Elsevier Editorial System(tm) for Science of the Total Environment
Manuscript Draft

Manuscript Number: STOTEN-D-18-08752R2

Title: Groundwater diffuse pollution in Functional Urban Areas: the need to define anthropogenic Diffuse Pollution Background Levels

Article Type: Research Paper

Keywords: Natural Background Level; Groundwater quality status; Urban areas; Diffuse Pollution; chlorinated solvents; hexavalent chromium

Corresponding Author: Dr. Arianna Azzellino, Ph.D

Corresponding Author's Institution: POLITECNICO DI MILANO

First Author: Arianna Azzellino, Ph.D

Order of Authors: Arianna Azzellino, Ph.D; Loris Colombo, PhD; Silvia Lombi; Valeria Marchesi; Alessandra Piana; Andrea Merri; Luca Alberti, PhD

Abstract: Groundwater status in highly urbanized areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. This makes very often challenging to determine whether the observed groundwater conditions are the result of localized pollutant sources (PS-Point Sources). In the EU legislative framework it is accepted that, when Natural Background Levels (NBLs) of undesirable elements are proven to be higher than specific Groundwater Quality Standards (GQSs), NBLs should be assumed as GQSs. No procedure is instead defined when anthropogenic Diffuse Pollution Background levels (DBPLs) are higher than GQSs and make unfeasible any remediation strategy. Among the many contaminants affecting groundwater, the chlorinated solvents, tetrachloroethylene (PCE), trichloroethylene (TCE) and trichloromethane (TCM) among the organics and hexavalent chromium, among the inorganics, having been widely used in several industries all over Europe, are very often the most prevalent contaminants in soil and groundwater. Aim of this paper is to discuss a multivariate statistical approach to address the issue of identification of anthropogenic Diffuse Pollution Background Levels. With such aim, an area of about 1600 km2, including the Functional Urban Area of Milan, was considered and 10 independent geochemical datasets, provided by local and regional agencies, and covering the period 2003-2014 were merged into a single database after homogenization and multiple quality checks. A total of 618258 chemical analyses from 3477 sampling wells were considered, being all samples collected and analyzed through internally consistent protocols. The analysis enabled to identify five main clusters, having specific hydrogeological characteristics, different temporal profiles and pollutant background concentration levels, which were also found to respond differently to meteo-climatic changes. This study offers a robust knowledge basis for drafting a diffuse pollution management plan of the area.

Response to Reviewers: ANSWERS TO REVIEWER N. 4

"Please make introduction easier to understand, this version is still not brief enough."

The introduction has been simplified and shortened according to the reviewer's recommendation.

ANSWERS TO REVIEWER N. 7
"Obviously, this paper did much improvement. However, tables are still too many, some tables are not necessary. For example, Table 3 is redundant, important information in table 3 was also shown in table 4. In my opinion, this paper still need a minor revision."

We agree with the comment. We deleted table 3 and table 5, including some comment lines concerning these two tables directly in the manuscript

Cover Letter

Politecnico di Milano - University of Technology
Civil and Environmental Engineering Department (DICA)
Environmental Division

POLITECNICO DI MILANO

Milano, 26th November 2018

Dr. José Virgílio Cruz

Associated Editor *Science of the Total Environment*

Dear Editor

Also on the behalf of my coauthors, I am sending the revised version of manuscript of the paper:

GROUNDWATER DIFFUSE POLLUTION IN FUNCTIONAL URBAN AREAS: THE NEED TO DEFINE DIFFUSE POLLUTION BACKGROUND LEVELS (DPBLS)TO BE SUBMITTED AS A RESEARCH PAPER TO THE SCIENCE OF THE TOTAL ENVIRONMENT

The manuscript has been thoroughly revised based on the reviewers' comments and suggestions. All the changes made have been reported in the word file as track changes. English language and grammar was also checked.

We believe that the reviewers comments have significantly contributed to improve the manuscript and we hope that it can be now acceptable to be published on STOTEN

Sincerely

Dr. Arianna Azzellino

Corresponding author

*Title Page

Groundwater diffuse pollution in Functional Urban Areas: the need to define anthropogenic Diffuse Pollution Background Levels

Azzellino Arianna[a], Colombo Loris [a], Lombi Silvia[a], Marchesi Valeria [b], Piana Alessandra[b], Andrea Merri[b], Alberti Luca [a,c]

[a] Politecnico di Milano, DICA, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy
[b] ARPA Lombardia, Via Rosellini n° 17, Milano, 20124, Italy.

Corresponding author:

Prof. Arianna Azzellino, PhD Environmental Engineering

Politecnico di Milano, University of Technology,

DICA Environmental Engineering Department

Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

ph. +39 02 23 99 64 31

fax. +39 02 23 99 64 99

e-mail: arianna.azzellino@polimi.it

Dear editors and reviewers,

We would like once again to thank the reviewers for the useful suggestions. We believe based on these comments the quality of the manuscript was further improved. All the changes we have made are reported in the word file as track changes. English language and grammar have been checked. A detailed response to the reviewers' comments is given.

Your sincerely,

Arianna Azzellino
On the behalf of the coauthors

ANSWERS TO REVIEWER N. 4

Please make introduction easier to understand, this version is still not brief enough.

The introduction has been simplified and shortened according to the reviewer's recommendation.

ANSWERS TO REVIEWER N. 7

Obviously, this paper did much improvement. However, tables are still too many, some tables are not necessary. For example, Table 3 is redundant, important information in table 3 was also shown in table 4. In my opinion, this paper still need a minor revision.

We agree with the comment. We deleted table 3 and table 5, including some comment lines concerning these two tables directly in the manuscript

Groundwater diffuse pollution in Functional Urban Areas: the need to define anthropogenic Diffuse Pollution Background Levels

Azzellino Arianna[a], Colombo Loris [a], Lombi Silvia[a], Marchesi Valeria [b], Piana Alessandra[b], Andrea Merri[b] , Alberti Luca [a, c]

[a] Politecnico di Milano, DICA, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy
[b] ARPA Lombardia, Via Rosellini n° 17, Milano, 20124, Italy.
[c] Scientific responsible of the research project

Abstract (250-300 words)

Groundwater status in highly urbanized areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. This makes very often challenging to determine whether the observed groundwater conditions are the result of localized pollutant sources (PS-Point Sources). In the EU legislative framework it is accepted that, when Natural Background Levels (NBLs) of undesirable elements are proven to be higher than specific Groundwater Quality Standards (GQSs), NBLs should be assumed as GQSs. No procedure is instead defined when anthropogenic Diffuse Pollution Background levels (DBPLs) are higher than GQSs and make unfeasible any remediation strategy. Among the many contaminants affecting groundwater, the chlorinated solvents, tetrachloroethylene (PCE), trichloroethylene (TCE) and trichloromethane (TCM) among the organics~~,~~ and hexavalent chromium, among the inorganics, having been widely used in several industries all over Europe, are very often the most prevalent contaminants in soil and groundwater.

Aim of this paper is to discuss a multivariate statistical approach to address the issue of identification of anthropogenic Diffuse Pollution Background Levels. With such aim, an area of about 1600 km^2, including the Functional Urban Area of Milan, was considered and 10 independent geochemical datasets, provided by local and regional agencies, and covering the period 2003–2014 were merged into a single database after homogenization and multiple quality checks. A total of 618258 chemical analyses from 3477 sampling wells were considered, being all samples collected and analyzed through internally consistent protocols. The analysis enabled to identify five main clusters, having specific hydrogeological characteristics, different temporal profiles and pollutant background concentration levels, which were also found to respond differently to meteo-climatic changes. This study offers a robust knowledge basis for drafting a diffuse pollution management plan of the area.

KEYWORDS max 6:

Natural Background Level; Groundwater quality status; Urban areas; Diffuse Pollution; chlorinated solvents; hexavalent chromium

Formatted: Line spacing: Multiple 1.15 li

Introduction

The management of contaminated sites as undertaken by municipal, regional or national authorities is not solely a matter of whether or not a site is contaminated and if a site should be remediated but also up to what extent the site should be remediated. The changes in the European land use patterns, in particular the urban sprawl phenomena, makes extremely complex the distinction between the point and diffuse sources of contamination (Balderacchi et al., 2014; Stevenazzi et al., 2015). According to Alberti et al. (2018a), the causes of contaminations in urban areas can be classified in three different classes: Point Sources (PS) or contamination hot spots corresponding to areas releasing plumes of high concentrations; Multiple Point Sources (MPS), where the contaminant load comes from a series of point sources that release a low contaminant mass, are clustered within a relatively large area (e.g. industrial district) and consequently are difficult to identify; Non-Point Sources (NPS), where the contaminant load comes from the development of anthropogenic activities over large areas (e.g. fertilizer contamination).

Urban sprawl phenomenon is a source of many groundwater pollutants coming from industrial and domestic sources; among those, the~~ well known~~ chlorinated solvents, tetrachloroethylene or perchloroethylene (PCE), trichloroethylene (TCE) and trichloromethane (TCM), are examples of the multi-source diffuse pollution (Balderacchi et al., 2013). Besides the environmental behavior, these compounds have also similar toxicity: TCM is suspected of being a human carcinogen under unfavorable environmental conditions (Barrlo-Lage et al., 1987; Vogel and McCarty, 1985), ~~excellent indicators of the groundwater pollution by multi-source diffuse type urban pollution because their environmental behavior and toxicity have been thoroughly studied (Balderacchi et al., 2013). Chlorinated solvents have been in fact widely used in several industries all over Europe, and, not surprisingly, are in fact known as the most prevalent organic contaminants found in groundwater (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroo et al., 2003),~~while PCE and TCE may undergo to reductive microbial dechlorination to dichloroethene (DCE) isomers and vinylchloride (VC), which are highly carcinogenic compounds. Chlorinated solvents have been widely used in several industries all over Europe, and, being capable to infiltrate rapidly into the subsurface, causing soil and groundwater pollution (Cortés et al., 2011; Kueper et al., 2003)~~.~~ ~~These compounds have been also found with very high concentrations (from hundreds to thousands of µg L^{-1}) in urban groundwater~~, very often are the most prevalent contaminants in groundwater (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroo et al., 2003). ~~affected by leaching of ponds filled with industrial residues and in groundwater contaminated by petrochemical activities (Pecoraino et al., 2008).~~

There are several thousand of PCE/TCE and TCM impacted sites throughout North America, continental Europe and other industrialized areas of the world (Carter et al., 2012; Cortés et al., 2011; Hunkeler and Aravena, 2000; Kueper et al., 2003; Parker et al., 2004). Many of these sites are affected by releases that took places in the first half of the 20th century (Kueper et al., 2003).

Very high concentrations (from hundreds to thousands of µg L^{-1}) of these compounds have been found in groundwaters affected by leaching of ponds filled with industrial residues and in groundwater contaminated by petrochemical activities (Pecoraino et al., 2008). ~~Among these, TCM is also commonly produced, as byproduct, during the chlorination of water and wastewater, together with three other trihalomethanes (THMs): bromodichloromethane, dibromochloromethane, and bromoform.~~

Besides the industrial origin, TCM may also be generated as byproduct by chlorination of waters and wastewaters. All these compounds share features that favor their persistence and areal diffusion in groundwater. They are typically mobile and recalcitrant (Guilbeault et al., 2005; Rivett and Feenstra, 2005) and they originate, at the source, as immiscible liquids: at many PCE, TCE and TCM spill sites, residual amounts of these compounds persist in a pure liquid phase, commonly

Field Code Changed

Formatted: French (Belgium)

Formatted: French (Belgium)

Formatted: English (United States)

Field Code Changed

Field Code Changed

Formatted: French (Belgium)

Formatted: French (Belgium)

referred to as dense non-aqueous-phase-liquids (DNAPLs), within pore spaces or fractures (Kao and Lei, 2000; Munholland et al., 2016; Rivett and Feenstra, 2005; U.S.EPA, 1992). Groundwater flowing through the DNAPL zones dissolves them, generating plumes that commonly achieve exceptionally large sizes (Mackay and Cherry, 1989; Schwille, 1984, 1988). ~~Moreover, , TCM is also suspected of being a human carcinogen,~~

Solvent DNAPLs may cross low permeability geological layers and~~ under unfavorable environmental conditions (Barrlo-Lage et al., 1987; Vogel and McCarty, 1985), PCE and TCE may undergo to reductive microbial dechlorination to dichloroethene (DCE) isomers and vinylchloride (VC), which are highly carcinogenic compounds. These compounds are also known to cross low permeability geological layers. Solvent DNAPLs can in fact~~ penetrate into or through most types of aquitards, even those with very low bulk hydraulic conductivity, due to naturally occurring preferential pathways (Parker et al., 2004). So, not only phreatic aquifers but also the confined ones are highly vulnerable toward this type of contamination. DNAPLs can closely interact with low-permeability deposits, especially those having a significant organic matter content, where they can be trapped and subsequently released into the aquifers (Chapman et al., 2012) through a back-flow diffusion process~~ so that these~~. These deposits may become a secondary source of pollution, persisting with low concentrations for times estimated up to hundreds of years (Chapman and Parker, 2005; Parker et al., 2004). Besides chlorinated ~~solvent~~solvents, chromium being one of the most abundant element in the earth's crust (Emsley, 2011) is also a very frequent ~~contaminant.~~urban sprawl contaminant and, as Cr(VI), may cause toxic and genotoxic effects on the human health (WHO, 1988).

~~Although Cr(VI) may have a geogenic origin (e.g. Nriagu and Nieboer, 1988; Reinmann and De Caritat, 1998; Izbicki et al., 2008),~~Although Cr(VI) may have a geogenic origin (e.g. Izbicki et al., 2008; Nriagu and Nieboer, 1988; Reinmann and De Caritat, 1998), in most cases, the presence of

Cr(VI) is anthropogenic ~~(Nriagu and Nieboer, 1988; Paine, 2001), and the~~(Izbicki et al., 2008; Nriagu and Nieboer, 1988; Paine, 2001). The level of Cr(VI) species in soil depends on the pH (CrO_4^{2-} at pH 6.5–14; $HCrO_4^{-}$ and $Cr_2O_7^{2-}$ at pH 0.7–6.5), the redox potential, and the presence of natural oxidants (e.g., manganese oxide) or reducing agents (e.g., Fe(II), phosphate, sulfide, and organic matter). ~~The chemical form of chromium in the environment~~Chromium speciation determines its mobility and bioavailability. It can be present as a solid mineral in association with several cations, leading to different chemical species with a large solubility range. The most insoluble compounds are those containing Pb, Ca, and Ba, whereas dichromates are highly soluble in soil-water systems (Unceta et al., 2010). ~~Cr(VI) causes toxic and genotoxic effects on the human health (WHO, 1988).~~

~~The occurrence and the persistence of these pollutants also support the hypothesis that they may be considered as representative of MPS.~~ Remediation of these pollutants can be very difficult~~ to conduct~~. Experience from the past 20 years has in fact demonstrated that sites contaminated by these pollutants are difficult to investigate and challenging to remediate (Kueper et al., 2003; Wanner et al., 2012). In spite of innovative technologies being continuously developed to overcome the technical impracticability of source treatment (e.g. ~~(Kueper et al., 2003)), very frequently it is not possible to locate and remove the residual pollutant concentrations, so remediation (e.g. Pumping&Treat) is often applied as preventing~~Kueper et al., 2003), very frequently it is not possible to locate the sources or to remove the residual concentrations. In these situations, remediation (e.g. Pumping&Treat) is often applied only to prevent further migration of dissolved contamination (Kao and Lei, 2000). Conventional groundwater treatment technologies, such as pump and treat or containment, are in fact able to control contaminant plumes emanating from DNAPL or other contaminants source zones, ~~however they~~

Formatted: English (United Kingdom)

~~involve~~nevertheless involving extended operating periods (decades) and potentially high life cycle costs (Fruchter, 2002; McGuire et al., 2006).

~~Decision-makers are in fact faced with a~~A broad range of different ~~technical~~cleanup approaches ~~for site cleanup~~, including biological, chemical and physical (thermal) technologies~~ that~~, can be implemented either ex situ or in situ, with efficiency and cleanup times that vary substantially as the associated costs and the environmental impact~~ of each method. ~~. In this context, it is extremely important to understand whether the source of pollution is diffuse or localized since completely different actions, with dramatically different associated costs, might be undertaken depending on the different cases.

~~In Italy, the legal framework to manage the local and diffuse contamination of water resources has three different levels: the European, the National and the Regional. From the EU side, the Water Framework Directive 2000/60/EC –(WFD), and the related Groundwater Directive (European Community , 2006), are the pillar for all the policies related to water, to be developed at local level.~~

~~Although clearly stated~~

Legislative voids and needs concerning the groundwater diffuse pollution

The EU Water Framework Directive (2000/60/EC – WFD), and the related Groundwater Directive (European Community, 2006 – GWD) clearly state that measures to recover diffuse pollution of groundwater contribute to the achievement of quality goals for both groundwater and surface water bodies~~ defined under the WFD, no~~; no discussion about recover possibilities and costs or specific mention about the need to identify diffuse pollution background levels is given in these directives.

~~At National level~~In Italy, the Legislative decree (D.lgs. 152/06 which enforce the WFD) defines the anthropogenic ~~Diffuse Pollution~~diffuse pollution as the “chemical, physical and biological alteration of environmental matrixes and contaminations determined by diffuse sources and not linked to a point source", and it designates Regional authorities to recognize and to enact actions when such diffuse contamination is identified. ~~Due to this~~Such legislative demand~~, there is~~ creates the need of scientific-based tools ~~as~~to support ~~for recognizing~~the identification of areas affected by anthropogenic diffuse pollution~~ and identifying~~, defining proper Diffuse Pollution Background Levels (DBPLs, i.e. background diffuse pollution level not attributable to specific point sources).

