Journal of Geochemical Exploration 143 (2014) 43–53

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Natural background level assessment in groundwaters: probability plot versus pre-selection method

Elisabetta Preziosi *, Daniele Parrone, Andrea Del Bon, Stefano Ghergo

CNR–IRSA, National Research Council–Water Research Institute, Via Salaria km 29,300, PB 10, 00015 Monterotondo, Rome, Italy

ARTICLE INFO

Article history:
Received 27 June 2013
Accepted 8 March 2014
Available online 24 March 2014

Keywords:
Geochemical background
Aquifer
Arsenic
Manganese
Pollution
Threshold value

ABSTRACT

The need for a sound methodology to establish a formal value separating natural geochemical background levels in groundwaters from the anthropogenic impact, is perceived especially from a legal point of view. In this research we compare the natural background levels derived for As, F, Mn and Fe in a volcanic-sedimentary aquifer in Central Italy, using probability plots and pre-selection methods. The derived values may differ by even two orders of magnitude for strongly skewed distributions. Results show that both methods have a degree of subjectivity, however the integration of the two methods reinforces the validity of the assessment. Critical issues include statistical data distribution choice, minimum population size, distance of the sampling sites from anthropized areas. Pre-selection methods appear more robust than probability plots when the hydrogeological asset is duly considered. The greater importance of the aquifer conceptual model of the system, than the choice of a statistical method, is pointed out.

Introduction

Groundwaters are threatened by anthropic activities and pollution is affecting a large number of aquifers worldwide. In order to distinguish anthropogenic pollution from natural origin contamination in groundwaters, it is important to assess the natural concentration of a substance or element, prior to the onset of anthropic activities. Groundwater geochemical composition, as well as that of surface water and soil, is mainly related to geological factors such as large-scale tectonic provinces and geochemical properties variation of lithological units, although contamination resulting from industrialized areas and intensive agriculture often overlaps on the natural features. As for Europe, Reimann and Birke (2010) provide an overview of the links between geological features and groundwater geochemistry at the continental scale, although limited to those groundwater bodies exploited for bottled waters, hence in good status, while poor quality groundwaters are not considered in this analysis. Despite that this overview is not fully exhaustive, it might represent a baseline reference for the inorganic compounds in groundwaters, along with the derived national scale published papers (e.g. Cicchella et al., 2010; Dinelli et al., 2010), to distinguish the anthropogenic contamination, including enhanced mineral dissolution due to redox conditions variation, from the natural background level (NBL). NBLs are the result, among others, of water–rock interaction, chemical and biological processes both in the vadose and saturated zone, interactions with other water bodies, residence time, and atmosphere and rainfall composition (Edmunds and Shand, 2004; European Commission, 2009; Wendland et al., 2005). How to separate the contribution due to human activities from the background is still under debate (Matschullat et al., 2000; Reimann et al., 2005). The need for a sound methodology to establish a formal value of NBL is perceived especially from a legal point of view.

In the European context, the European Community (2006) requests the EU Member States to derive appropriate threshold values (TV) for several potentially harmful substances, taking into account NBLs when necessary, in order to assess the chemical status of groundwater bodies. In addition to this, the possible negative effects of human activities are evaluated at the site scale, when groundwaters are threatened by potentially polluting activities such as a dump filling. As for Italy, specific compliance values have been set at the national level by the national legislation ("Contamination Threshold Concentration", CSC in the following) as the local binding levels (National Decree D.Lgs.152/2006). In Latium, central Italy, values exceeding the CSC have been detected in groundwaters at several municipal dumps, for some substances including As, F, Fe and Mn. However, the natural origin of these elements is well known in these aquifers hosted by Pleistocene volcanic formations and Quaternary alluvial deposits (Dinelli et al., 2010; Preziosi et al., 2010). In this case CSCs can be locally modified to account for the natural background composition, and the determination of local NBLs is crucial to detect the overlap of anthropogenic contamination and eventually quantify it. The extent of natural contamination for inorganic elements such as F, Fe and Mn in highly polluted areas and the importance of NBL determination in the Italian legislative context is also addressed by Corniello and Ducci (2014).

* Corresponding author. Tel.: +39 06 90672 778; fax: +39 06 90672787.
E-mail addresses: preziosi@irsa.cnr.it (E. Preziosi), parrone@irsa.cnr.it (D. Parrone), delbon@irsa.cnr.it (A. Del Bon), ghergo@irsa.cnr.it (S. Ghergo).

http://dx.doi.org/10.1016/j.gexplo.2014.03.015
0375-6742/

The presence of natural origin trace elements, including arsenic and fluoride, in tap waters and bottled waters extracted from aquifers in Italy, was documented at the national scale by Cicchella et al. (2010) and Dinelli et al. (2012). The relations of the groundwater composition with the regional geological asset, particularly the role of alkaline volcanites in central Italy, are also described by Dinelli et al. (2010).

