport of As in groundwater may be slow but steady. It will be necessary to see if changes of As

concentrations in depth follow a typical profile of increasing with depth at shallow levels and decreasing at greater depths, as referred by Kim et al. (2009) for alluvial aquifers where Fe-(hydr)oxide reduction is the major As release mechanism. The inexistence of such a profile may provide additional hints as to the main sources of the high As concentrations in the groundwater of the study site.

6. Conclusions

Multivariate geostatistical modeling has proven a useful tool in providing insights to the problem of As contamination of shallow alluvial groundwater under agricultural land, based on the case study in Central Portugal. These tools have improved the understanding of the role of land use and intrinsic factors such as aquifer lithology and depth, as well as water depth, on the As contamination degree. At the study site, high As concentrations are seen to occur preferably in the summer season areas with rice culture, particularly when associated with vegetable crops, and where the unsaturated zone is thin or almost absent. According to the model the highest concentrations in the shallow groundwater do not occur near the surface, but a few meters deep (more than 4 m). Aquifer lithology does not correlate to contamination levels, which could result from a lack of resolution in the well classification process.

Although the main source of As released into groundwater is believed to be reductive dissolution of naturally occurring Fe (hydr)oxides, such reduced conditions are clearly promoted by human activities, in this case specifically linked to the cultivation of rice. The introduction of As from agricultural products, such as fertilizers, is considered a potentially relevant additional source, despite the negative correlation with NO_3^- and the absence of correlation with PO_4^{3-} . Whereas the former is thought to be a consequence of different fates of NO_3^- and As under aerobic or reduced environments, the latter is possibly a result of site competition between As and PO_4^{3-} for adsorption and of the low mobility of PO_4^{3-} .

The parameters that reveal significant correlation to As were included in factorial regression models that predict As concentration with a known error on the basis of established relationships. One such model based solely on field measured parameters proved adequate to predict As concentrations under these environments. Adding Eh and NO_3^- concentrations to the model constructions resulted in models with lower error. These findings can provide valuable indications for the occurrence of As-rich shallow groundwaters in similar context areas. In addition, factorial regression models could be used when evaluating "if-then" scenarios, i.e. predicting how As concentrations will respond to changes in any of the explanatory variables, particularly those related to human intervention. Finally, including the probability aspect through the geostatistical concept of indicator kriging, allowed associating observed and predicted concentrations to the risk of exceeding the drinking water guideline. It further permitted accounting for the uncertainty inherent to monitoring and analyzing the representativeness of the monitoring network.

Acknowledgments

The authors wish to express their thanks to the *Centro de Geofísica da Universidade de Coimbra*, the *Departamento de Ciências da Terra*, the *Centro de Geociências da Universidade de Coimbra* and the *Instituto de Investigação da Água*, for their support during the progress of this study, to the farmers who allowed the installation of the observation wells on their lands, to the *Escola Superior Agrária de Coimbra* for providing the equipment for well installation, to Eng. Filipe Melo and Sr. José Borralho for their help during the installation of the wells and for useful agricultural information and finally to the *Associação de Beneficiários da Obra de Fomento Hidroagrícola do Baixo Mondego* and the *Direcção-Geral de Agricultura e Desenvolvimento Rural—Projecto Mondego* for their support in the field work and supply of relevant

data. The second author holds a post-doctoral research position under the Ciencia 2007 program financed by the FCT – Fundação para a Ciência e a Tecnologia.