~~In the WFD and GWD (European Community, 2006) it is accepted that, when the Natural Background Level (NBL) of undesirable elements is proven to be higher than the specific Groundwater Quality Standard (GQS), NBLs should be assumed as GQSs. No clear procedure is yet defined instead on how to manage when Diffuse Pollution Background Levels (DPBLs) are higher than GQSs and make challenging if not unfeasible any remediation strategy, starting from the DPBLs definition itself. In this respect, it should be reminded that, if several and well-established methodological approaches exist on the identification of NBLs (Matschullat et al., 2000) of groundwater contaminants: some parametric (e.g. (Carral et al., 1995; Reimann and Filzmoser, 2000; Wendland et al., 2005) and some non-parametric (Hinsby et al., 2008; Molinari et al., 2014, 2012; Muller et al., 2006; Wendland et al., 2008); consolidated methods are still lacking for the identification of the DPBLs. Aim of this paper is to discuss a multivariate statistical approach to address the issue of identification of Diffuse Pollution Background Levels.~~

When the Natural Background Level (NBL) of undesirable elements is proven to be higher than the specific Groundwater Quality Standard (GQS), WFD and GWD accept that NBLs are assumed as GQSs. No clear procedure is yet defined instead on how to manage when Diffuse Pollution Background Levels (DPBLs) are higher than GQSs and make challenging if not unfeasible any remediation strategy, starting from the DPBLs definition itself. In this respect, if several and well-established methodological approaches exist on the identification of NBLs (Matschullat et al.,

2000) of groundwater contaminants: some parametric (e.g. Carral et al., 1995; Reimann and Filzmoser, 2000; Wendland et al., 2005) and some non-parametric (Hinsby et al., 2008; Molinari et al., 2014, 2012; Muller et al., 2006; Wendland et al., 2008); consolidated methods are still lacking for the identification of the DPBLs. Aim of this paper is to discuss a multivariate statistical approach to address the issue of identification of Diffuse Pollution Background Levels.

Materials and methods

Study Area

The study area (Fig. 1) is located in the Po Plain (Lombardy region, Northern Italy). It includes the Milano Functional Urban Area (FUA after ~~OECD, 2013) (i.e.~~OECD, 2012) (i.e. Milan metropolitan area and 34 surrounding municipalities) and it spans over the territory of other 74 municipalities.

FIGURE 1

The area is about 1600 km^2 wide, and it lies at the center of one of the most urbanized and industrialized areas in Europe, hosting about 5.6% of the whole Italian population (over 3400000 people, (ISTAT, 2017)~~)~~. The climate is continental, and the mean annual precipitation is about 960 mm/year. The valleys of the main rivers (Ticino, Adda, Lambro) are deeply incised due to erosion of post-glacial deposits, leading to several orders of fluvial terraces, with the river bed lower than the regional groundwater level. For this reason, the main rivers play a dominant draining action permitting only small natural oscillations in time (Alberti et al., 2016; Alberti et al., 2018b; Giudici et al., 2007). The Milan – Po plain aquifer system see Fig.2 is composed of three main aquifers made up of Pliocene - Pleistocene continental sediments overlying marine depositional sequences (Carcano and Piccin, 2002; Perego et al., 2014). The unconfined (A) aquifer consists of coarse lithology, mainly gravel with a sandy matrix (gravel sand unit in Fig. 2c). The aquifer, 20–100 m thick, overlays a clayey-silty aquitard. This shows a good continuity in the southern portion of the study area (South of Milan), whereas (Fig. 2) the aquitard becomes discontinuous and then disappear moving northward.

FIGURE 2

The underlying semi-confined (B) aquifer is 50 to 150 meter thick and consists of a sequence of gravel and medium-coarse sand in a sandy matrix with discontinuous confining layers of clay and silt. The base of the semi-confined aquifer consists of clay and silt layers and locally of conglomeratic units. The deep confined aquifer (C) consists of sandy lenses within clay and silt units representing the lower Pleistocene continental-marine transition facies (Colombo et al., 2018; Francani and Beretta, 1995; Pedretti et al., 2013). Concerning the same area, (Bonomi, 2009) provided a 3D detailed definition of subsoil parameters proposing a data process method aimed to increase the value of the stratigraphic well logs~~.~~. De Caro et al., (2017) more recently have provided a geochemical characterization of the area based on mapping of naturally controlled species, providing also NBLs for the area as a detailed mapping of contaminant trends and patterns. According to these authors, CrVI in this area cannot be attributed to geogenic sources due to the absence of ophiolites and serpentinites which are the main sources of natural chromium in other Italian regions (e.g. ~~De Giusti, 2003; Rotiroti et al., 2015).~~De Giusti et al., 2003; Rotiroti et al., 2015).

Formatted: Italian (Italy)

NBLs vs DPBLs determination

The NBL or Baseline level is defined as ‘‘the range of concentration of a given element, isotope or chemical compound in solution, derived entirely from natural, geological, biological or atmospheric sources, under conditions not perturbed by anthropogenic activity” (Edmunds and Shand, 2009). Groundwaters from aquifers that are part of the active water cycle are influenced by human activities. Water changes chemistry from the moment it enters the system through rainfall infiltration, exchanges with surface water bodies or through other sources, until it leaves through runoff, evaporation or withdrawal (Vázquez-Suñé et al., 2005). Shallow aquifers in fact rarely reflect true natural concentration levels, whereas deep aquifers may be more likely free from anthropogenic impacts (Muller et al., 2006). Groundwater status in highly urbanized and farm areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. This makes challenging to determine whether the observed groundwater conditions are the result of a natural chemical status according to the WFD directive or not (Wendland et al., 2005). Most of the methods used to determine NBLs rely on the separation of the natural from anthropogenic components based on some indicator chemical species, such as NO_3, Cl or SO_4 (Hinsby et al., 2008; Matschullat et al., 2000; Muller et al., 2006) and the identification of the NBL as a fixed percentile value (e.g. 90^{th} or 97^{th}) of the observed distribution of the indicator concentrations. Another common characteristic of these methods is the univariate approach since, even when considering a set of parameters, the NBLs are evaluated on the basis of univariate frequency distributions. Moreover, even though many of the studies dealing with NBLs determination have also come up with a mapping of the contamination patterns (De Caro et al., 2017), only very few studies (Busico et al., 2018; Hasenmueller and Criss, 2013; Hwang et al., 2015) have attempted to determine some kind of anthropogenic DPBLs. One of the most innovative aspects of this study is the use of a multivariate approach to determine DPBLs.

Used Data and pre-processing

Ten independent geochemical datasets, provided by local and regional agencies, and covering the period 2003–2014 were merged into a single database after homogenization and multiple quality checks. A total of 618258 chemical analyses from 3477 sampling wells (Fig. 3) were available.

All samples were collected and analyzed through internally consistent protocols (e.g. ASTM Standard methods). Measurements distribution across the sampling wells is quite heterogeneous. About half of the sampling points presents a number of measurements spanning from to 2 to 20. Only 10% of the sampling wells had more than 34 measurements. Figure 3 shows the spatial distribution of sampling wells across the FUA and the data availability at each sampling well.

FIGURE 3

The database included sampling point coordinates, screens depth and a code indicating the aquifer to which each record belongs. Data quality and consistency were checked by considering: (i) data out of the normal range, (ii) presence of outliers (iii) absence of depth or other aquifer information. Errors in the dataset were corrected only in the few cases where obvious data entry errors were identified (i.e. manual correction of wrong measure unit or magnitude of the sampled parameter) or where it was possible to obtain the correct information from the dataset source. The analysis of missing values and non-detects was also performed. The frequency of valid values (i.e. non missing) of all retained constituents was generally larger than 75-65 percent (see Table 1). Concentrations below the Limit of Quantification (LOQ) were replaced with a value of half the LOQ.

Table 1

Moreover, the study area's lithology and precipitation regime were considered. Particularly the lithology layer, extracted from the Regione Lombardia geoportal was used (http://www.geoportale.regione.lombardia.it/), which contains the following information: the genetic classification of surface deposits; the classification of the rocky substratum based on the characteristics of petrographic composition, structure and texture; the main structural features and layers as the depths of the soils reported in classes (e.g. 0-50 cm; 50-100 cm; 100-200 cm). Moreover, precipitation and temperature data available at 15 meteorological stations included in the FUA, were used to estimate the rainfall depurated by evapotranspiration based on the Thornthwaite method (Thornthwaite and Mather, 1955).

Then the net precipitation monthly values have been interpolated through a kriging technique and used for the analysis. Moreover, a grid of 1x1 km^2 was used to associate both lithology and precipitation to spatial units and to clusters. Particularly lithology has been evaluated either in terms of areal fraction of the different lithological types, either as overall similarity of the cluster. As similarity metric the Percentage Similarity Index, (i.e. PSC, after Brock, (1977) see Eq.1) was used:

$$PSC = 100 - 0.5 \sum_{i=1}^{k} |a - b| \qquad \text{Eq.1}$$

as the sum of the percent differences of the lithology i between a pair of grid units.

Statistical Analysis

All the statistical computations were made using the statistical package IBM SPSS Statistics 24.0. Principal Component and Factor Analysis (hereinafter FA, cfr. Afifi et al., (2003)) were performed based on the correlation matrix of the concentration measurements. Particularly, Factor Analysis was obtained through the preliminary Principal Components Analysis (hereinafter PCA) which extracted the eigenvalues and eigenvectors from the covariance matrix of the original variances. Factor analysis was chosen to reduce the contribution of the less significant parameters within each component, by extracting a new set of varifactors through rotating the axes defined by the PCA extraction. The Varimax rotation criterion was used to rotate the PCA axes allowing to maintain the axes orthogonality. The number of factors to be retained was chosen on the basis of the "*eigenvalue higher than 1*" criterion (i.e. all the factors that explained less than the variance of one of the original variables were discarded). That allowed to select few factors able to describe the whole dataset with minimum loss of original information. Moreover, K-means Cluster Analysis (hereinafter CA, cfr. Afifi et al., (2003)) was used to analyze the similarities among the water quality profiles at the different monitoring stations, using the Euclidean Distance as distance metric (see Eq. 2).

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^{p}\left(x_{ik} - y_{jk}\right)^2} \qquad \text{Eq.2}$$

where i and j refer to a couple of stations, and p to the considered parameters.

CA was run based on the FA extracted varifactors. Due to the fact that the k-means procedure is somewhat sensitive to the initial choice of seeds, CA was run twice using the final cluster centroids obtained from the first CA analysis as initial seeds for the second run (cfr. Afifi et al., (2003)). Different CA trials were run to identify the optimal K for the cluster solution. The final choice was made based on cluster interpretability and stability across different CA results.

The Pearson's linear correlation coefficient (r), where adequate sample size was available, and the Spearman's rank order correlation coefficient (rho) were used for the correlation analysis. Kruskal-Wallis test by Rank was instead used to test the difference among cluster medians of lithological class fractions and similarities; Bonferroni's correction was applied for multiple comparisons.

Results

Preliminary analysis of multivariate dataset

In consideration of the fact that PCA/FA had to be applied to the whole dataset using a listwise deletion criterion of missing values (i.e. case units with missing values in one or more variables are discarded from the analysis), the parameters having a percent of missing values higher than 35% were preliminary excluded. Hexavalent chromium, although a priority pollutant, was excluded in reason of the poor availability across the study area and in consideration of its high correlation with total chromium (r: 0.979, $P<0.01$).

The application of the listwise deletion criterion reduced the size of multivariate dataset to about 60% (N: 25800) of the mean size of whole dataset concerning the investigated parameters (N: about 42300).

In consideration of the sample size reduction, the representativeness of the multivariate dataset was investigated. As shown in Table 2, the multivariate dataset, even though characterized by a smaller sample size, resulted to be fully representative of the whole dataset.

TABLE 2

As it can be observed, in fact, the distribution quartiles of the four pollutants of main interest for this study are quite comparable in the two situations; mean and standard deviation, being strongly affected by outliers that are only present in the whole dataset, do not correspond. Outliers in the whole dataset mostly refer to extremely high values measured at some specific pollution hotspot sites. In these sites, very often, only the information about pollutants was available, often lacking data about the other parameters of the multivariate dataset. In consideration of the fact that multivariate dataset had to be used to determine DPBLs, the absence of extreme outliers in the multivariate dataset was considered beneficial for the analysis. Moreover, in order to assess whether the spatial coverage of the multivariate dataset was about the same of the whole dataset considering the single parameters independently, the centroids of the measurement site coordinates, weighted by the corresponding frequency of measurements, obtained from either the multivariate or the whole dataset were compared. As it can be observed in Figure 4, the two centroids are quite close, being the distance between them less than 500 m, so it can be concluded that the multivariate dataset has about the same spatial coverage of the whole dataset.

FIGURE 4

Principal Component and Factor Analysis

PCA applied to the multivariate dataset led to the extraction of five principal components explaining overall the 78% of the total variance (see Table 3). As shown in Table 3, the first two principal components bring most of the information (about 53%).

~~TABLE 3~~

Table 43 shows the comparison between the factor loadings obtained by PCA and the loadings of the FA rotated solution.

TABLE 3

~~TABLE 4~~

Formatted: Indent: Left: 0.63 cm

The two sets of loadings are very similar, although in the rotated solution some parameters have "migrated" from a component to another. PCA solution had in fact a first component constituted by all the main hydrochemical parameters (i.e. conductivity and the main ions) explaining alone 40% of the total variance, a second component strongly linked to PCE and TCE explaining about 13% of the total variance, and the following components, each explaining less than 10% of the total variance. Moreover, in the PCA solution there are some parameters, namely Na, pH, Total Cr and TCM which load on two different components making more difficult the component interpretation.

On the other hand, the Varimax rotated solution (see Table 43) makes the PCA loadings either large or small on single varifactors, facilitating the interpretation. Particularly, in the rotated solution the hydrochemical parameters, which in the PCA solution almost all loaded the first component, are now loading both the first (i.e. explaining 25% of total variance) and the second varifactor (i.e. explaining about 24% of total variance); the pollutants Cr and TCM show also a cleaner distribution across the rotated varifactors, being total chromium and TCM both respectively loaded only on the fourth (i.e. 8.5% of the total variance) and on the fifth (i.e. 8% of the total variance) varifactor.

In summary the extracted varifactors are the following:

Varifactor 1: Accounts 25.1% of the total variance and is loaded by Conductivity, Cl, Na, K, SO_4; this varifactor represents the conductivity component that is most influenced by anthropic activity particularly concerning the salinization effect as described later; the high correlation between SO_4, Cl and Na has been related to the leakage of agricultural and municipal wastes ~~(Sikora et al. 1976).~~(Sikora et al., 1976).

Varifactor 2: Accounts 23.8% of the total variance and is loaded by Conductivity, Ca, Mg, NO_3 and pH; this varifactor represents the conductivity-hardness component of the groundwater. The correlation between Ca and NO_3 can be associated with the use of fertilizers such as $NH_4NO_3{\cdot}CaCO_3$ (22% N and 33% $NCaCO_3$) which is very common in cultivated regions of the study area ~~(Tisdale and Nelson 1975).~~(Tisdale and Nelsson, 1975).

Varifactor 3: Accounts 12.6% of the total variance and is loaded by PCE and TCE. Due to their widespread use and subsequent disposal, and their chemical affinity these two pollutants are commonly found in groundwaters and often correlated.

Varifactor 4: Accounts 8.5% of the total variance and is loaded by Cr. In this area Cr(VI) represents in average the 74% of Total-Cr.

Varifactor 5: Accounts 8% of the total variance and is loaded by TCM which is also a common groundwater contaminant of industrial origin.

K-means Cluster Analysis

K-means cluster analysis was applied to the extracted FA varifactors. K was set to 15 and the analysis was run twice, using the final cluster centroids obtained from the first analysis as initial seeds for the second. Figure 5.a) shows the 15 clusters' composition in terms of number of cases included. It can be observed that most of the clusters contain a very small amount of data (i.e. less than 1% of the total cases) while few clusters, namely cluster 3, 5, 6, 7 and 13, contain more than 98% of the records including most of the background value measurements. Figure 5.b) shows the cluster characteristics in terms of varifactors.

FIGURE 5

It can be observed that clusters representing less than 1% of the total measurements are also the ones characterized by the higher values of the varifactors. Those clusters include in fact measurements that are outliers with respect of one or more of the considered parameters. Figure 6 shows the characteristics of the five main clusters:

- *Cluster 3* is characterized by PCE and TCE values higher than the average (less than a standard deviation), and TCM values much higher than the average (almost two standard deviations);
- *Cluster 5* is characterized by average or below average values concerning all the parameters;
- *Cluster 6* is also characterized by average values concerning all the parameters with the exception of those loading the first varifactor, namely conductivity, Cl, Na, and sulfates that is much higher than the average;

- *Cluster 7* is characterized by average values concerning all the parameters with the exception of Total Chromium which loads the fourth varifactor, which has a mean value much higher than the average.
- *Cluster 13* is characterized by average values concerning all the parameters with the exception of those loading the second varifactor, namely conductivity, Ca, Mg, nitrates and pH, which is higher than the average.

FIGURE 6

Cluster interpretation and trend

Cluster analysis has been applied to the whole dataset without separating boreholes located in different aquifers so the composition of the cluster in terms of aquifers was also investigated. As shown in Figure 7, the five main clusters contain boreholes with screens located in different aquifers since they present a water quality profile similar in both the unconfined (A) and semiconfined (B) aquifer.

FIGURE 7

Figure 8 shows as an example the PCE annual average concentration profiles in the three bigger clusters (i.e. 5, 6 and 13) comparing the unconfined (A) with the semiconfined aquifer (B) profile.