Statistical techniques for the separation of a background and a non-background population (i.e. the "anomalous" part of the dataset) were originally developed in exploration geochemistry. They were directed to the identification of the anomaly which could correspond to an ore deposit: in that case, the attention was not focused on the background but on the outliers. In this sense, since the early 1960's (Hawkes and Webb, 1962) statistical procedures for the identification of data outliers were developed. Among these, the calculation of the mean plus two standard deviations was commonly used, implicitly assuming a normal distribution of the dataset. Thence, about 2.5% of the upper extreme values was selected, on which the investigations should concentrate. However, geochemical data rarely follow a normal distribution, and the mean plus two standard deviations method, strongly influenced by the presence of extreme values, could be unsuitable. The most appropriate distribution of geochemical data, is still under discussion: a lognormal distribution has been claimed (Ahrens, 1954; Sinclair, 1974) while more recently others have argued that natural geochemical distributions are typically neither normal nor lognormal (Edmunds and Shand, 2004; Reimann and Filzmoser, 2000). They are usually skewed (Edmunds and Shand, 2004; Edmunds et al., 2003; Reimann and Filzmoser, 2000; Reimann et al., 2005) and have one or several outliers. For this reason the use of the median plus twofold the median of the absolute deviations from the median of all the data (MAD, Tukey, 1977) has been suggested for skewed distributions, as MAD is little influenced by the presence of outliers, thus it is considered more suitable for the background assessment (Reimann et al., 2005). The use of the boxplot (in particular, the upper whisker) for the detection of outliers and the estimation of the natural background was also suggested by Reimann et al. (2005), who perform a comparison of the results obtained with different methods and conclude that the boxplot is most informative if the number of outliers is below 10% while the median + 2MAD is preferable if the proportion of outliers is above 15%.

Besides these methods "routinely used in geochemistry" (Reimann et al., 2005), other approaches based on the use of cumulative probability plots were developed (Sinclair, 1974; Tennant and White, 1959). Probability plots allow the display of the data, and graphically observe the main trends, discontinuities, and outliers. They are grounded on the principle that different sources generate different populations of data that can be separated by statistical procedures. The distribution must be decided a priori, most often the lognormal is used (Panno et al., 2006; Sinclair, 1974), although normal, gamma or others are also used (Molinari et al., 2012; Wendland et al., 2005). Aligned points on a probability plots may indicate that samples belong to a single population that originates from a unique process. Changes in the slope of the curves may correspond to the transition between different populations, hence they might separate the natural component from the anthropic one, but they could also mark natural variation of geochemical facies as well or a local geochemical anomaly.

Today two groups of methods are essentially proposed, the first ascribed to the above mentioned probability plots (PP method) (Edmunds and Shand, 2008; Edmunds et al., 2003; Panno et al., 2006; Walter, 2008; Wendland et al., 2005). The second, more recently developed, is based on the selection of the uninfluenced water samples corresponding to the natural population (Pre-Selection method, "PS"). The PS methods (Coetsiers et al., 2009; Ducci and Sellerino, 2012; European Commission, 2009; Gemitzi, 2012; Griffioen et al., 2008; Hinsby et al., 2008; Marandi and Karro, 2008; Molinari et al., 2012; Muller et al., 2006; Preziosi et al., 2010; Rotiroti and Fumagalli, 2013; Wendland et al., 2008) propose to select only those samples which are not, or very little, influenced by human activities, e.g. removing those with high nitrate content or other markers of anthropic contamination such as synthetic compounds. In the residual dataset one value, usually one percentile, is chosen as a representative of the NBL, meaning that all concentrations exceeding that level should be ascribed to anthropogenic sources. In the examined literature different percentiles (e.g. 90th, 95th, and 97.7th) have been proposed so far, broadly depending on the degree of knowledge of the hydrogeochemical system.

In this paper we compare the NBLs derived with the two methods for As, F, Mn and Fe in a groundwater body hosted by a volcanic-sedimentary aquifer in central Italy, where groundwaters are threatened by an important municipal dump and other industrial activities. The derived NBLs should be applied to detect the overlap of the anthropic contamination to the natural background in the industrial area, called "focus area" in the following. The NBLs derived for the focus area are then compared to the NBLs derived for the whole groundwater body. As claimed by Salminen and Tarvainen (1997) about sediment, natural background concentrations can vary widely due to differences in basic geology and this is equally true for groundwater. The main aim of this research is to evaluate how far the NBLs estimated from field data are sensitive to the chosen procedure. The importance of the conceptual model of the aquifer for the NBLs' derivation is also discussed.

Data and methods

Study area description

The study area is located in northern Latium (central Italy) and extends for about 340 km^2 (Fig. 1). It is characterized by K-alkaline volcanic rocks (middle Pleistocene), which rest on continental alluvial deposits (Lower Pleistocene–Middle Pleistocene), and marine sediments of the Plio-Pleistocene. In the northern part of the area groundwaters circulate in the volcanic rocks, which are dominant in the north but they decrease in thickness and extension toward south. In the central and southern part of the area groundwaters circulate mainly in the sedimentary formations, whose deposition happened prior to the volcanism onset, hence they have a quite different geochemical composition as they derive mainly from the degradation of Meso-Cenozoic carbonatic sediments. Hydraulic heads decrease from north to south from about 120 m a.s.l. in the northern sector to altitudes close to sea level near the valley of the River Tiber. See Parrone et al. (2013) and Preziosi et al. (2013) for a more detailed description of the hydrogeological setting of the region. The industrial activities are concentrated in the southern end of the study area.

Sampling and laboratory methods

Aiming at a homogenous spatial distribution of water points, 50 samples were collected at 43 private wells or piezometers, 1 spring and 6 fountains from February to June 2011 (Fig. 1), including four samples collected within or very close to the landfill site to be used for comparison ("influenced samples"). The final target was a statistically significant number of samples, not influenced by industrial activities of the site nor by other anthropogenic disturbances (e.g. upgradient and/or far enough from the site). A practical limitation was the availability of already existing wells, as the drilling of new boreholes was not planned in the project, for budget reasons. The sampling strategy was based on the need to obtain a dataset consisting of a significant number of samples, with the constraint of being not too close to the focus area, and within the same hydrogeological context at the same time. The final density of the sampling network is a result of this compromise.

Water samples were filtered at 0.4 μm in nitrogen flux, collected in HNO_3 1% treated polyethylene bottles, and stored in refrigerated containers. One fraction was acidified in the field after filtration for major cation and trace element determination (1 mL HNO_3); additionally two fractions were frozen for future record. T, Eh, pH, EC, DO, were