References

- Achryya SK. Arsenic levels in groundwater from Quaternary alluvium in the Ganga Plain and the Bengal Basin, Indian subcontinent: insights into influence of stratigraphy. *Gondwana Res* 2005;8:55–66.
- Almeida AC, Soares AF, Cunha L, Marques JF. Proémio ao estudo do baixo Mondego. *Biblos*; 1990. p. 17–47. [LXVI].
- Almeida C, Mendonça JL, Jesus MR, Gomes AJ. Sistemas aquíferos de Portugal Continental. Lisboa: Instituto da Água/Centro de Geologia da Universidade de Lisboa; 2000 [3 volumes].
- Anawar HM, Akai J, Mostofa KMG, Safiullah S, Tareq SM. Arsenic poisoning in groundwater health risk and geochemical sources in Bangladesh. *Environ Int* 2002;27:597–604.
- Andrade AIASS. Hidrogeoquímica e caracterização da contaminação hídrica nos corpos aluvionares do Mondego entre Coimbra e Montemor-o-Velho. PhD Thesis, Universidade de Coimbra, Portugal; 2005.
- Andrade AIASS, Stigter TY. Multi-method assessment of nitrate and pesticide contamination in shallow alluvial groundwater as a function of hydrogeological setting and land use. *Agric Water Manag* 2009;96:1751–65.
- Andrade AIASS, Stigter TY. Hydrogeochemical controls on shallow alluvial groundwater under agricultural land: case study in central Portugal. *Environ Earth Sci* 2011;63:809–25.
- Antunes IMHR, Neiva AMR, Silva MMVG. The mineralized veins and the impact of old mine workings on the environment at Segura, central Portugal. *Chem Geol* 2002;190:417–31.
- Appelo CAJ, Postma D. *Geochemistry, groundwater and pollution*. 2nd ed. A.A. Balkema Publishers; 2005.
- Ávila P, Pereira MR, Novais H, Ferreira A. Arsénio nos solos, sedimentos e águas na envolvente da mina do Tuela (Vinhais, NE Portugal). In: *Mirão J, Balbino A, coord. Livro de resumos do VII Congresso Nacional de Geologia vol. II*; 2006. pp. 443–446.
- Barbosa BP, Soares AF, Rocha RB, Manupella G, Henriques MH. Carta Geológica de Portugal. Notícia explicativa da folha 19-A Cantanhede. *Serviços Geológicos de Portugal*; 1988.
- Benzécri JP. L'analyse des donnees, v.2-L'analyse des correspondences. *Dunod*; 1973.
- BGS, MacDonald M. Phase I: groundwater studies of arsenic contamination in Bangladesh. Executive summary, main report; 2000 [Accessed at: <http://bicn.com/acic/resources/infobank/bgs-mmi/risumm.htm> on 06.07.2012].
- Brown CE. *Applied multivariate statistics in geohydrology and related sciences*. Springer-Verlag; 1998.
- Cama J, Rovira M, Ávila P, Pereira MR, Asta MP, Grandia F, et al. Distribución de arsénico en la región Ibérica. In: *Bundschuh J, Pérez Carrera A, Litter MI, editors. IBEROARSEN Distribución del arsénico en las regiones Ibérica e Iberoamericana*. Argentina: CYTED; 2008. p. 95–136. [accessed at: <http://www.cnea.gov.ar/xxi/ambiental/iberoarsen/publicaciones.asp> on 24.11.2011].
- Campos V. Arsenic in groundwater affected by phosphate fertilizers at São Paulo, Brazil. *Environ Geol* 2002;42:83–7.
- Chakraborti D, Rahman MM, Das B, Murrill M, Dey S, Mukherjee SC, et al. Status of groundwater arsenic contamination in Bangladesh: a 14-year study report. *Water Res* 2010;44:5789–802.
- Clark I. *Practical geostatistics*. Applied Science Publishers Ltd.; 1979.
- Cristo FP. Águas subterrâneas no Baixo Mondego. *Projecto Praxis XXI 2/2.1/CTA – 156/94 – O Baixo Mondego: organização, geossistemática e recursos naturais*, Coimbra, Portugal; 1998. p. 105–23.
- Deutsch WJ. *Groundwater geochemistry*. Lewis Publishers; 1997.
- Deutsch CV, Journel AG. *Geostatistical software library and user's guide*. New York, USA: Oxford University Press; 1998 [369 pp.].
- DGADR. Aproveitamento Hidroagrícola do Baixo Mondego. *Direção-Geral de Agricultura e Desenvolvimento Rural*; 2008 [Accessed at http://www.dgadr.pt/ar/a_hidroagricolas/exploracao/ahmond.htm on 01.11.2008].
- Duker AA, Carranza EJM, Hale M. Arsenic geochemistry and health. *Environ Int* 2005;31:631–41.
- EC. Council Directive 98/83/CE of 3 November 1998 on the quality of water intended for human consumption. *Off J Eur Communities* 1998;L330:32–54.
- EPA. EPA national primary drinking water regulations. United States Environmental Protection Agency; 2009 [Accessed at <http://www.epa.gov/safewater/consumer/pdf/mcl.pdf> on 05.12.2011].
- Farnham IM, Johannesson KH, Singh AK, Hodge VF, Stetzenbach KJ. Factor analytical approaches for evaluating groundwater trace element chemistry data. *Anal Chim Acta* 2003;490:123–38.
- Farooqi A, Masuda H, Kusakabe M, Naseem M, Firdous N. Distribution of highly arsenic and fluoride contaminated groundwater from east Punjab, Pakistan, and the controlling role of anthropogenic pollutants in the natural hydrological cycle. *Geochem J* 2007;41:213–34.
- Fetter CW. *Applied hydrogeology*. 4th ed. Prentice-Hall Inc.; 2001.
- García-Sánchez A, Moyano A, Mayorga P. High arsenic contents in groundwater in central Spain. *Environ Geol* 2005;47:847–54.
- Garelick H, Jones H, Dybowska A, Valsami-Jones E. Arsenic pollution sources. In: *Whitacre DM, editor. Reviews of environmental contamination*, 197. Springer Science + Business Media, LLC; 2008. p. 17–60.
- Gómez JJ, Lillo J, Sahún B. Naturally occurring arsenic in groundwater and identification of the geochemical sources in the Duero Cenozoic Basin, Spain. *Environ Geol* 2006;50:1151–70.
- Goulding K. Nitrate leaching from arable and horticulture land. *Soil Use Manag* 2000;16:145–51.