FIGURE 8

It can be observed that they present a comparable temporal trend. Such trends have been also tested through a Spearman's rho rank correlation analysis revealing that the PCE concentration profiles in the unconfined (A) and semiconfined (B) aquifers are correlated at a 0.05 and a 0.10 significance level~~ (see Table 5).~~. The Spearman's correlation coefficients between A and B aquifers in the three bigger clusters (i.e. 5, 6 and 13) respectively are: 0.519 (P: 0.084; n: 12), 0.627 (P: 0.039; n: 11) and 0.591 (P: 0.056; n: 11).

~~TABLE 5~~

Considering the limited size of the analyzed time series, this result strongly supports the hypothesis that the different aquifers in this area present the same temporal variability. Figure 9 shows the cluster temporal profiles in terms of the five varifactor components. It can be observed that the five clusters present different temporal trends:

- *Cluster 3*: presents an increasing annual trend of the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.359 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO_3, pH r: - 0.139 P<0.01). No trend is shown by the PCE/TCE and TCM varifactors whereas the fourth varifactor shows a weak decreasing trend (i.e. Total Cr; r: - 0.118 P<0.01).
- *Cluster 5*: as far as the conductivity components are concerned, cluster 5 presents a weak increasing trend of the first (i.e. Cl, Na, K, SO_4; r: 0.162 P<0.01) and a very weak decreasing trend of the second varifactor (i.e. Ca, Mg, N-NO_3; r: -0.03 P<0.01). Concerning the pollutant components, cluster 5 presents a weak increasing trend of the PCE/TCE varifactor (r: 0.217 P<0.01) and decreasing trends for the Total chromium and the TCM varifactors (respectively: r: -0.191 P<0.01; r: -0.168 P<0.01).
- *Cluster 6*: cluster 6 presents no trend concerning the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.162 P<0.01) and a weak increasing trend of the second varifactor (i.e. Ca, Mg, N-NO_3; r: 0.132 P<0.01). Moreover, cluster 6 presents weak increasing trends of both the PCE/TCE varifactor (r: 0.116 P<0.01) and TCM varifactors (r: 0.052 P<0.01). No significant trend is shown by the Total-chromium varifactor.
- *Cluster 7*: cluster 7 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.420 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO_3, pH r: - 0.225 P<0.01). Concurrently, the trend is very weakly decreasing for the PCE/TCE and Total

chromium varifactors (respectively: r: -0.066 $P<0.01$; r: -0.07 $P<0.01$), while the TCM varifactor does not show any significant trend ($P>0.05$).

- *Cluster 13*: cluster 13 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.225 $P<0.01$) and a decreasing trend of the second (i.e. Ca, Mg, N-NO_3, pH r: - 0.169 $P<0.01$). Moreover, while the trend is increasing for the PCE/TCE varifactor (r: 0.163 $P<0.01$), Total chromium and TCM varifactor trends are significantly decreasing (respectively: r: -0.237 $P<0.01$; r: -0.248 $P<0.01$).

FIGURE 9

Besides the different temporal trends, the clusters have also a well defined different spatial distribution (see Fig.10).

FIGURE 10

The lithological composition was compared among the different clusters. In order to simplify the analysis, only the main lithological classes (i.e. clays, gravels, silts, sands, no soil) were considered and compared among the clusters. The lithological composition of the five clusters (Fig.11) resulted to be significantly different at the KW test concerning the gravel ($P < 0.05$), silt ($P<0.01$) and sand fractions ($P < 0.05$).

FIGURE 11

No significant difference was found in terms of clay ($P > 0.60$) and no soil ($P > 0.30$) fractions among the clusters. Percentage similarity indexes (see Fig. 12) of the cluster lithology was also found statistically different ($P<0.001$).

FIGURE 12

Table ~~6~~4 shows the multiple comparison test results.

TABLE 4

~~TABLE 6~~

Formatted: Indent: Left: 0.63 cm

The correlation with net precipitation (i.e. depurated by evapotranspiration) was also considered. Table ~~7~~5 shows the correlation analysis of the five varifactors with net precipitation, cluster by cluster. It can be observed that most of the varifactors presents a correlation with both the year and net precipitation.

TABLE ~~7~~5

Towards the DPBLs determination

The 5 identified clusters, representing the 98% of the sample, offer a robust support for the definition of DPBLs for the study area. As shown in Figure 13 in fact the five main clusters are very different also in terms of pollutant background values (see Supplementary tables S1, S2, S3) and their own spatial distribution suggest that in the study area coexist different water quality profiles and different pollution background levels.

FIGURE 13

Particularly, clusters 3 and 7 (see Figure 10) have a more localized spatial distribution and show higher DPBLs in all the aquifers. Particularly, cluster 3 is characterized by solvents concentrations of one magnitude higher than the GQS (e.g. PCE: 1.1 ug/l; TCE 1.5 ug/l; TCM: 0.15 ug/l), while cluster 7 has chromium and nitrates levels that respectively are almost double or very close to the corresponding GQS (respectively Cr-IV: 5 ug/l and NO_3: 50 mg/l). On the other hand, clusters 5 and 13 have a more diffuse spatial distribution and tend to have lower background values concerning all the pollutants. Cluster 6 is instead mostly concentrated within the municipality of Milan and it shows also well-defined characteristics: chromium levels that are higher than the ones of cluster 5 and 13 but lower than the GQS, solvents and nitrates levels comparable with cluster 7 levels, being the PCE and TCM levels (TCE levels are higher than the GQS only in the semiconfined aquifer) higher than the corresponding GQS all over the area. Overall, it is relevant to observe, that both

PCE and TCE present a wide logarithmic variability in their concentration values, concerning all the clusters, being respectively 3, 7 and 6 the cluster with the higher levels, which on some spots are many orders of magnitude higher than the GQS. Overall, background pollution levels in the different aquifers follow the same pattern of the cluster, being generally the background levels in the semiconfined (B) and in the undifferentiated aquifer higher that the corresponding levels in the unconfined aquifer (A).

Discussion

A widespread pattern of pollution occurrence in groundwaters has been documented in many urban areas all around Europe (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroo et al., 2003). In time plumes coming from single point-sources (PS), mix with each other and end into a diffuse pattern of pollution linked to unidentifiable Multi-Point Sources, which, in most cases, started decades ago (Cortés et al., 2011).

Remediation of diffuse pollution to GQS levels, which very often are much lower than the background levels, is very difficult if not ~~totally~~ unfeasible, and decision-makers, in many situations concerning Functional Urban Areas, are in fact urged to define new remediation goals, different from the GQS which are very often determined through a conservative risk assessment process. However, such a procedure is generally conducted by each Regional authority on case-by-case basis process while a national and European legislation about the management of diffuse pollution is still lacking. There is no doubt that relying on site-specific risk analysis in a FUAs with large patterns of diffuse contamination, it substantially increases clean-up associated costs compared to a very small improvement of groundwater quality. On the other hand, assessing the type of pollution, whether diffuse or localized, is critical as Public Authorities need to face those contaminations applying completely different actions, with dramatically different associated costs.

If well-established methodological approaches exist for the identification of Natural Background Levels (Matschullat et al., 2000) of groundwater contaminants, no consolidated method either exist for identifying the anthropogenic groundwater diffuse pollution. The methodology proposed in this study is not intrinsically innovative, since multivariate techniques of the kind applied have been also used in several monitoring studies concerning both surface and groundwaters (Busico et al., 2018; Gourdol et al., 2013; Marcelli et al., 2010; Mendizabal et al., 2011; Olsen et al., 2012; Selle et al., 2013; Sheikhy Narany et al., 2014; Yidana et al., 2008; Yu et al., 2014); however, to our knowledge, this is the first time that these multivariate methods are employed to define DPBLs in FUA where MPS are present. The strength of this data driven approach is the possibility to identify hidden patterns in monitoring data (e.g. the hydrosomes as in Mendizabal et al., (2011)) directly from the dataset. The analysis showed clearly that the Milan FUA is characterized by different ~~hydro-chemical~~hydrochemical groundwater subsystems with well differentiated water quality characteristics, which vary in time and space. The five main clusters identified through the analysis were in fact found to have different temporal profiles and background concentration levels. Moreover, the clusters were found to be characterized by different lithological characteristics. The acquired evidence is not enough to support the hypothesis that these clusters are a sort of hydrosomes as the ones described by (Mendizabal et al., 2011) (e.g. *hydrosome* is defined as a coherent, three-dimensional unit of groundwater with a specific origin) but it certainly supports the hypothesis that the FUA should not be seen or managed as an homogeneous groundwater body. Concerning the hydrogeochemical constitutes, in four of the five main clusters, the analysis reveals the presence of an increasing trend of the first varifactor (i.e. Na, K, Cl, SO_4) and a weak but highly significant ($P<0.01$) decreasing trend of the second varifactor (i.e. Ca, Mg, NO_3, pH). While the increasing trend of the first varifactor is a quite widespread pattern that has been found in many urbanised and cultivated territories, which to date concerns more than one-third of the world's irrigated land (Abbas et al., 2013; Heuperman et al., 2002; Houk et al., 2006; Li et al., 2015), the weak decreasing trend of the second varifactor might be due to cation exchange. As shown in Figure 14a, the dissolution of calcite, dolomite and gypsum seems to be the dominant reaction in the system as indicated in the plot of (Ca^{2+}+ Mg^{2+}) vs. (HCO_3^{-} + SO_4^{2-}), which is close to a 1:1 line.

The plot of Ca^{2+} and Mg^{2+}concentrations compared with those of $HCO3^{-}$ mostly in cluster 6 and 13 may indicate the existence of an additional Ca^{2+} and Mg^{2+} source. Such a result seems to be consistent with De Caro et al., 2017 findings which revealed higher calcium concentrations in the northeastern sector of their study area which partially corresponds to ours. It is in fact worthwhile to observe that those high ratios between Ca^{2+} and Mg^{2+} and $HCO3^{-}$ concentrations may not be attributed to HCO_3 depletion because of the existing neutral to slightly alkaline conditions (e.g. groundwater pH ranges from 6.85 to 8.10, with an average of 7.52 and a standard deviation of 0.25) which do not favour the formation of carbonic acid (H_2CO_3, see Spears, (1986)). Moreover, according to Jankowski et al. (1998), if active cation exchange between Na^{+} and Ca^{2++} Mg^{2+} is occurring in the aquifer, the slope of the bivariate plot between Cl^{-} corrected Na^{+} + K^{+} and (Ca^{2+} + Mg^{2+}) corrected HCO_3^{-} + SO_4^{2-} (see Figure 14c) would be −1 (i.e., y = −x). The slope obtained for this plot in the study area (−1.72) seems to indicate that cation exchange is acting as an important process.

FIGURE 14

It is also worthwhile to underline that correlations with precipitation are weak, being insignificant (i.e. cluster 3) or weakly direct concerning the first varifactor, and weakly indirect concerning the second varifactor. The correlation between Na+ and Cl− in the five main clusters is also different, being always significant at the $P<0.01$ level, and higher than 0.6-0.7 in the smaller clusters (i.e. 0.76, 0.63 and 0.73 respectively for clusters 3, 6 and 7 distributed mostly within the Milan municipality) and lower than 0.5 for the bigger clusters (i.e. 0.38, and 0.55 respectively for clusters 5, and 13). In this respect it has been suggested that the poor correlation (r2 < 0.54 that would correspond to an r < 0.74) between Na+ and Cl might indicate a possible evaporative concentration of these ions (Rose, 2002). Na^{+}/Cl^{-} ratio has been used, especially in semiarid or arid regions, to identify the mechanism for acquiring salinity, since the Cl^{-} ion is not affected by soil retention while the Na^{+} ion is retained (Jalali and Khanlari, 2008; Tiwari and Singh, 2014); enrichment of Na+ may also result from reactions taking place in the clay mineral of the surface soil horizons. As shown in Figure 14d the clusters, especially the bigger ones, have a marked fingerprint with respect to the Na+/Cl− ratio vs Cl− concentrations being cluster 5 the one of the highest Na+ enrichment and cluster 6 the one with the lowest.

As far as the pollutant varifactors are concerned, it’s interesting to observe that FA clearly separates their variability from the variability of the hydrogeochemical characteristics, since the pollutant components show also specific trends which are also different within the clusters. Particularly, while the third varifactor (i.e. PCE and TCE) appears to be slightly increasing in time in all the bigger clusters (5, 6 and 13) and to be stable or slightly decreasing in the smaller clusters 3 and 7, the fourth and fifth varifactors (i.e respectively Total-Cr and TCM) show instead different patterns: total Cr is stable (i.e. cluster 6) or a decreasing in all the other clusters while TCM is stable in clusters 3 and 7, weakly increasing in cluster 6 and decreasing in clusters 5 and 13.

Correlations with precipitation for the pollutant varifactors are all very weak although significant suggesting that meteo-climatic variability may also affect the observed patterns. As pointed out earlier, cluster 3 and 7 both have higher average concentrations of the studied contaminants: cluster 3 having PCE and TCM levels respectively 2 times and 5 times higher than the whole sample average, and cluster 7 showing PCE and Total-Cr levels 2 times higher the whole sample average. The wells belonging to these clusters are mostly inside the boundaries of Milan municipality and are positioned downgradient to a couple of historical industrial district areas where many brownfields were remediated in the last 20 years. So, despite the effort done in the past to find the contamination hot spots (i.e. PS) and to remediate soil and groundwater, there are still traces of the past production activities whose impact can still be detected. This is because many small unidentifiable sources (i.e. MPS) are still present, but since they are releasing a low mass of contaminants there is no chance to locate their position and to apply a specific remediation action. Luckily concentrations in these two clusters are showing a decreasing trend in time, probably due to Natural Attenuation Processes that could be assessed, monitored and maybe enhanced to accelerate the achievement of the groundwater quality status. Territorial authorities need support in the process of identifying such trends, and their awareness of the existence of portions of the groundwater body that may have different quality characteristics and trends need to be arisen. For the same reason appropriate DPBLs should be identified for these portions since only through this process it would be possible to properly manage these areas applying adequate planning policies and actions (e.g. investigations and monitoring). We showed that diffuse pollution is present within the FUA with different water quality levels and different distribution: cluster 13 being the dominant water quality profile representing the 35% of the boreholes, followed by cluster 5 representing the 31% and by cluster 6 representing the 21%. Among the contaminants considered in this study only PCE and TCM systematically exceed their corresponding ~~GQS~~GQSs with median and interquartile values respectively of 3, 1÷7 µg/l, and 0.5, 0.5÷2 µg/l, while TCE and Cr-VI present only spots of higher contamination, with median and interquartile values respectively of 1, 0.5÷2.7 µg/l and 3.9, 2.3÷6.6 µg/l. Defining appropriate DPBLs for these contaminants is both a political and a scientific-technical issue and is beyond the scope of this paper, however, we the evidence provided by this study about the FUA not being a whole homogeneous groundwater body strongly support the hypothesis that different DBPLs should be defined within FUA subareas. Moreover, as weak trends were found, revealing a slow temporal evolution of the contamination (i.e. in average less than 0.06 ug/l per year for the considered decade). Reliable decennial background levels should be drawn every 5-10 years with the caveat of periodically revising. Furthermore, being the clusters robust and stable in time, they allow to overcome the problem of spatial and temporal heterogeneity of measurements which particularly affects this area and offer a robust data series to define spatially the DPBLs through geostatistical approaches. Deterministic fate and transport models will also be useful to simulate the most important plumes deriving from a PS contamination in the FUA, and to identify the subareas affected by these plumes that should be excluded from the geostatistical interpolation.

In summary, it can be concluded that the picture offered by the applied multivariate analysis was able to synthetize the main hydrogeochemical processes acting in the study area (i.e. salinization, ion exchanges processes and the main diffuse pollution patterns), allowing to identify different temporal profiles and to lay the basis for the definition of Diffuse Pollution Background Levels.

Conclusions

Groundwater status in highly urbanized and farm areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. In the EU legislative framework there is no indication about how to define Diffuse Pollution Background Levels, which in many situations can be higher than GQSs and make unfeasible any remediation strategy. The methodology applied in this study enabled to identify five main clusters, having specific hydrogeological characteristics and different temporal profiles and pollutant background concentration levels. The evidence provided by this study strongly suggests that the FUA should not be managed as a homogeneous groundwater body, and ~~it~~ outlines the need of defining Diffuse Pollution Background levels at least for PCE and TCM which were systematically found to exceed their corresponding GQSs. The clusters described in this study offer a robust knowledge basis for the drafting a diffuse pollution management plan of the area.

Acknowledgements

This study has been funded within the framework of the AMIIGA Project (Central Europe Interreg (AMIIGA - CE32) and of the PLUME project funded by Regione Lombardia. We are deeply grateful to Elisabetta Preziosi who gave us the right perspective to describe the legislative context of WFD and GWD, and to the many anonymous reviewers who greatly helped us with their valuable comments to significantly improve the original manuscript.

References

Abbas, A., Khan, S., Hussain, N., Hanjra, M.A., Akbar, S., 2013. Characterizing soil salinity in irrigated agriculture using a remote sensing approach. Phys. Chem. Earth. doi:10.1016/j.pce.2010.12.004

Afifi, A., May, S., Clark, V.A., 2003. Cluster analysis, in: Computer-Aided Multivariate Analysis. pp. 428–444.

Alberti, L., Cantone, M., Colombo, L., Lombi, S., Piana, A., 2016. Numerical modeling of regional groundwater flow in the Adda-Ticino Basin: Advances and new results. Rend. Online Soc. Geol. Ital. 41. doi:10.3301/ROL.2016.80

Alberti, L., Colombo, L., Formentin, G., 2018. Null-space Monte Carlo particle tracking to assess groundwater PCE (Tetrachloroethene) diffuse pollution in north-eastern Milan functional urban area. Sci. Total Environ. 621. doi:10.1016/j.scitotenv.2017.11.253

Alberti, L., Colombo, L., La Licata, I., Cantone, M., 2018. Model calibration using the automatic parameter estimation procedure (PEST) of the North-eastern zone of the Milan Functional Urban Area (Italy). Acque Sotter. - Ital. J. Groundw. 7, 27–38. doi:10.7343/as-2018-336

Balderacchi, M., Benoit, P., Cambier, P., Eklo, O.M., Gargini, A., Gemitzi, A., Gurel, M., Kløve, B., Nakic, Z., Predaa, E., Ruzicic, S., Wachniew, P., Trevisan, M., 2013. Groundwater pollution and quality monitoring approaches at the European level. Crit. Rev. Environ. Sci. Technol.

doi:10.1080/10643389.2011.604259

Balderacchi, M., Filippini, M., Gemitzi, A., ~~Kløve~~KlÃ¶ve, B., Petitta, M., Trevisan, M., Wachniew, P., Witczak, S., Gargini, A., 2014. Does groundwater protection in Europe require new EU-wide environmental quality standards? Front. Chem. doi:10.3389/fchem.2014.00032

Barrlo-Lage, G., Parsons, F.Z., Nassar, R.S., 1987. Kinetics of the Depletion of Trichloroethene. Environ. Sci. Technol. 21, 366–370. doi:10.1021/es00158a005

Bonomi, T., 2009. Database development and 3D modeling of textural variations in heterogeneous, unconsolidated aquifer media: Application to the Milan plain. Comput. Geosci. doi:10.1016/j.cageo.2007.09.006

Brock, D.A., 1977. Comparison of community similarity indexes. Water Pollut. Control Fed. 49, 2488–2494.

Busico, G., Cuoco, E., Kazakis, N., Colombani, N., Mastrocicco, M., Tedesco, D., Voudouris, K., 2018. Multivariate statistical analysis to characterize/discriminate between anthropogenic and geogenic trace elements occurrence in the Campania Plain, Southern Italy. Environ. Pollut. doi:10.1016/j.envpol.2017.11.053

Carcano, C., Piccin, A., 2002. Padans aquifer geology of the Lombardy Plain. Firenze.

Carral, E., Puente, X., Villares, R., Carballeira, A., 1995. Background heavy metal levels in estuarine sediments and organisms in Galicia (northwest Spain) as determined by modal analysis. Sci. Total Environ. doi:10.1016/0048-9697(95)04788-3

Carter, J.M., Moran, M.J., Zogorski, J.S., Price, C. V., 2012. Factors associated with sources, transport, and fate of chloroform and three other trihalomethanes in untreated groundwater used for drinking water. Environ. Sci. Technol. doi:10.1021/es301839p

Chapman, S.W., Parker, B.L., 2005. Plume persistence due to aquitard back diffusion following dense nonaqueous phase liquid source removal or isolation. Water Resour. Res. doi:10.1029/2005WR004224

Chapman, S.W., Parker, B.L., Sale, T.C., Doner, L.A., 2012. Testing high resolution numerical models for analysis of contaminant storage and release from low permeability zones. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2012.04.006

Colombo, L., Gattinoni, P., Scesi, L., 2018. Stochastic modelling of groundwater flow for hazard assessment along the underground infrastructures in Milan (northern Italy). Tunn. Undergr. Sp. Technol. 79. doi:10.1016/j.tust.2018.05.007

Cortés, A., Puigserver, D., Carmona, J.M., Viladevall, M., 2011. Biological remediation approach involving soils and groundwaters polluted with chlorinated solvents in a Mediterranean context. Recent Adv. Pharm. Sci. 661, 223–246.

De Caro, M., Crosta, G.B., Frattini, P., 2017. Hydrogeochemical characterization and Natural Background Levels in urbanized areas: Milan Metropolitan area (Northern Italy). J. Hydrol. doi:10.1016/j.jhydrol.2017.02.025

De Giusti, F., ~~Dal ~~Piaz, ~~G.V~~D., Massironi, M., Schiavo, A., 2003. Carta geotettonica della Valle

~~d'Aosta (Geotectonic map of~~d ' Aosta~~ Valley)~~. Mem. Sci. Geol~~. 55,~~ 129–149.

Edmunds, W.M., Shand, P., 2009. Natural Groundwater Quality, Natural Groundwater Quality. doi:10.1002/9781444300345

Emsley, J., 2011. Nature's Building Blocks: An A-Z Guide to the Elements, Oxford University Press. doi:978-0-19-960563-7

European Community, 2006. Groundwater directive 2006/118/CE. Directive of the European Parliament and of the Council on the Protection of Groundwater Against Pollution and Deterioration~~, OJ L372, 27/12/2006 pp. 19–31~~. Off. J. Eur. Union 19, 19–31. doi:http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006L0118

~~Emsley, J., 2011. Nature's Building Blocks: An A-Z Guide to the Elements, Oxford University Press. doi:978-0-19-960563-7~~

Francani, V., Beretta, G. Pietro, 1995. Preliminary survey about sustainable use of deeper groundwater in Milano province. Milan.

Fruchter, J., 2002. In-Situ Treatment of Chromium-Contaminated Groundwater. Environ. Sci. Technol. doi:10.1021/es022466i

Giudici, M., Ponzini, G., Romano, E., Vassena, C., 2007. Some Lessons from Modeling Ground Water Flow in the Metropolitan Area of Milano at Different Scales. Mem. Descr. della Cart. Geol. d'Italia LXXVI, 207–218.

Gourdol, L., Hissler, C., Hoffmann, L., Pfister, L., 2013. On the potential for the Partial Triadic Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry. Appl. Geochemistry. doi:10.1016/j.apgeochem.2013.10.002

Guilbeault, M.A., Parker, B.L., Cherry, J.A., 2005. Mass and flux distributions from DNAPL zones in sandy aquifers. Ground Water. doi:10.1111/j.1745-6584.2005.tb02287.x

Hasenmueller, E.A., Criss, R.E., 2013. Multiple sources of boron in urban surface waters and groundwaters. Sci. Total Environ. doi:10.1016/j.scitotenv.2013.01.001

Heuperman, A.F., Kapoor, A.S., Denecke, H.W., 2002. Biodrainage - Principles, Experiences and Applications. Rome.

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

Houk, E., Frasier, M., Schuck, E., 2006. The agricultural impacts of irrigation induced waterlogging and soil salinity in the Arkansas Basin. Agric. Water Manag. 85, 175–183. doi:10.1016/j.agwat.2006.04.007

Hunkeler, D., Aravena, R., 2000. Determination of compound-specific carbon isotope ratios of chlorinated methanes, ethanes, and ethenes in aqueous samples. Environ. Sci. Technol. doi:10.1021/es991178s

Formatted: Justified

Hwang, H.-H., Panno, S. V., Hackley, K.C., 2015. Sources and Changes in Groundwater Quality with Increasing Urbanization, Northeastern Illinois. Environ. Eng. Geosci. doi:10.2113/gseegeosci.21.2.75

ISTAT, 2017. Annual survey on movement and calculation of the resident population. Rome.

Izbicki, J.A., Ball, J.W., Bullen, T.D., Sutley, S.J., 2008. Chromium, chromium isotopes and selected trace elements, western Mojave Desert, USA. Appl. Geochemistry. doi:10.1016/j.apgeochem.2007.11.015

Jalali, M., Khanlari, Z. V., 2008. Major ion chemistry of groundwaters in the Damagh area, Hamadan, western Iran. Environ. Geol. doi:10.1007/s00254-007-0795-6

Kao, C.M., Lei, S.E., 2000. Using a peat biobarrier to remediate PCE/TCE contaminated aquifers. Water Res. doi:10.1016/S0043-1354(99)00213-4

Kueper, B.H., Wealthall, G.P., Smith, J.W.N., Leharne, S. a, Lerner, D.N., 2003. An illustrated handbook of DNAPL transport and fate in the subsurface, R&D Publication 133.

Li, W., Wang, M.-Y., Liu, L.-Y., Yan, Y., 2015. Assessment of Long-Term Evolution of Groundwater Hydrochemical Characteristics Using Multiple Approaches: A Case Study in Cangzhou, Northern China. Water 7, 1109–1128. doi:10.3390/w7031109

Mackay, D.M., Cherry, J.A., 1989. Groundwater contamination: Pump-and-treat remediation. Environ. Sci. Technol. doi:10.1021/es00064a001

Marcelli, R., Bartolucci, G., Papaioannu, G., De Angelis, G., Lucibello, A., Proietti, E., Margesin, B., Giacomozzi, F., Deborgies, F., 2010. Reliability of RF MEMS switches due to charging effects and their circuital modelling, in: Microsystem Technologies. pp. 1111–1118. doi:10.1007/s00542-009-1006-z

Matschullat, J., Ottenstein, R., Reimann, C., 2000. Geochemical background - Can we calculate it? Environ. Geol. doi:10.1007/s002549900084

McGuire, T.M., McDade, J.M., Newell, C.J., 2006. Performance of DNAPL source depletion technologies at 59 chlorinated solvent-impacted sites. Gr. Water Monit. Remediat. doi:10.1111/j.1745-6592.2006.00054.x

~~Meffe, R., de Bustamante, I., 2014. Emerging organic contaminants in surface water and groundwater: A first overview of the situation in Italy. Sci. Total Environ. doi:10.1016/j.scitotenv.2014.02.053~~

Mendizabal, I., Stuyfzand, P.J., Wiersma, A.P., 2011. Hydrochemical system analysis of public supply well fields, to reveal water-quality patterns and define groundwater bodies: The Netherlands. Hydrogeol. J. doi:10.1007/s10040-010-0614-0

Molinari, A., Chidichimo, F., Straface, S., Guadagnini, A., 2014. Assessment of natural background levels in potentially contaminated coastal aquifers. Sci. Total Environ. doi:10.1016/j.scitotenv.2013.12.125

Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2012. Natural background levels and threshold values of chemical species in three large-scale groundwater bodies in Northern

Italy. Sci. Total Environ. doi:10.1016/j.scitotenv.2012.03.015

Muller, D., Blum, A., Hart, A., Hookey, J., Kunkel, R., Scheidleder, A., 2006. Final Report - BRIDGE (Background criteria for the identification of groundwater thresholds) 13030.

Munholland, J.L., Mumford, K.G., Kueper, B.H., 2016. Factors affecting gas migration and contaminant redistribution in heterogeneous porous media subject to electrical resistance heating. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2015.10.011

Nriagu, J.O., Nieboer, E., 1988. Chromium in the Natural and Human Environments. John Wiley & Sons.

OECD, 2012. Redefining "Urban": A New Way to Measure Metropolitan Areas. OECD Publ. 1–9. doi:10.1787/9789264174108-en

Olsen, R.L., Chappell, R.W., Loftis, J.C., 2012. Water quality sample collection, data treatment and results presentation for principal components analysis e literature review and Illinois River watershed case study. Water Res.

Paine, A.J., 2001. Mechanisms of chromium toxicity, carcinogenicity and allergenicity: Review of the literature from 1985 to 2000. Hum. Exp. Toxicol. doi:10.1191/096032701682693062

Parker, B.L., Cherry, J.A., Chapman, S.W., 2004. Field study of TCE diffusion profiles below DNAPL to assess aquitard integrity. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2004.02.011

Pecoraino, G., Scalici, L., Avellone, G., Ceraulo, L., Favara, R., Candela, E.G., Provenzano, M.C., Scaletta, C., 2008. Distribution of volatile organic compounds in Sicilian groundwaters analysed by head space-solid phase micro extraction coupled with gas chromatography mass spectrometry (SPME/GC/MS). Water Res. doi:10.1016/j.watres.2008.07.022

Pedretti, D., Masetti, M., Beretta, G.P., Vitiello, M., 2013. A revised conceptual model to reproduce the distribution of chlorinated solvents in the Rho Aquifer (Italy). Groundw. Monit. Remediat. doi:10.1111/gwmr.12017

Perego, R., Bonomi, T., Fumagalli, M.L., Benastini, V., Aghib, F., Rotiroti, M., Cavallin, A., 2014. 3D reconstruction of the multi-layer aquifer in a Po Plain area. Rend. Online Soc. Geol. Ital. 30, 41–44. doi:10.3301/ROL.2014.09

Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: Death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environ. Geol. doi:10.1007/s002549900081

Reinmann, C., De Caritat, P., 1998. Chemical Elements in the Environment. sPRINGER-vERLAG, Berlin.

Rivett, M.O., Feenstra, S., 2005. Dissolution of an emplaced source of DNAPL in a natural aquifer setting. Environ. Sci. Technol. doi:10.1021/es040016f

Rose, S., 2002. Comparative major ion geochemistry of Piedmont streams in the Atlanta, Georgia region: Possible effects of urbanization. Environ. Geol. doi:10.1007/s00254-002-0545-8

Rotiroti, M., ~~DiMauro~~Di Mauro, B., Bonomi, T., Fumagalli, M.L., ~~Bonomi, T.~~ 2015. COMPSEC, a new tool to derive natural background levels by the component separation ~~approach:~~

Formatted: Justified

~~application~~approach : Application in two different hydrogeological contexts in northern Italy~~.~~ ~~Journal of Geochemical Exploration 158, 44–54~~ COMPSEC , a new tool to derive natural background levels by the component separation. doi:10.1016/j.gexplo.2015.06.017

Schwille, F. (1984), 1984. Migration of Organic Fluids Immiscible with Water in the unsaturated zone 2–3.

Schwille, F. (1988), 1988. Dense Chlorinated Solvents in Porous and Fractured Media: Model Experiments, translated from the German by J. F. Pankow.

Selle, B., Schwientek, M., Lischeid, G., 2013. Understanding processes governing water quality in catchments using principal component scores. J. Hydrol. doi:10.1016/j.jhydrol.2013.01.030

Sheikhy Narany, T., Ramli, M.F., Aris, A.Z., Sulaiman, W.N.A., Fakharian, K., 2014. Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol-Babol Plain, Iran. Environ. Monit. Assess. doi:10.1007/s10661-014-3820-8

Sikora~~ LG,~~, L., Bent~~ MG,~~, M., Corey~~ RB,~~, R., Keeney~~ DR (~~, D., 1976~~)~~. Septic nitrogen and phosphorus removal test system. Groundwater 14~~/5:~~, 309–314.

Spears, D.A., 1986. Mineralogical control of the chemical evolution of groundwter. Solute Process. 512.

Stevenazzi, S., Masetti, M., Nghiem, S. V., Sorichetta, A., 2015. Groundwater vulnerability maps derived from a time-dependent method using satellite scatterometer data. Hydrogeol. J. 23, 631–647. doi:10.1007/s10040-015-1236-3

Stroo, H.F., Unger, M., Ward, C.H., Kavanaugh, M.C., Vogel, C., Leeson, A., Marqusee, J.A., Smith, B.P., 2003. Peer Reviewed: Remediating Chlorinated Solvent Source Zones. Environ. Sci. Technol. 37, 224A–230A. doi:10.1021/es032488k

Thornthwaite, C.W., Mather, J.R., 1955. The water balance. Publ. Climatol. 8, 1–104.

~~Tisdal~~Tisdale, S~~,~~., Nelsson, W~~ (~~., 1975~~)~~. Soil fertility and fertilizers. Macmillan Publishing Co Inc, New ~~York, p 240~~Jork.

Tiwari, A.K., Singh, A.K., 2014. Hydrogeochemical investigation and groundwater quality assessment of Pratapgarh district, Uttar Pradesh. J. Geol. Soc. India. doi:10.1007/s12594-014-0045-y

U.S.EPA, 1992. Dense Non Aqueous Phase Liquids - A Workshop Summary. Dallas.

Unceta, N., Séby, F., Malherbe, J., Donard, O.F.X., 2010. Chromium speciation in solid matrices and regulation: A review. Anal. Bioanal. Chem. doi:10.1007/s00216-009-3417-1

Vázquez-Suñé, E., Sánchez-Vila, X., Carrera, J., 2005. Introductory review of specific factors influencing urban groundwater, an emerging branch of hydrogeology, with reference to Barcelona, Spain. Hydrogeol. J. doi:10.1007/s10040-004-0360-2

Vogel, T.M., McCarty, P.L., 1985. Biotransformation of tetrachloroethylene to trichloroethylene, dichloroethylene, vinyl chloride, and carbon dioxide under methanogenic conditions. Appl. Environ. Microbiol.

Wanner, C., Zink, S., Eggenberger, U., Mäder, U., 2012. Assessing the Cr(VI) reduction efficiency of a permeable reactive barrier using Cr isotope measurements and 2D reactive transport modeling. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2012.01.007

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

Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H.J., Wolter, R., 2005. A procedure to define natural groundwater conditions of groundwater bodies in Germany. Water Sci. Technol.