- Griffioen J. Extent of immobilisation of phosphate during aeration of nutrient-rich, anoxic groundwater. *J Hydrol* 2007;320:359–69.
- Guo H, Yang S, Tang X, Li Y, Shen Z. Groundwater geochemistry and its implications for arsenic mobilization in shallow aquifers of the Hetao Basin, Inner Mongolia. *Sci Total Environ* 2008;393:131–44.
- Hazen A. Discussion: dams on sand foundations. *Trans Am Soc Civ Eng* 1911;73:199–203.
- Hopenhayn C. Arsenic in drinking water: impact on human health. *Element* 2006;2:103–7.
- Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 1933;24:417–41.
- Hudak PF. Distribution of arsenic concentrations in groundwater of the Seymour Aquifer, Texas, USA. *Int J Environ Health Res* 2008;18:79–82.
- IARC. *Monographs on the evaluation of carcinogenic risks to humans: arsenic, metals, fibres, and dusts. A review of human carcinogens, 100 C*. Lyon: International Agency for Research on Cancer; 2012. p. 41–94.
- INAG. Estabelecimento de limiares nas águas subterrâneas. *Ministério do Ambiente e do Ordenamento do Território, Instituto da Água*; 2009.
- Johannesson KH, Stetzenbach KJ, Kreamer DK, Hodge VF. Multivariate statistical analysis of arsenic and selenium concentrations in groundwaters from south-central Nevada and Death Valley, California. *J Hydrol* 1996;178:181–204.
- Kar S, Maity JP, Jean J-S, Liu C-C, Nath B, Yang H-J, et al. Arsenic-enriched aquifers: occurrences and mobilization of arsenic in groundwater of Ganges Delta Plain, Barasat, West Bengal, India. *Appl Geochem* 2010;25:1805–14.
- Kim K, Moon JT, Kim SH, Ko KS. Importance of surface geologic condition in regulating As concentration of groundwater in the alluvial plain. *Chemosphere* 2009;77:478–84.
- Klump S, Kipfer R, Cirpka OA, Harvey CF, Brennwald MS, Ashfaq KN, et al. Groundwater dynamics and arsenic mobilization in Bangladesh assessed using noble gases and tritium. *Environ Sci Technol* 2006;40:243–50.
- Lachance M, Bobee B, Gouin D. Characterization of the water-quality in the Saint-Lawrence river—determination of homogeneous zones by correspondence-analysis. *Water Resour Res* 1979;15:1451–62.
- Liu C-W, Lin K-H, Kuo Y-M. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci Total Environ* 2003;313:77–89.
- Liu C-W, Jang C-S, Liao C-M. Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). *Sci Total Environ* 2004;321:173–88.
- Manahan SE. *Environmental chemistry*. 6th ed. Lewis Publishers; 1994.
- Meliker JR, Avruskin GA, Slotnick MJ, Goovaerts P, Schottenfeld D, Jacquez GM, et al. Validity of spatial models of arsenic concentrations in private well water. *Environ Res* 2008;321:42–50.
- Mendes B, Oliveira JFS. *Qualidade da água para consumo humano*. Lidel: edições técnicas, lda; 2004.
- Moore JN, Woessner WW. Arsenic contamination in the water supply of Milltown, Montana. In: *Welch AH, Stollenwerk KG, editors. Arsenic in ground water. Geochemistry and occurrence*. Springer; 2003. p. 329–50.
- Morgada ME, Mateu M, Bundschuh J, Litter MI. Arsenic in the Iberoamerican region. The IBEROARSEN network and a possible economic solution for arsenic removal in isolated rural zones, 5(5). e-Terra; 2008 [11 pp.].
- Moyano A, García-Sánchez A, Mayorga P, Anawar HM, Alvarez-Ayuso E. Impact of irrigation with arsenic-rich groundwater on soils and crops. *J Environ Monitor* 2009;11:498–502.
- Otero N, Vitória L, Soler A, Canals A. Fertiliser characterisation: major, trace and rare earth elements. *Appl Geochem* 2005;20:1473–88.
- Pearson K. On lines and planes of closest fit to systems of points in space. *Phil Mag* 1901;6(2):559–72.
- Pereira HJ, Sousa AJ. *Análise de dados para o tratamento de quadros multidimensionais: textos de apoio ao Curso Intensivo de Análise de Dados 1988–2000*. Lisbon, Portugal: CVRM, Instituto Superior Técnico; 2000.
- Pereira MR, Montes R, Gomes E, Ferreira A, Ávila P. Geochemistry of soils and waters from the abandoned Freixeda Gold Mine, Northeast Portugal, 10. *Macla*; 2008. p. 136–7.
- Phuong NM, Kang Y, Sakurai K, Sugihara M, Kien CN, Bang ND, et al. Arsenic contamination in groundwater and its possible sources in Hanam, Vietnam. *Environ Monit Assess* 2012;184:4501–15.
- Ramos C, Agut A, Lidón AL. Nitrate leaching in important crops of the Valencian Community region (Spain). *Environ Pollut* 2002;118:215–23.
- Rebello F, Cunha L, Almeida AC. Contribuição da geografia física para a inventariação das potencialidades turísticas do Baixo Mondego. *Cadernos de Geografia*, 9. ; 1990. p. 3–34.
- Reimann C, Caritat P. *Chemical elements in the environment*. Springer; 1998.
- Saha D, Sarangam SS, Dwivedi SN, Bhartiya KG. Evaluation of hydrogeochemical processes in arsenic-contaminated alluvial aquifers in parts of Mid-Ganga Basin, Bihar, Eastern India. *Environ Earth Sci* 2010;61:799–811.
- Selim Reza AHM, Jean J-S, Lee M-K, Yang H-J, Liu C-C. Arsenic enrichment and mobilization in the Holocene alluvial aquifers of the Chapai-Nawabganj district, Bangladesh: a geochemical and statistical study. *Appl Geochem* 2010;25:1280–9.
- Siegel FR. *Environmental geochemistry of potentially toxic metals*. Springer-Verlag; 2002.
- Smedley PL, Kinniburgh DG. A review of the source, behaviour and distribution of arsenic in natural waters. *Appl Geochem* 2002;17:517–68.