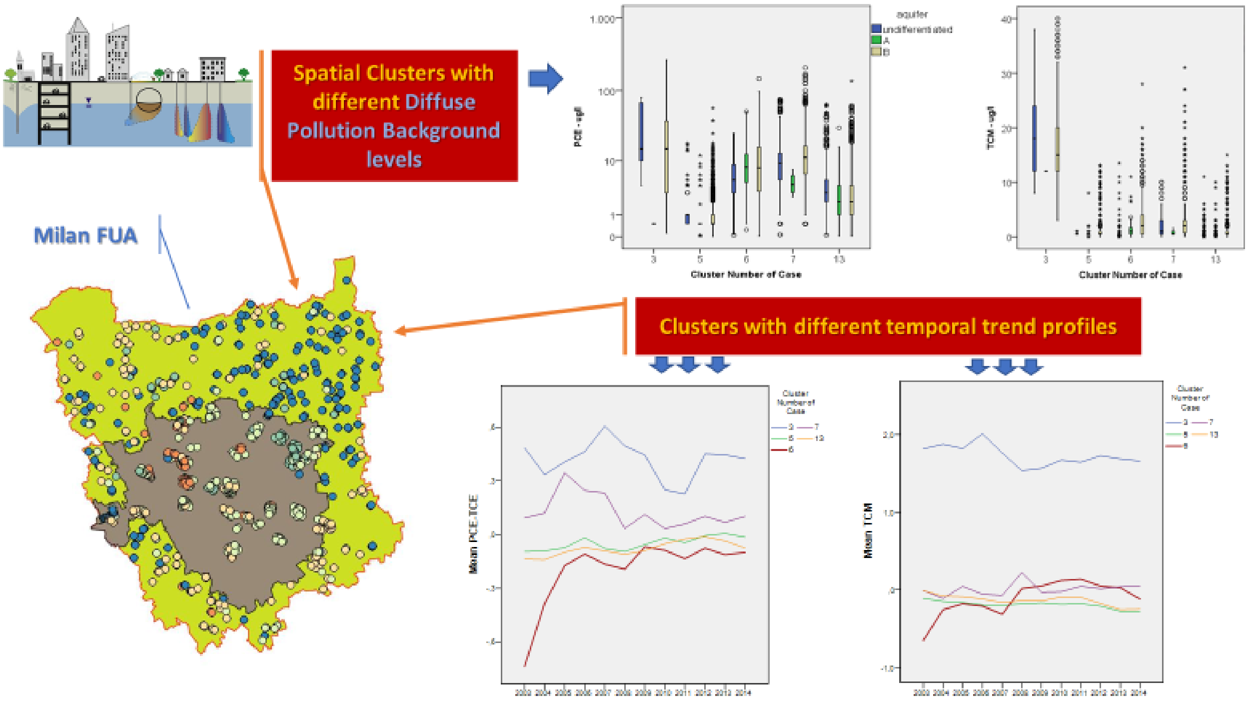
WHO, 1988. Environmental Health Criteria. Geneve.

Yidana, S.M., Ophori, D., Banoeng-Yakubo, B., 2008. A multivariate statistical analysis of surface water chemistry data-The Ankobra Basin, Ghana. J. Environ. Manage. 86, 80–87. doi:10.1016/j.jenvman.2006.11.023

Yu, Y., Song, X., Zhang, Y., Zheng, F., Liang, J., Liu, L., 2014. Identifying spatio-temporal variation and controlling factors of chemistry in groundwater and river water recharged by reclaimed water at Huai River, North China. Stoch. Environ. Res. Risk Assess. doi:10.1007/s00477-013-0803-1

***Graphical Abstract**

Graphical abstract



***Highlights (for review : 3 to 5 bullet points (maximum 85 characters including spaces per bullet point)**

Highlights

- Knowledge of Diffuse Pollution is needed for the management groundwater remediation
- Groundwater chemistry after decades of urbanization is investigated in the Milan area.
- Temporal and spatial patterns of groundwater diffuse pollution were documented
- Diffuse pollution background levels may be determined based on the observed patterns

Groundwater diffuse pollution in Functional Urban Areas: the need to define anthropogenic Diffuse Pollution Background Levels

Azzellino Arianna[a], Colombo Loris [a], Lombi Silvia[a], Marchesi Valeria [b], Piana Alessandra[b], Andrea Merri[b] , Alberti Luca [a, c]

[a] Politecnico di Milano, DICA, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy
[b] ARPA Lombardia, Via Rosellini n° 17, Milano, 20124, Italy.
[c] Scientific responsible of the research project

Abstract (250-300 words)

Groundwater status in highly urbanized areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. This makes very often challenging to determine whether the observed groundwater conditions are the result of localized pollutant sources (PS-Point Sources). In the EU legislative framework it is accepted that, when Natural Background Levels (NBLs) of undesirable elements are proven to be higher than specific Groundwater Quality Standards (GQSs), NBLs should be assumed as GQSs. No procedure is instead defined when anthropogenic Diffuse Pollution Background levels (DBPLs) are higher than GQSs and make unfeasible any remediation strategy. Among the many contaminants affecting groundwater, the chlorinated solvents, tetrachloroethylene (PCE), trichloroethylene (TCE) and trichloromethane (TCM) among the organics and hexavalent chromium, among the inorganics, having been widely used in several industries all over Europe, are very often the most prevalent contaminants in soil and groundwater. Aim of this paper is to discuss a multivariate statistical approach to address the issue of identification of anthropogenic Diffuse Pollution Background Levels. With such aim, an area of about 1600 km^2, including the Functional Urban Area of Milan, was considered and 10 independent geochemical datasets, provided by local and regional agencies, and covering the period 2003–2014 were merged into a single database after homogenization and multiple quality checks. A total of 618258 chemical analyses from 3477 sampling wells were considered, being all samples collected and analyzed through internally consistent protocols. The analysis enabled to identify five main clusters, having specific hydrogeological characteristics, different temporal profiles and pollutant background concentration levels, which were also found to respond differently to meteo-climatic changes. This study offers a robust knowledge basis for drafting a diffuse pollution management plan of the area.

KEYWORDS max 6:

Natural Background Level; Groundwater quality status; Urban areas; Diffuse Pollution; chlorinated solvents; hexavalent chromium

Introduction

The management of contaminated sites as undertaken by municipal, regional or national authorities is not solely a matter of whether or not a site is contaminated and if a site should be remediated but also up to what extent the site should be remediated. The changes in the European land use patterns, in particular the urban sprawl phenomena, makes extremely complex the distinction between the point and diffuse sources of contamination (Balderacchi et al., 2014; Stevenazzi et al., 2015). According to Alberti et al. (2018a), the causes of contaminations in urban areas can be classified in three different classes: Point Sources (PS) or contamination hot spots corresponding to areas releasing plumes of high concentrations; Multiple Point Sources (MPS), where the contaminant load comes from a series of point sources that release a low contaminant mass, are clustered within a relatively large area (e.g. industrial district) and consequently are difficult to identify; Non-Point Sources (NPS), where the contaminant load comes from the development of anthropogenic activities over large areas (e.g. fertilizer contamination).

Urban sprawl phenomenon is a source of many groundwater pollutants coming from industrial and domestic sources; among those, the chlorinated solvents, tetrachloroethylene or perchloroethylene (PCE), trichloroethylene (TCE) and trichloromethane (TCM), are examples of the multi-source diffuse pollution (Balderacchi et al., 2013). Besides the environmental behavior, these compounds have also similar toxicity: TCM is suspected of being a human carcinogen under unfavorable environmental conditions (Barrlo-Lage et al., 1987; Vogel and McCarty, 1985), while PCE and TCE may undergo to reductive microbial dechlorination to dichloroethene (DCE) isomers and vinylchloride (VC), which are highly carcinogenic compounds. Chlorinated solvents have been widely used in several industries all over Europe, and, being capable to infiltrate rapidly into the subsurface, causing soil and groundwater pollution (Cortés et al., 2011; Kueper et al., 2003), very often are the most prevalent contaminants in groundwater (Kao and Lei, 2000; Rivett and

Feenstra, 2005; Stroo et al., 2003). There are several thousand of PCE/TCE and TCM impacted sites throughout North America, continental Europe and other industrialized areas of the world (Carter et al., 2012; Cortés et al., 2011; Hunkeler and Aravena, 2000; Kueper et al., 2003; Parker et al., 2004). Many of these sites are affected by releases that took places in the first half of the 20^{th} century (Kueper et al., 2003). Very high concentrations (from hundreds to thousands of $\mu g\ L^{-1}$) of these compounds have been found in groundwaters affected by leaching of ponds filled with industrial residues and in groundwater contaminated by petrochemical activities (Pecoraino et al., 2008). Besides the industrial origin, TCM may also be generated as byproduct by chlorination of waters and wastewaters. All these compounds share features that favor their persistence and areal diffusion in groundwater. They are typically mobile and recalcitrant (Guilbeault et al., 2005; Rivett and Feenstra, 2005) and they originate, at the source, as immiscible liquids: at many PCE, TCE and TCM spill sites, residual amounts of these compounds persist in a pure liquid phase, commonly referred to as dense non-aqueous-phase-liquids (DNAPLs), within pore spaces or fractures (Kao and Lei, 2000; Munholland et al., 2016; Rivett and Feenstra, 2005; U.S.EPA, 1992). Groundwater flowing through the DNAPL zones dissolves them, generating plumes that commonly achieve exceptionally large sizes (Mackay and Cherry, 1989; Schwille, 1984, 1988).

Solvent DNAPLs may cross low permeability geological layers and penetrate into or through most types of aquitards, even those with very low bulk hydraulic conductivity, due to naturally occurring preferential pathways (Parker et al., 2004). So, not only phreatic aquifers but also the confined ones are highly vulnerable toward this type of contamination. DNAPLs can closely interact with low-permeability deposits, especially those having a significant organic matter content, where they can be trapped and subsequently released into the aquifers (Chapman et al., 2012) through a back-flow diffusion process. These deposits may become a secondary source of pollution, persisting with low concentrations for times estimated up to hundreds of years (Chapman and

Parker, 2005; Parker et al., 2004). Besides chlorinated solvents, chromium being one of the most abundant element in the earth's crust (Emsley, 2011) is also a very frequent urban sprawl contaminant and, as Cr(VI), may cause toxic and genotoxic effects on the human health (WHO, 1988).

Although Cr(VI) may have a geogenic origin (e.g. Izbicki et al., 2008; Nriagu and Nieboer, 1988; Reinmann and De Caritat, 1998), in most cases, the presence of Cr(VI) is anthropogenic (Izbicki et al., 2008; Nriagu and Nieboer, 1988; Paine, 2001). The level of Cr(VI) species in soil depends on the pH (CrO_4^{2-} at pH 6.5–14; $HCrO_4^-$ and $Cr_2O_7^{2-}$ at pH 0.7–6.5), the redox potential, and the presence of natural oxidants (e.g., manganese oxide) or reducing agents (e.g., Fe(II), phosphate, sulfide, and organic matter). Chromium speciation determines its mobility and bioavailability. It can be present as a solid mineral in association with several cations, leading to different chemical species with a large solubility range. The most insoluble compounds are those containing Pb, Ca, and Ba, whereas dichromates are highly soluble in soil-water systems (Unceta et al., 2010).

Remediation of these pollutants can be very difficult. Experience from the past 20 years has in fact demonstrated that sites contaminated by these pollutants are difficult to investigate and challenging to remediate (Kueper et al., 2003; Wanner et al., 2012). In spite of innovative technologies being continuously developed to overcome the technical impracticability of source treatment (e.g. Kueper et al., 2003), very frequently it is not possible to locate the sources or to remove the residual concentrations. In these situations, remediation (e.g. Pumping&Treat) is often applied only to prevent further migration of dissolved contamination (Kao and Lei, 2000). Conventional groundwater treatment technologies, such as pump and treat or containment, are in fact able to control contaminant plumes emanating from DNAPL or other contaminants source zones, nevertheless involving extended operating periods (decades) and potentially high life cycle costs (Fruchter, 2002; McGuire et al., 2006).

A broad range of different cleanup approaches, including biological, chemical and physical (thermal) technologies, can be implemented either ex situ or in situ, with efficiency and cleanup times that vary substantially as the associated costs and the environmental impact. In this context, it is extremely important to understand whether the source of pollution is diffuse or localized since completely different actions, with dramatically different associated costs, might be undertaken depending on the different cases.

Legislative voids and needs concerning the groundwater diffuse pollution

The EU Water Framework Directive (2000/60/EC – WFD), and the related Groundwater Directive (European Community, 2006 – GWD) clearly state that measures to recover diffuse pollution of groundwater contribute to the achievement of quality goals for both groundwater and surface water bodies; no discussion about recover possibilities and costs or specific mention about the need to identify diffuse pollution background levels is given in these directives.

In Italy, the Legislative decree (D.lgs. 152/06 which enforce the WFD) defines the anthropogenic diffuse pollution as the "chemical, physical and biological alteration of environmental matrixes and contaminations determined by diffuse sources and not linked to a point source", and it designates Regional authorities to recognize and to enact actions when such diffuse contamination is identified. Such legislative demand creates the need of scientific-based tools to support the identification of areas affected by anthropogenic diffuse pollution, defining proper Diffuse Pollution Background Levels (DBPLs, i.e. background diffuse pollution level not attributable to specific point sources).

When the Natural Background Level (NBL) of undesirable elements is proven to be higher than the specific Groundwater Quality Standard (GQS), WFD and GWD accept that NBLs are assumed as GQSs. No clear procedure is yet defined instead on how to manage when Diffuse Pollution

Background Levels (DPBLs) are higher than GQSs and make challenging if not unfeasible any remediation strategy, starting from the DPBLs definition itself. In this respect, if several and well-established methodological approaches exist on the identification of NBLs (Matschullat et al., 2000) of groundwater contaminants: some parametric (e.g. Carral et al., 1995; Reimann and Filzmoser, 2000; Wendland et al., 2005) and some non-parametric (Hinsby et al., 2008; Molinari et al., 2014, 2012; Muller et al., 2006; Wendland et al., 2008); consolidated methods are still lacking for the identification of the DPBLs. Aim of this paper is to discuss a multivariate statistical approach to address the issue of identification of Diffuse Pollution Background Levels.

Materials and methods

Study Area

The study area (Fig. 1) is located in the Po Plain (Lombardy region, Northern Italy). It includes the Milano Functional Urban Area (FUA after OECD, 2012) (i.e. Milan metropolitan area and 34 surrounding municipalities) and it spans over the territory of other 74 municipalities.

FIGURE 1

The area is about 1600 km^2 wide, and it lies at the center of one of the most urbanized and industrialized areas in Europe, hosting about 5.6% of the whole Italian population (over 3400000 people, (ISTAT, 2017). The climate is continental, and the mean annual precipitation is about 960 mm/year. The valleys of the main rivers (Ticino, Adda, Lambro) are deeply incised due to erosion of post-glacial deposits, leading to several orders of fluvial terraces, with the river bed lower than the regional groundwater level. For this reason, the main rivers play a dominant draining action permitting only small natural oscillations in time (Alberti et al., 2016; Alberti et al., 2018b; Giudici et al., 2007). The Milan – Po plain aquifer system see Fig.2 is composed of three main aquifers made up of Pliocene - Pleistocene continental sediments overlying marine depositional sequences (Carcano and Piccin, 2002; Perego et al., 2014). The unconfined (A) aquifer consists of coarse lithology, mainly gravel with a sandy matrix (gravel sand unit in Fig. 2c). The aquifer, 20–100 m thick, overlays a clayey-silty aquitard. This shows a good continuity in the southern portion of the study area (South of Milan), whereas (Fig. 2) the aquitard becomes discontinuous and then disappear moving northward.

FIGURE 2

The underlying semi-confined (B) aquifer is 50 to 150 meter thick and consists of a sequence of gravel and medium-coarse sand in a sandy matrix with discontinuous confining layers of clay and silt. The base of the semi-confined aquifer consists of clay and silt layers and locally of conglomeratic units. The deep confined aquifer (C) consists of sandy lenses within clay and silt units representing the lower Pleistocene continental-marine transition facies (Colombo et al., 2018; Francani and Beretta, 1995; Pedretti et al., 2013). Concerning the same area, (Bonomi, 2009) provided a 3D detailed definition of subsoil parameters proposing a data process method aimed to increase the value of the stratigraphic well logs. De Caro et al., (2017) more recently have provided a geochemical characterization of the area based on mapping of naturally controlled species, providing also NBLs for the area as a detailed mapping of contaminant trends and patterns. According to these authors, CrVI in this area cannot be attributed to geogenic sources due to the absence of ophiolites and serpentinites which are the main sources of natural chromium in other Italian regions (e.g. De Giusti et al., 2003; Rotiroti et al., 2015).

NBLs vs DPBLs determination

The NBL or Baseline level is defined as ‘‘the range of concentration of a given element, isotope or chemical compound in solution, derived entirely from natural, geological, biological or atmospheric sources, under conditions not perturbed by anthropogenic activity” (Edmunds and Shand, 2009). Groundwaters from aquifers that are part of the active water cycle are influenced by human activities. Water changes chemistry from the moment it enters the system through rainfall infiltration, exchanges with surface water bodies or through other sources, until it leaves through runoff, evaporation or withdrawal (Vázquez-Suñé et al., 2005). Shallow aquifers in fact rarely reflect true natural concentration levels, whereas deep aquifers may be more likely free from anthropogenic impacts (Muller et al., 2006). Groundwater status in highly urbanized and farm areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. This makes challenging to determine whether the observed groundwater conditions are the result of a natural chemical status according to the WFD directive or not (Wendland et al., 2005). Most of the methods used to determine NBLs rely on the separation of the natural from anthropogenic components based on some indicator chemical species, such as NO_3, Cl or SO_4 (Hinsby et al., 2008; Matschullat et al., 2000; Muller et al., 2006) and the identification of the NBL as a fixed percentile value (e.g. 90^{th} or 97^{th}) of the observed distribution of the indicator concentrations. Another common characteristic of these methods is the univariate approach since, even when considering a set of parameters, the NBLs are evaluated on the basis of univariate frequency distributions. Moreover, even though many of the studies dealing with NBLs determination have also come up with a mapping of the contamination patterns (De Caro et al., 2017), only very few studies (Busico et al., 2018; Hasenmueller and Criss, 2013; Hwang et al., 2015) have attempted to determine some kind of anthropogenic DPBLs. One of the most innovative aspects of this study is the use of a multivariate approach to determine DPBLs.

Used Data and pre-processing

Ten independent geochemical datasets, provided by local and regional agencies, and covering the period 2003–2014 were merged into a single database after homogenization and multiple quality checks. A total of 618258 chemical analyses from 3477 sampling wells (Fig. 3) were available.
All samples were collected and analyzed through internally consistent protocols (e.g. ASTM Standard methods). Measurements distribution across the sampling wells is quite heterogeneous. About half of the sampling points presents a number of measurements spanning from to 2 to 20. Only 10% of the sampling wells had more than 34 measurements. Figure 3 shows the spatial distribution of sampling wells across the FUA and the data availability at each sampling well.

FIGURE 3

The database included sampling point coordinates, screens depth and a code indicating the aquifer to which each record belongs. Data quality and consistency were checked by considering: (i) data out of the normal range, (ii) presence of outliers (iii) absence of depth or other aquifer information. Errors in the dataset were corrected only in the few cases where obvious data entry errors were identified (i.e. manual correction of wrong measure unit or magnitude of the sampled parameter) or where it was possible to obtain the correct information from the dataset source. The analysis of missing values and non-detects was also performed. The frequency of valid values (i.e. non missing) of all retained constituents was generally larger than 75-65 percent (see Table 1). Concentrations below the Limit of Quantification (LOQ) were replaced with a value of half the LOQ.

Table 1

Moreover, the study area’s lithology and precipitation regime were considered. Particularly the lithology layer, extracted from the Regione Lombardia geoportal was used (http://www.geoportale.regione.lombardia.it/), which contains the following information: the genetic classification of surface deposits; the classification of the rocky substratum based on the characteristics of petrographic composition, structure and texture; the main structural features and layers as the depths of the soils reported in classes (e.g. 0-50 cm; 50-100 cm; 100-200 cm). Moreover, precipitation and temperature data available at 15 meteorological stations included in the FUA, were used to estimate the rainfall depurated by evapotranspiration based on the Thornthwaite method (Thornthwaite and Mather, 1955).

Then the net precipitation monthly values have been interpolated through a kriging technique and used for the analysis. Moreover, a grid of 1x1 km^2 was used to associate both lithology and precipitation to spatial units and to clusters. Particularly lithology has been evaluated either in terms of areal fraction of the different lithological types, either as overall similarity of the cluster. As similarity metric the Percentage Similarity Index, (i.e. PSC, after Brock, (1977) see Eq.1) was used:

$$PSC = 100 - 0.5 \sum_{i=1}^{k} |a - b| \qquad \text{Eq.1}$$

as the sum of the percent differences of the lithology i between a pair of grid units.

Statistical Analysis

All the statistical computations were made using the statistical package IBM SPSS Statistics 24.0. Principal Component and Factor Analysis (hereinafter FA, cfr. Afifi et al., (2003)) were performed based on the correlation matrix of the concentration measurements. Particularly, Factor Analysis was obtained through the preliminary Principal Components Analysis (hereinafter PCA) which extracted the eigenvalues and eigenvectors from the covariance matrix of the original variances. Factor analysis was chosen to reduce the contribution of the less significant parameters within each component, by extracting a new set of varifactors through rotating the axes defined by the PCA extraction. The Varimax rotation criterion was used to rotate the PCA axes allowing to maintain the axes orthogonality. The number of factors to be retained was chosen on the basis of the "*eigenvalue higher than 1*" criterion (i.e. all the factors that explained less than the variance of one of the original variables were discarded). That allowed to select few factors able to describe the whole dataset with minimum loss of original information. Moreover, K-means Cluster Analysis (hereinafter CA, cfr. Afifi et al., (2003)) was used to analyze the similarities among the water quality profiles at the different monitoring stations, using the Euclidean Distance as distance metric (see Eq. 2).

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^{p}\left(x_{ik} - y_{jk}\right)^2} \qquad \text{Eq.2}$$

where i and j refer to a couple of stations, and p to the considered parameters.

CA was run based on the FA extracted varifactors. Due to the fact that the k-means procedure is somewhat sensitive to the initial choice of seeds, CA was run twice using the final cluster centroids obtained from the first CA analysis as initial seeds for the second run (cfr. Afifi et al., (2003)). Different CA trials were run to identify the optimal K for the cluster solution. The final choice was made based on cluster interpretability and stability across different CA results.

The Pearson's linear correlation coefficient (r), where adequate sample size was available, and the Spearman's rank order correlation coefficient (rho) were used for the correlation analysis. Kruskal-Wallis test by Rank was instead used to test the difference among cluster medians of lithological class fractions and similarities; Bonferroni's correction was applied for multiple comparisons.

Results

Preliminary analysis of multivariate dataset

In consideration of the fact that PCA/FA had to be applied to the whole dataset using a listwise deletion criterion of missing values (i.e. case units with missing values in one or more variables are discarded from the analysis), the parameters having a percent of missing values higher than 35% were preliminary excluded. Hexavalent chromium, although a priority pollutant, was excluded in reason of the poor availability across the study area and in consideration of its high correlation with total chromium (r: 0.979, $P<0.01$).

The application of the listwise deletion criterion reduced the size of multivariate dataset to about 60% (N: 25800) of the mean size of whole dataset concerning the investigated parameters (N: about 42300).

In consideration of the sample size reduction, the representativeness of the multivariate dataset was investigated. As shown in Table 2, the multivariate dataset, even though characterized by a smaller sample size, resulted to be fully representative of the whole dataset.

TABLE 2

As it can be observed, in fact, the distribution quartiles of the four pollutants of main interest for this study are quite comparable in the two situations; mean and standard deviation, being strongly affected by outliers that are only present in the whole dataset, do not correspond. Outliers in the whole dataset mostly refer to extremely high values measured at some specific pollution hotspot sites. In these sites, very often, only the information about pollutants was available, often lacking data about the other parameters of the multivariate dataset. In consideration of the fact that multivariate dataset had to be used to determine DPBLs, the absence of extreme outliers in the multivariate dataset was considered beneficial for the analysis. Moreover, in order to assess whether the spatial coverage of the multivariate dataset was about the same of the whole dataset considering the single parameters independently, the centroids of the measurement site coordinates, weighted by the corresponding frequency of measurements, obtained from either the multivariate or the whole dataset were compared. As it can be observed in Figure 4, the two centroids are quite close, being the distance between them less than 500 m, so it can be concluded that the multivariate dataset has about the same spatial coverage of the whole dataset.

FIGURE 4

Principal Component and Factor Analysis

PCA applied to the multivariate dataset led to the extraction of five principal components explaining overall the 78% of the total variance (see Table 3). As shown in Table 3, the first two principal components bring most of the information (about 53%).

Table 3 shows the comparison between the factor loadings obtained by PCA and the loadings of the FA rotated solution.

TABLE 3

The two sets of loadings are very similar, although in the rotated solution some parameters have “migrated” from a component to another. PCA solution had in fact a first component constituted by all the main hydrochemical parameters (i.e. conductivity and the main ions) explaining alone 40% of the total variance, a second component strongly linked to PCE and TCE explaining about 13% of the total variance, and the following components, each explaining less than 10% of the total variance. Moreover, in the PCA solution there are some parameters, namely Na, pH, Total Cr and TCM which load on two different components making more difficult the component interpretation.

On the other hand, the Varimax rotated solution (see Table 3) makes the PCA loadings either large or small on single varifactors, facilitating the interpretation. Particularly, in the rotated solution the hydrochemical parameters, which in the PCA solution almost all loaded the first component, are now loading both the first (i.e. explaining 25% of total variance) and the second varifactor (i.e. explaining about 24% of total variance); the pollutants Cr and TCM show also a cleaner distribution across the rotated varifactors, being total chromium and TCM both respectively loaded only on the fourth (i.e. 8.5% of the total variance) and on the fifth (i.e. 8% of the total variance) varifactor.

In summary the extracted varifactors are the following:

Varifactor 1: Accounts 25.1% of the total variance and is loaded by Conductivity, Cl, Na, K, SO_4; this varifactor represents the conductivity component that is most influenced by anthropic activity particularly concerning the salinization effect as described later; the high correlation between SO_4, Cl and Na has been related to the leakage of agricultural and municipal wastes (Sikora et al., 1976).

Varifactor 2: Accounts 23.8% of the total variance and is loaded by Conductivity, Ca, Mg, NO_3 and pH; this varifactor represents the conductivity-hardness component of the groundwater. The correlation between Ca and NO_3 can be associated with the use of fertilizers such as $NH_4NO_3{\cdot}CaCO_3$ (22% N and 33% $NCaCO_3$) which is very common in cultivated regions of the study area (Tisdale and Nelsson, 1975).

Varifactor 3: Accounts 12.6% of the total variance and is loaded by PCE and TCE. Due to their widespread use and subsequent disposal, and their chemical affinity these two pollutants are commonly found in groundwaters and often correlated.

Varifactor 4: Accounts 8.5% of the total variance and is loaded by Cr. In this area Cr(VI) represents in average the 74% of Total-Cr.

Varifactor 5: Accounts 8% of the total variance and is loaded by TCM which is also a common groundwater contaminant of industrial origin.

K-means Cluster Analysis

K-means cluster analysis was applied to the extracted FA varifactors. K was set to 15 and the analysis was run twice, using the final cluster centroids obtained from the first analysis as initial seeds for the second. Figure 5.a) shows the 15 clusters' composition in terms of number of cases included. It can be observed that most of the clusters contain a very small amount of data (i.e. less than 1% of the total cases) while few clusters, namely cluster 3, 5, 6, 7 and 13, contain more than 98% of the records including most of the background value measurements. Figure 5.b) shows the cluster characteristics in terms of varifactors.

FIGURE 5

It can be observed that clusters representing less than 1% of the total measurements are also the ones characterized by the higher values of the varifactors. Those clusters include in fact measurements that are outliers with respect of one or more of the considered parameters. Figure 6 shows the characteristics of the five main clusters:

- *Cluster 3* is characterized by PCE and TCE values higher than the average (less than a standard deviation), and TCM values much higher than the average (almost two standard deviations);
- *Cluster 5* is characterized by average or below average values concerning all the parameters;
- *Cluster 6* is also characterized by average values concerning all the parameters with the exception of those loading the first varifactor, namely conductivity, Cl, Na, and sulfates that is much higher than the average;
- *Cluster 7* is characterized by average values concerning all the parameters with the exception of Total Chromium which loads the fourth varifactor, which has a mean value much higher than the average.
- *Cluster 13* is characterized by average values concerning all the parameters with the exception of those loading the second varifactor, namely conductivity, Ca, Mg, nitrates and pH, which is higher than the average.

FIGURE 6

Cluster interpretation and trend

Cluster analysis has been applied to the whole dataset without separating boreholes located in different aquifers so the composition of the cluster in terms of aquifers was also investigated. As shown in Figure 7, the five main clusters contain boreholes with screens located in different aquifers since they present a water quality profile similar in both the unconfined (A) and semiconfined (B) aquifer.

FIGURE 7

Figure 8 shows as an example the PCE annual average concentration profiles in the three bigger clusters (i.e. 5, 6 and 13) comparing the unconfined (A) with the semiconfined aquifer (B) profile.

FIGURE 8

It can be observed that they present a comparable temporal trend. Such trends have been also tested through a Spearman's rho rank correlation analysis revealing that the PCE concentration profiles in the unconfined (A) and semiconfined (B) aquifers are correlated at a 0.05 and a 0.10 significance level. The Spearman's correlation coefficients between A and B aquifers in the three bigger clusters (i.e. 5, 6 and 13) respectively are: 0.519 (P: 0.084; n: 12), 0.627 (P: 0.039; n: 11) and 0.591 (P: 0.056; n: 11).

Considering the limited size of the analyzed time series, this result strongly supports the hypothesis that the different aquifers in this area present the same temporal variability. Figure 9 shows the cluster temporal profiles in terms of the five varifactor components. It can be observed that the five clusters present different temporal trends:

- *Cluster 3*: presents an increasing annual trend of the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.359 $P<0.01$) and a decreasing trend of the second (i.e. Ca, Mg, N-NO_3, pH r: - 0.139 $P<0.01$). No trend is shown by the PCE/TCE and TCM varifactors whereas the fourth varifactor shows a weak decreasing trend (i.e. Total Cr; r: - 0.118 $P<0.01$).

- *Cluster 5*: as far as the conductivity components are concerned, cluster 5 presents a weak increasing trend of the first (i.e. Cl, Na, K, SO_4; r: 0.162 $P<0.01$) and a very weak decreasing trend of the second varifactor (i.e. Ca, Mg, N-NO_3; r: -0.03 $P<0.01$). Concerning the pollutant components, cluster 5 presents a weak increasing trend of the PCE/TCE varifactor (r: 0.217 $P<0.01$) and decreasing trends for the Total chromium and the TCM varifactors (respectively: r: -0.191 $P<0.01$; r: -0.168 $P<0.01$).
- *Cluster 6*: cluster 6 presents no trend concerning the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.162 $P<0.01$) and a weak increasing trend of the second varifactor (i.e. Ca, Mg, N-NO_3; r: 0.132 $P<0.01$). Moreover, cluster 6 presents weak increasing trends of both the PCE/TCE varifactor (r: 0.116 $P<0.01$) and TCM varifactors (r: 0.052 $P<0.01$). No significant trend is shown by the Total-chromium varifactor.
- *Cluster 7*: cluster 7 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.420 $P<0.01$) and a decreasing trend of the second (i.e. Ca, Mg, N-NO_3, pH r: - 0.225 $P<0.01$). Concurrently, the trend is very weakly decreasing for the PCE/TCE and Total chromium varifactors (respectively: r: -0.066 $P<0.01$; r: -0.07 $P<0.01$), while the TCM varifactor does not show any significant trend ($P>0.05$).
- *Cluster 13*: cluster 13 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO_4; r: 0.225 $P<0.01$) and a decreasing trend of the second (i.e. Ca, Mg, N-NO_3, pH r: - 0.169 $P<0.01$). Moreover, while the trend is increasing for the PCE/TCE varifactor (r: 0.163 $P<0.01$), Total chromium and TCM varifactor trends are significantly decreasing (respectively: r: -0.237 $P<0.01$; r: -0.248 $P<0.01$).

FIGURE 9

Besides the different temporal trends, the clusters have also a well defined different spatial distribution (see Fig.10).

FIGURE 10

The lithological composition was compared among the different clusters. In order to simplify the analysis, only the main lithological classes (i.e. clays, gravels, silts, sands, no soil) were considered and compared among the clusters. The lithological composition of the five clusters (Fig.11) resulted to be significantly different at the KW test concerning the gravel ($P < 0.05$), silt ($P<0.01$) and sand fractions ($P < 0.05$).

FIGURE 11

No significant difference was found in terms of clay ($P > 0.60$) and no soil ($P > 0.30$) fractions among the clusters. Percentage similarity indexes (see Fig. 12) of the cluster lithology was also found statistically different ($P<0.001$).

FIGURE 12

Table 4 shows the multiple comparison test results.

TABLE 4

The correlation with net precipitation (i.e. depurated by evapotranspiration) was also considered. Table 5 shows the correlation analysis of the five varifactors with net precipitation, cluster by cluster. It can be observed that most of the varifactors presents a correlation with both the year and net precipitation.

TABLE 5

Towards the DPBLs determination

The 5 identified clusters, representing the 98% of the sample, offer a robust support for the definition of DPBLs for the study area. As shown in Figure 13 in fact the five main clusters are very different also in terms of pollutant background values (see Supplementary tables S1, S2, S3) and their own spatial distribution suggest that in the study area coexist different water quality profiles and different pollution background levels.

FIGURE 13

Particularly, clusters 3 and 7 (see Figure 10) have a more localized spatial distribution and show higher DPBLs in all the aquifers. Particularly, cluster 3 is characterized by solvents concentrations of one magnitude higher than the GQS (e.g. PCE: 1.1 ug/l; TCE 1.5 ug/l; TCM: 0.15 ug/l), while cluster 7 has chromium and nitrates levels that respectively are almost double or very close to the corresponding GQS (respectively Cr-IV: 5 ug/l and NO_3: 50 mg/l). On the other hand, clusters 5 and 13 have a more diffuse spatial distribution and tend to have lower background values concerning all the pollutants. Cluster 6 is instead mostly concentrated within the municipality of Milan and it shows also well-defined characteristics: chromium levels that are higher than the ones of cluster 5 and 13 but lower than the GQS, solvents and nitrates levels comparable with cluster 7 levels, being the PCE and TCM levels (TCE levels are higher than the GQS only in the semiconfined aquifer) higher than the corresponding GQS all over the area. Overall, it is relevant to observe, that both PCE and TCE present a wide logarithmic variability in their concentration values, concerning all the clusters, being respectively 3, 7 and 6 the cluster with the higher levels, which on some spots are many orders of magnitude higher than the GQS. Overall, background pollution levels in the different aquifers follow the same pattern of the cluster, being generally the background levels in the semiconfined (B) and in the undifferentiated aquifer higher that the corresponding levels in the unconfined aquifer (A).

Discussion

A widespread pattern of pollution occurrence in groundwaters has been documented in many urban areas all around Europe (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroo et al., 2003). In time plumes coming from single point-sources (PS), mix with each other and end into a diffuse pattern of pollution linked to unidentifiable Multi-Point Sources, which, in most cases, started decades ago (Cortés et al., 2011).