- Soares AF, Marques JF, Rocha RB. Contribuição para o conhecimento geológico de Coimbra. *Memórias e Notícias*, 100. Publ Mus Lab Mineral Geol Univ Coimbra; 1985. p. 41–71.
- Stigter TY, Almeida P, Carvalho Dill AMM, Ribeiro L. Influence of irrigation on groundwater concentrations in areas considered to have low vulnerability to contamination. In: Bocanegra EM, Hernández MA, Usunoff E, editors. *Selected papers on hydrogeology 6 – groundwater and human development*. Leiden, the Netherlands: A.A. Balkema; 2005. p. 69–85.
- Stigter TY, Ribeiro L, Carvalho Dill AMM. Factorial regression models of groundwater contamination by agricultural practices—a case study in the south of Portugal. *J Hydrol* 2008;357:42–56.
- Stüben D, Berner Z, Chandrasekharam D, Karmakar J. Arsenic enrichment in groundwater of West Bengal India: geochemical evidence for mobilization of As under reducing conditions. *Appl Geochem* 2003;18:1417–34.
- Wang Y, Jiao JJ, Cherry JA. Occurrence and geochemical behavior of arsenic in a coastal aquifer–aquitard system of the Pearl River Delta, China. *Sci Total Environ* 2012;427–428:286–97.
- WHO. Arsenic, drinking-water and health risks substitution in arsenic mitigation: a discussion paper. World Health Organization WHO/SDE/WSH/03.06; 2003 [accessed at http://www.who.int/water_sanitation_health/dwq/WSH03.06fulltext.pdf on 09.07.2012].
- WHO. Guidelines for drinking-water quality. 4th ed. World Health Organization; 2011 [accessed at http://www.who.int/water_sanitation_health/publications/2011/dwq_guidelines/en/ on 09.11.2011].
- Zahid A, Hassan MQ, Balke K-D, Flegr M, Clark DW. Groundwater chemistry and occurrence of arsenic in the Meghna floodplain aquifer, southeastern Bangladesh. *Environ Geol* 2008;54:1247–60.