Remediation of diffuse pollution to GQS levels, which very often are much lower than the background levels, is very difficult if not very unfeasible, and decision-makers, in many situations concerning Functional Urban Areas, are in fact urged to define new remediation goals, different from the GQS which are very often determined through a conservative risk assessment process. However, such a procedure is generally conducted by each Regional authority on case-by-case basis process while a national and European legislation about the management of diffuse pollution is still lacking. There is no doubt that relying on site-specific risk analysis in a FUAs with large patterns of diffuse contamination, it substantially increases clean-up associated costs compared to a very small improvement of groundwater quality. On the other hand, assessing the type of pollution, whether diffuse or localized, is critical as Public Authorities need to face those contaminations applying completely different actions, with dramatically different associated costs.

If well-established methodological approaches exist for the identification of Natural Background Levels (Matschullat et al., 2000) of groundwater contaminants, no consolidated method either exist for identifying the anthropogenic groundwater diffuse pollution. The methodology proposed in this study is not intrinsically innovative, since multivariate techniques of the kind applied have been also used in several monitoring studies concerning both surface and groundwaters (Busico et al., 2018; Gourdol et al., 2013; Marcelli et al., 2010; Mendizabal et al., 2011; Olsen et al., 2012; Selle et al., 2013; Sheikhy Narany et al., 2014; Yidana et al., 2008; Yu et al., 2014); however, to our knowledge, this is the first time that these multivariate methods are employed to define DPBLs in FUA where MPS are present. The strength of this data driven approach is the possibility to identify hidden patterns in monitoring data (e.g. the hydrosomes as in Mendizabal et al., (2011)) directly

from the dataset. The analysis showed clearly that the Milan FUA is characterized by different hydrochemical groundwater subsystems with well differentiated water quality characteristics, which vary in time and space. The five main clusters identified through the analysis were in fact found to have different temporal profiles and background concentration levels. Moreover, the clusters were found to be characterized by different lithological characteristics. The acquired evidence is not enough to support the hypothesis that these clusters are a sort of hydrosomes as the ones described by (Mendizabal et al., 2011) (e.g. *hydrosome* is defined as a coherent, three-dimensional unit of groundwater with a specific origin) but it certainly supports the hypothesis that the FUA should not be seen or managed as an homogeneous groundwater body. Concerning the hydrogeochemical constitutes, in four of the five main clusters, the analysis reveals the presence of an increasing trend of the first varifactor (i.e. Na, K, Cl, SO_4) and a weak but highly significant ($P<0.01$) decreasing trend of the second varifactor (i.e. Ca, Mg, NO_3, pH). While the increasing trend of the first varifactor is a quite widespread pattern that has been found in many urbanised and cultivated territories, which to date concerns more than one-third of the world's irrigated land (Abbas et al., 2013; Heuperman et al., 2002; Houk et al., 2006; Li et al., 2015), the weak decreasing trend of the second varifactor might be due to cation exchange. As shown in Figure 14a, the dissolution of calcite, dolomite and gypsum seems to be the dominant reaction in the system as indicated in the plot of (Ca^{2+}+ Mg^{2+}) vs. (HCO_3^{-} + SO_4^{2-}), which is close to a 1:1 line.
The plot of Ca^{2+} and Mg^{2+}concentrations compared with those of $HCO3^{-}$ mostly in cluster 6 and 13 may indicate the existence of an additional Ca^{2+} and Mg^{2+} source. Such a result seems to be consistent with De Caro et al., 2017 findings which revealed higher calcium concentrations in the northeastern sector of their study area which partially corresponds to ours. It is in fact worthwhile to observe that those high ratios between Ca^{2+} and Mg^{2+} and $HCO3^{-}$ concentrations may not be attributed to HCO_3 depletion because of the existing neutral to slightly alkaline conditions (e.g.

groundwater pH ranges from 6.85 to 8.10, with an average of 7.52 and a standard deviation of 0.25) which do not favour the formation of carbonic acid (H_2CO_3, see Spears, (1986)). Moreover, according to Jankowski et al. (1998), if active cation exchange between Na^{+} and Ca^{2++} Mg^{2+} is occurring in the aquifer, the slope of the bivariate plot between Cl^{-} corrected Na^{+} + K^{+} and (Ca^{2+} + Mg^{2+}) corrected HCO_3^{-} + SO_4^{2-} (see Figure 14c) would be −1 (i.e., y = −x). The slope obtained for this plot in the study area (−1.72) seems to indicate that cation exchange is acting as an important process.

FIGURE 14

It is also worthwhile to underline that correlations with precipitation are weak, being insignificant (i.e. cluster 3) or weakly direct concerning the first varifactor, and weakly indirect concerning the second varifactor. The correlation between Na+ and Cl− in the five main clusters is also different, being always significant at the $P<0.01$ level, and higher than 0.6-0.7 in the smaller clusters (i.e. 0.76, 0.63 and 0.73 respectively for clusters 3, 6 and 7 distributed mostly within the Milan municipality) and lower than 0.5 for the bigger clusters (i.e. 0.38, and 0.55 respectively for clusters 5, and 13). In this respect it has been suggested that the poor correlation ($r2 < 0.54$ that would correspond to an $r < 0.74$) between Na+ and Cl might indicate a possible evaporative concentration of these ions (Rose, 2002). Na^{+}/Cl^{-} ratio has been used, especially in semiarid or arid regions, to identify the mechanism for acquiring salinity, since the Cl^{-} ion is not affected by soil retention while the Na^{+} ion is retained (Jalali and Khanlari, 2008; Tiwari and Singh, 2014); enrichment of Na+ may also result from reactions taking place in the clay mineral of the surface soil horizons. As shown in Figure 14d the clusters, especially the bigger ones, have a marked fingerprint with respect to the Na+/Cl− ratio vs Cl− concentrations being cluster 5 the one of the highest Na+ enrichment and cluster 6 the one with the lowest.

As far as the pollutant varifactors are concerned, it's interesting to observe that FA clearly separates their variability from the variability of the hydrogeochemical characteristics, since the pollutant components show also specific trends which are also different within the clusters. Particularly, while the third varifactor (i.e. PCE and TCE) appears to be slightly increasing in time in all the bigger clusters (5, 6 and 13) and to be stable or slightly decreasing in the smaller clusters 3 and 7, the fourth and fifth varifactors (i.e respectively Total-Cr and TCM) show instead different patterns: total Cr is stable (i.e. cluster 6) or a decreasing in all the other clusters while TCM is stable in clusters 3 and 7, weakly increasing in cluster 6 and decreasing in clusters 5 and 13.

Correlations with precipitation for the pollutant varifactors are all very weak although significant suggesting that meteo-climatic variability may also affect the observed patterns. As pointed out earlier, cluster 3 and 7 both have higher average concentrations of the studied contaminants: cluster 3 having PCE and TCM levels respectively 2 times and 5 times higher than the whole sample average, and cluster 7 showing PCE and Total-Cr levels 2 times higher the whole sample average. The wells belonging to these clusters are mostly inside the boundaries of Milan municipality and are positioned downgradient to a couple of historical industrial district areas where many brownfields were remediated in the last 20 years. So, despite the effort done in the past to find the contamination hot spots (i.e. PS) and to remediate soil and groundwater, there are still traces of the past production activities whose impact can still be detected. This is because many small unidentifiable sources (i.e. MPS) are still present, but since they are releasing a low mass of contaminants there is no chance to locate their position and to apply a specific remediation action. Luckily concentrations in these two clusters are showing a decreasing trend in time, probably due to Natural Attenuation Processes that could be assessed, monitored and maybe enhanced to accelerate the achievement of the groundwater quality status. Territorial authorities need support in the process of identifying such trends, and their awareness of the existence of portions of the groundwater body that may have different quality characteristics and trends need to be arisen. For the same reason appropriate DPBLs should be identified for these portions since only through this process it would be possible to properly manage these areas applying adequate planning policies and actions (e.g. investigations and monitoring). We showed that diffuse pollution is present within the FUA with different water quality levels and different distribution: cluster 13 being the dominant water quality profile representing the 35% of the boreholes, followed by cluster 5 representing the 31% and by cluster 6 representing the 21%. Among the contaminants considered in this study only PCE and TCM systematically exceed their corresponding GQSs with median and interquartile values respectively of 3, 1÷7 μg/l, and 0.5, 0.5÷2 μg/l, while TCE and Cr-VI present only spots of higher contamination, with median and interquartile values respectively of 1, 0.5÷2.7 μg/l and 3.9, 2.3÷6.6 μg/l. Defining appropriate DPBLs for these contaminants is both a political and a scientific-technical issue and is beyond the scope of this paper, however, we the evidence provided by this study about the FUA not being a whole homogeneous groundwater body strongly support the hypothesis that different DBPLs should be defined within FUA subareas. Moreover, as weak trends were found, revealing a slow temporal evolution of the contamination (i.e. in average less than 0.06 ug/l per year for the considered decade). Reliable decennial background levels should be drawn every 5-10 years with the caveat of periodically revising. Furthermore, being the clusters robust and stable in time, they allow to overcome the problem of spatial and temporal heterogeneity of measurements which particularly affects this area and offer a robust data series to define spatially the DPBLs through geostatistical approaches. Deterministic fate and transport models will also be useful to simulate the most important plumes deriving from a PS contamination in the FUA, and to identify the subareas affected by these plumes that should be excluded from the geostatistical interpolation.

In summary, it can be concluded that the picture offered by the applied multivariate analysis was able to synthetize the main hydrogeochemical processes acting in the study area (i.e. salinization, ion exchanges processes and the main diffuse pollution patterns), allowing to identify different temporal profiles and to lay the basis for the definition of Diffuse Pollution Background Levels.

Conclusions

Groundwater status in highly urbanized and farm areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from many sources. In the EU legislative framework there is no indication about how to define Diffuse Pollution Background Levels, which in many situations can be higher than GQSs and make unfeasible any remediation strategy. The methodology applied in this study enabled to identify five main clusters, having specific hydrogeological characteristics and different temporal profiles and pollutant background concentration levels. The evidence provided by this study strongly suggests that the FUA should not be managed as a homogeneous groundwater body, and outlines the need of defining Diffuse Pollution Background levels at least for PCE and TCM which were systematically found to exceed their corresponding GQSs. The clusters described in this study offer a robust knowledge basis for the drafting a diffuse pollution management plan of the area.

Acknowledgements

This study has been funded within the framework of the AMIIGA Project (Central Europe Interreg (AMIIGA - CE32) and of the PLUME project funded by Regione Lombardia. We are deeply grateful to Elisabetta Preziosi who gave us the right perspective to describe the legislative context of WFD and GWD, and to the many anonymous reviewers who greatly helped us with their valuable comments to significantly improve the original manuscript.

References

Abbas, A., Khan, S., Hussain, N., Hanjra, M.A., Akbar, S., 2013. Characterizing soil salinity in irrigated agriculture using a remote sensing approach. Phys. Chem. Earth. doi:10.1016/j.pce.2010.12.004

Afifi, A., May, S., Clark, V.A., 2003. Cluster analysis, in: Computer-Aided Multivariate Analysis. pp. 428–444.

Alberti, L., Cantone, M., Colombo, L., Lombi, S., Piana, A., 2016. Numerical modeling of regional groundwater flow in the Adda-Ticino Basin: Advances and new results. Rend. Online Soc. Geol. Ital. 41. doi:10.3301/ROL.2016.80

Alberti, L., Colombo, L., Formentin, G., 2018. Null-space Monte Carlo particle tracking to assess groundwater PCE (Tetrachloroethene) diffuse pollution in north-eastern Milan functional urban area. Sci. Total Environ. 621. doi:10.1016/j.scitotenv.2017.11.253

Alberti, L., Colombo, L., La Licata, I., Cantone, M., 2018. Model calibration using the automatic parameter estimation procedure (PEST) of the North-eastern zone of the Milan Functional Urban Area (Italy). Acque Sotter. - Ital. J. Groundw. 7, 27–38. doi:10.7343/as-2018-336

Balderacchi, M., Benoit, P., Cambier, P., Eklo, O.M., Gargini, A., Gemitzi, A., Gurel, M., Kløve, B., Nakic, Z., Predaa, E., Ruzicic, S., Wachniew, P., Trevisan, M., 2013. Groundwater pollution and quality monitoring approaches at the European level. Crit. Rev. Environ. Sci. Technol. doi:10.1080/10643389.2011.604259

Balderacchi, M., Filippini, M., Gemitzi, A., KlÃ¶ve, B., Petitta, M., Trevisan, M., Wachniew, P., Witczak, S., Gargini, A., 2014. Does groundwater protection in Europe require new EU-wide environmental quality standards? Front. Chem. doi:10.3389/fchem.2014.00032

Barrlo-Lage, G., Parsons, F.Z., Nassar, R.S., 1987. Kinetics of the Depletion of Trichloroethene. Environ. Sci. Technol. 21, 366–370. doi:10.1021/es00158a005

Bonomi, T., 2009. Database development and 3D modeling of textural variations in heterogeneous, unconsolidated aquifer media: Application to the Milan plain. Comput. Geosci. doi:10.1016/j.cageo.2007.09.006

Brock, D.A., 1977. Comparison of community similarity indexes. Water Pollut. Control Fed. 49, 2488–2494.

Busico, G., Cuoco, E., Kazakis, N., Colombani, N., Mastrocicco, M., Tedesco, D., Voudouris, K., 2018. Multivariate statistical analysis to characterize/discriminate between anthropogenic and geogenic trace elements occurrence in the Campania Plain, Southern Italy. Environ. Pollut. doi:10.1016/j.envpol.2017.11.053

Carcano, C., Piccin, A., 2002. Padans aquifer geology of the Lombardy Plain. Firenze.

Carral, E., Puente, X., Villares, R., Carballeira, A., 1995. Background heavy metal levels in estuarine sediments and organisms in Galicia (northwest Spain) as determined by modal analysis. Sci. Total Environ. doi:10.1016/0048-9697(95)04788-3

Carter, J.M., Moran, M.J., Zogorski, J.S., Price, C. V., 2012. Factors associated with sources, transport, and fate of chloroform and three other trihalomethanes in untreated groundwater used for drinking water. Environ. Sci. Technol. doi:10.1021/es301839p

Chapman, S.W., Parker, B.L., 2005. Plume persistence due to aquitard back diffusion following dense nonaqueous phase liquid source removal or isolation. Water Resour. Res. doi:10.1029/2005WR004224

Chapman, S.W., Parker, B.L., Sale, T.C., Doner, L.A., 2012. Testing high resolution numerical models for analysis of contaminant storage and release from low permeability zones. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2012.04.006

Colombo, L., Gattinoni, P., Scesi, L., 2018. Stochastic modelling of groundwater flow for hazard assessment along the underground infrastructures in Milan (northern Italy). Tunn. Undergr. Sp. Technol. 79. doi:10.1016/j.tust.2018.05.007

Cortés, A., Puigserver, D., Carmona, J.M., Viladevall, M., 2011. Biological remediation approach involving soils and groundwaters polluted with chlorinated solvents in a Mediterranean context. Recent Adv. Pharm. Sci. 661, 223–246.

De Caro, M., Crosta, G.B., Frattini, P., 2017. Hydrogeochemical characterization and Natural Background Levels in urbanized areas: Milan Metropolitan area (Northern Italy). J. Hydrol. doi:10.1016/j.jhydrol.2017.02.025

De Giusti, F., Piaz, D., Massironi, M., Schiavo, A., 2003. Carta geotettonica della Valle d ’ Aosta. Mem. Sci. Geol 129–149.

Edmunds, W.M., Shand, P., 2009. Natural Groundwater Quality, Natural Groundwater Quality. doi:10.1002/9781444300345

Emsley, J., 2011. Nature’s Building Blocks: An A-Z Guide to the Elements, Oxford University Press. doi:978-0-19-960563-7

European Community, 2006. Groundwater directive 2006/118/CE. Directive of the European Parliament and of the Council on the Protection of Groundwater Against Pollution and Deterioration. Off. J. Eur. Union 19, 19–31. doi:http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006L0118

Francani, V., Beretta, G. Pietro, 1995. Preliminary survey about sustainable use of deeper groundwater in Milano province. Milan.

Fruchter, J., 2002. In-Situ Treatment of Chromium-Contaminated Groundwater. Environ. Sci. Technol. doi:10.1021/es022466i

Giudici, M., Ponzini, G., Romano, E., Vassena, C., 2007. Some Lessons from Modeling Ground Water Flow in the Metropolitan Area of Milano at Different Scales. Mem. Descr. della Cart. Geol. d’Italia LXXVI, 207–218.

Gourdol, L., Hissler, C., Hoffmann, L., Pfister, L., 2013. On the potential for the Partial Triadic Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry. Appl. Geochemistry. doi:10.1016/j.apgeochem.2013.10.002

Guilbeault, M.A., Parker, B.L., Cherry, J.A., 2005. Mass and flux distributions from DNAPL zones in sandy aquifers. Ground Water. doi:10.1111/j.1745-6584.2005.tb02287.x

Hasenmueller, E.A., Criss, R.E., 2013. Multiple sources of boron in urban surface waters and groundwaters. Sci. Total Environ. doi:10.1016/j.scitotenv.2013.01.001

Heuperman, A.F., Kapoor, A.S., Denecke, H.W., 2002. Biodrainage - Principles, Experiences and Applications. Rome.

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

Houk, E., Frasier, M., Schuck, E., 2006. The agricultural impacts of irrigation induced waterlogging and soil salinity in the Arkansas Basin. Agric. Water Manag. 85, 175–183. doi:10.1016/j.agwat.2006.04.007

Hunkeler, D., Aravena, R., 2000. Determination of compound-specific carbon isotope ratios of chlorinated methanes, ethanes, and ethenes in aqueous samples. Environ. Sci. Technol. doi:10.1021/es991178s

Hwang, H.-H., Panno, S. V., Hackley, K.C., 2015. Sources and Changes in Groundwater Quality with Increasing Urbanization, Northeastern Illinois. Environ. Eng. Geosci. doi:10.2113/gseegeosci.21.2.75

ISTAT, 2017. Annual survey on movement and calculation of the resident population. Rome.

Izbicki, J.A., Ball, J.W., Bullen, T.D., Sutley, S.J., 2008. Chromium, chromium isotopes and selected trace elements, western Mojave Desert, USA. Appl. Geochemistry. doi:10.1016/j.apgeochem.2007.11.015

Jalali, M., Khanlari, Z. V., 2008. Major ion chemistry of groundwaters in the Damagh area, Hamadan, western Iran. Environ. Geol. doi:10.1007/s00254-007-0795-6

Kao, C.M., Lei, S.E., 2000. Using a peat biobarrier to remediate PCE/TCE contaminated aquifers. Water Res. doi:10.1016/S0043-1354(99)00213-4

Kueper, B.H., Wealthall, G.P., Smith, J.W.N., Leharne, S. a, Lerner, D.N., 2003. An illustrated handbook of DNAPL transport and fate in the subsurface, R&D Publication 133.

Li, W., Wang, M.-Y., Liu, L.-Y., Yan, Y., 2015. Assessment of Long-Term Evolution of Groundwater Hydrochemical Characteristics Using Multiple Approaches: A Case Study in Cangzhou, Northern China. Water 7, 1109–1128. doi:10.3390/w7031109

Mackay, D.M., Cherry, J.A., 1989. Groundwater contamination: Pump-and-treat remediation. Environ. Sci. Technol. doi:10.1021/es00064a001

Marcelli, R., Bartolucci, G., Papaioannu, G., De Angelis, G., Lucibello, A., Proietti, E., Margesin, B., Giacomozzi, F., Deborgies, F., 2010. Reliability of RF MEMS switches due to charging effects and their circuital modelling, in: Microsystem Technologies. pp. 1111–1118. doi:10.1007/s00542-009-1006-z

Matschullat, J., Ottenstein, R., Reimann, C., 2000. Geochemical background - Can we calculate it?

Environ. Geol. doi:10.1007/s002549900084

McGuire, T.M., McDade, J.M., Newell, C.J., 2006. Performance of DNAPL source depletion technologies at 59 chlorinated solvent-impacted sites. Gr. Water Monit. Remediat. doi:10.1111/j.1745-6592.2006.00054.x

Mendizabal, I., Stuyfzand, P.J., Wiersma, A.P., 2011. Hydrochemical system analysis of public supply well fields, to reveal water-quality patterns and define groundwater bodies: The Netherlands. Hydrogeol. J. doi:10.1007/s10040-010-0614-0

Molinari, A., Chidichimo, F., Straface, S., Guadagnini, A., 2014. Assessment of natural background levels in potentially contaminated coastal aquifers. Sci. Total Environ. doi:10.1016/j.scitotenv.2013.12.125

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

Muller, D., Blum, A., Hart, A., Hookey, J., Kunkel, R., Scheidleder, A., 2006. Final Report - BRIDGE (Background criteria for the identification of groundwater thresholds) 13030.

Munholland, J.L., Mumford, K.G., Kueper, B.H., 2016. Factors affecting gas migration and contaminant redistribution in heterogeneous porous media subject to electrical resistance heating. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2015.10.011

Nriagu, J.O., Nieboer, E., 1988. Chromium in the Natural and Human Environments. John Wiley & Sons.

OECD, 2012. Redefining “Urban”: A New Way to Measure Metropolitan Areas. OECD Publ. 1–9. doi:10.1787/9789264174108-en

Olsen, R.L., Chappell, R.W., Loftis, J.C., 2012. Water quality sample collection, data treatment and results presentation for principal components analysis e literature review and Illinois River watershed case study. Water Res.

Paine, A.J., 2001. Mechanisms of chromium toxicity, carcinogenicity and allergenicity: Review of the literature from 1985 to 2000. Hum. Exp. Toxicol. doi:10.1191/096032701682693062

Parker, B.L., Cherry, J.A., Chapman, S.W., 2004. Field study of TCE diffusion profiles below DNAPL to assess aquitard integrity. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2004.02.011

Pecoraino, G., Scalici, L., Avellone, G., Ceraulo, L., Favara, R., Candela, E.G., Provenzano, M.C., Scaletta, C., 2008. Distribution of volatile organic compounds in Sicilian groundwaters analysed by head space-solid phase micro extraction coupled with gas chromatography mass spectrometry (SPME/GC/MS). Water Res. doi:10.1016/j.watres.2008.07.022

Pedretti, D., Masetti, M., Beretta, G.P., Vitiello, M., 2013. A revised conceptual model to reproduce the distribution of chlorinated solvents in the Rho Aquifer (Italy). Groundw. Monit. Remediat. doi:10.1111/gwmr.12017

Perego, R., Bonomi, T., Fumagalli, M.L., Benastini, V., Aghib, F., Rotiroti, M., Cavallin, A., 2014. 3D reconstruction of the multi-layer aquifer in a Po Plain area. Rend. Online Soc. Geol. Ital. 30, 41–44. doi:10.3301/ROL.2014.09

Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: Death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environ. Geol. doi:10.1007/s002549900081

Reinmann, C., De Caritat, P., 1998. Chemical Elements in the Environment. sPRINGER-vERLAG, Berlin.

Rivett, M.O., Feenstra, S., 2005. Dissolution of an emplaced source of DNAPL in a natural aquifer setting. Environ. Sci. Technol. doi:10.1021/es040016f

Rose, S., 2002. Comparative major ion geochemistry of Piedmont streams in the Atlanta, Georgia region: Possible effects of urbanization. Environ. Geol. doi:10.1007/s00254-002-0545-8

Rotiroti, M., Di Mauro, B., Bonomi, T., Fumagalli, M.L., 2015. COMPSEC , a new tool to derive natural background levels by the component separation approach : Application in two different hydrogeological contexts in northern Italy COMPSEC , a new tool to derive natural background levels by the component separation. doi:10.1016/j.gexplo.2015.06.017

Schwille, F. (1984), 1984. Migration of Organic Fluids Immiscible with Water in the unsaturated zone 2–3.

Schwille, F. (1988), 1988. Dense Chlorinated Solvents in Porous and Fractured Media: Model Experiments, translated from the German by J. F. Pankow.

Selle, B., Schwientek, M., Lischeid, G., 2013. Understanding processes governing water quality in catchments using principal component scores. J. Hydrol. doi:10.1016/j.jhydrol.2013.01.030

Sheikhy Narany, T., Ramli, M.F., Aris, A.Z., Sulaiman, W.N.A., Fakharian, K., 2014. Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol-Babol Plain, Iran. Environ. Monit. Assess. doi:10.1007/s10661-014-3820-8

Sikora, L., Bent, M., Corey, R., Keeney, D., 1976. Septic nitrogen and phosphorus removal test system. Groundwater 14, 309–314.

Spears, D.A., 1986. Mineralogical control of the chemical evolution of groundwter. Solute Process. 512.

Stevenazzi, S., Masetti, M., Nghiem, S. V., Sorichetta, A., 2015. Groundwater vulnerability maps derived from a time-dependent method using satellite scatterometer data. Hydrogeol. J. 23, 631–647. doi:10.1007/s10040-015-1236-3

Stroo, H.F., Unger, M., Ward, C.H., Kavanaugh, M.C., Vogel, C., Leeson, A., Marqusee, J.A., Smith, B.P., 2003. Peer Reviewed: Remediating Chlorinated Solvent Source Zones. Environ. Sci. Technol. 37, 224A–230A. doi:10.1021/es032488k

Thornthwaite, C.W., Mather, J.R., 1955. The water balance. Publ. Climatol. 8, 1–104.

Tisdale, S., Nelsson, W., 1975. Soil fertility and fertilizers. Macmillan Publishing Co Inc, New Jork.

Tiwari, A.K., Singh, A.K., 2014. Hydrogeochemical investigation and groundwater quality assessment of Pratapgarh district, Uttar Pradesh. J. Geol. Soc. India. doi:10.1007/s12594-014-0045-y

U.S.EPA, 1992. Dense Non Aqueous Phase Liquids - A Workshop Summary. Dallas.

Unceta, N., Séby, F., Malherbe, J., Donard, O.F.X., 2010. Chromium speciation in solid matrices and regulation: A review. Anal. Bioanal. Chem. doi:10.1007/s00216-009-3417-1

Vázquez-Suñé, E., Sánchez-Vila, X., Carrera, J., 2005. Introductory review of specific factors influencing urban groundwater, an emerging branch of hydrogeology, with reference to Barcelona, Spain. Hydrogeol. J. doi:10.1007/s10040-004-0360-2

Vogel, T.M., McCarty, P.L., 1985. Biotransformation of tetrachloroethylene to trichloroethylene, dichloroethylene, vinyl chloride, and carbon dioxide under methanogenic conditions. Appl. Environ. Microbiol.

Wanner, C., Zink, S., Eggenberger, U., Mäder, U., 2012. Assessing the Cr(VI) reduction efficiency of a permeable reactive barrier using Cr isotope measurements and 2D reactive transport modeling. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2012.01.007

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

Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H.J., Wolter, R., 2005. A procedure to define natural groundwater conditions of groundwater bodies in Germany. Water Sci. Technol.

WHO, 1988. Environmental Health Criteria. Geneve.

Yidana, S.M., Ophori, D., Banoeng-Yakubo, B., 2008. A multivariate statistical analysis of surface water chemistry data-The Ankobra Basin, Ghana. J. Environ. Manage. 86, 80–87. doi:10.1016/j.jenvman.2006.11.023

Yu, Y., Song, X., Zhang, Y., Zheng, F., Liang, J., Liu, L., 2014. Identifying spatio-temporal variation and controlling factors of chemistry in groundwater and river water recharged by reclaimed water at Huai River, North China. Stoch. Environ. Res. Risk Assess. doi:10.1007/s00477-013-0803-1

Table 1
Click here to download Table: Table 1_rev.docx

Table 1 – Summary statistics of the main parameters in the used dataset. Parameters used in the PCA/FA analysis are highlighted in bold.

								Percentiles		
	Valid	**% Missing**	**Mean**	**Median**	**Std. Dev.**	**Min**	**Max**	**25**	**50**	**75**
Ca	**40316**	**32.6**	**73.34**	**72.00**	**48.69**	**0.50**	**4000.00**	**52.00**	**72.00**	**92.00**
Cl	**43773**	**26.8**	**16.31**	**13.20**	**13.05**	**0.01**	**303.00**	**6.00**	**13.20**	**22.98**
Conductivity	**39468**	**34.0**	**460.82**	**455.93**	**170.41**	**0.35**	**2896.00**	**336.00**	**455.93**	**578.00**
Cr(VI)	14794	75.2	16.29	4.00	198.53	0.00	8022.90	1.94	4.00	6.96
Total Cr	**45552**	**23.8**	**9.09**	**2.50**	**149.99**	**0.00**	**12300.00**	**2.50**	**2.50**	**3.00**
Mg	**39396**	**34.1**	**15.85**	**16.00**	**5.72**	**0.50**	**249.00**	**12.00**	**16.00**	**20.00**
Nitrates	**46716**	**21.8**	**22.75**	**21.30**	**14.43**	**0.01**	**190.06**	**10.00**	**21.30**	**34.00**
pH	**39603**	**33.7**	**7.66**	**7.70**	**0.27**	**5.56**	**9.56**	**7.50**	**7.70**	**7.84**
K	**38667**	**35.3**	**0.91**	**0.50**	**0.71**	**0.10**	**40.00**	**0.50**	**0.50**	**1.00**
Dry residue	36352	39.2	318.01	313.14	106.00	0.25	909.00	245.00	313.14	398.00
Na	**39259**	**34.3**	**8.74**	**6.50**	**7.42**	**0.50**	**214.00**	**5.00**	**6.50**	**10.90**
Sulfates	**43915**	**26.5**	**30.99**	**30.00**	**27.49**	**0.01**	**1363.00**	**14.00**	**30.00**	**41.00**
Temperature	4041	93.2	14.92	14.90	1.51	7.70	26.60	14.00	14.90	15.80
PCE	**44912**	**24.9**	**14.87**	**3.00**	**310.32**	**0.00**	**37800.00**	**0.60**	**3.00**	**7.70**
TCE	**44569**	**25.4**	**13.97**	**1.00**	**261.90**	**0.00**	**14000.00**	**0.50**	**1.00**	**2.32**
TCM	**43541**	**27.2**	**3.55**	**0.50**	**88.30**	**0.00**	**17214.00**	**0.50**	**0.50**	**1.20**
Listwise N	**25800**									

Table 2
Click here to download Table: Table 2_rev.docx

Table 2 – Summary statistics of the four pollutants of main interest in the used dataset. The multivariate dataset summary statistics are representative of the whole dataset

		PCE		TCE		TCM		Total-Cr	
		whole	multivariate	whole	multivariate	whole	multivariate	whole	multivariate
N		45602	25757	45190	25757	44083	25757	46042	25757
Mean		15.316	6.193	13.816	13.818	3.542	2.535	9.114	3.854
Median		3.000	2.000	1.000	1.000	0.500	0.500	2.500	2.500
Std. Deviation		311.685	28.802	260.100	250.020	87.811	8.919	149.219	5.087
Minimum		0.003	0.010	0.003	0.010	0.005	0.010	0.005	0.200
Maximum		37800	3000	14000	14000	17214	341	12300	190
Percentiles	**25**	0.610	0.500	0.500	0.500	0.500	0.500	2.500	2.500
	50	3.000	2.000	1.000	1.000	0.500	0.500	2.500	2.500
	75	7.805	5.100	2.300	2.200	1.160	1.000	3.000	2.500

Table 3

Click here to download Table: Table 3 ex4.docx

Table 3 – Comparison of PCA and FA factor loadings. The absolute loading values higher than 0.5 are highlighted in bold.

Component Matrix

	Component				
	1	2	3	4	5
Ca	**0.895**	-0.250	-0.194	-0.012	-0.053
Cl	**0.839**	0.076	0.318	0.029	0.016
Conductivity	**0.960**	-0.128	0.020	-0.010	-0.017
Total Cr	0.151	-0.022	-0.219	**0.689**	**0.606**
Mg	**0.691**	-0.318	-0.311	-0.052	-0.130
NO3	**0.714**	-0.050	-0.194	0.189	0.057
pH	-0.485	0.123	**0.550**	0.246	0.214
K	**0.571**	-0.015	0.394	-0.146	0.102
Na	**0.642**	0.350	**0.506**	-0.137	0.097
Sulfates	**0.860**	0.119	0.100	0.090	0.014
PCE	0.203	**0.782**	-0.298	0.058	-0.015
TCE	0.143	**0.821**	-0.283	-0.203	0.054
TCM	0.037	0.195	0.158	**0.650**	**-0.693**
Eigenvalue	5.212	1.664	1.239	1.090	0.939
Explained variance (%)	40.094	12.802	9.535	8.381	7.226
Cumulative (%) of variance	40.094	52.896	62.430	70.811	78.037

Extraction Method: Principal Component Analysis.

Rotated Component Matrix

	Component				
	1	2	3	4	5
Ca	0.462	**0.822**	-0.056	0.116	0.003
Cl	**0.821**	0.345	0.045	0.077	0.104
Conductivity	**0.675**	**0.686**	-0.016	0.105	0.032
Total Cr	-0.005	0.043	0.021	**0.954**	-0.005
Mg	0.219	**0.798**	-0.096	0.035	-0.005
NO3	0.366	**0.589**	0.089	0.305	0.080
pH	0.051	**-0.774**	-0.166	0.147	0.089
K	**0.691**	0.142	-0.077	-0.042	-0.088
Na	**0.875**	0.002	0.223	-0.060	-0.005
Sulfates	**0.695**	0.472	0.160	0.164	0.120
PCE	0.071	0.057	**0.842**	0.095	0.133
TCE	0.075	0.001	**0.896**	-0.062	-0.085
TCM	0.013	-0.030	0.038	0.000	**0.982**
Eigenvalue	3.262	3.098	1.645	1.102	1.037
Explained variance (%)	25.096	23.833	12.652	8.478	7.979
Cumulative (%) of variance	25.096	48.929	61.581	70.059	78.037

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Table 4

Click here to download Table: Table 4 ex6.docx

Table 4 – Multiple comparison of the cluster lithology fractions. Bonferroni's correction, i.e. 0.05/(k-1), was used to evaluate the level of significance.

Gravel	3	5	6	7	13
3		**	n.s.	n.s.	n.s.
5	**		n.s.	**	n.s.
6	n.s.	n.s.		n.s.	n.s.
7	n.s.	**	n.s.		n.s.
13	n.s.	n.s.	n.s.	n.s.	
Silt					
	3	5	6	7	13
3		**	**	**	**
5	**		**	n.s.	**
6	**	**		n.s.	n.s.
7	**	n.s.	n.s.		n.s.
13	**	**	n.s.	n.s.	
Sand					
	3	5	6	7	13
3		n.s.	n.s.	n.s.	n.s.
5	n.s.		n.s.	**	n.s.
6	n.s.	n.s.		n.s.	n.s.
7	n.s.	**	n.s.		**
13	n.s.	n.s.	n.s.	**	
PSC					
	3	5	6	7	13
3		**	**	n.s.	n.s.
5	**		**	**	**
6	**	**		**	**
7	n.s.	**	**		**
13	n.s.	**	**	**	

** Correlation is significant at the P < 0.0125 level (2-tailed).

n.s. Correlation is not significant

Table 5

Click here to download Table: Table 5 ex7.docx

Table 5 – Correlation Analysis (Pearson's correlation coefficients) of the 5 varifactors vs precipitation and year.

Pearson Correlation

	Cluster	F1	F2	F3	F4	F5	precipitation	year
fac1_3	3	1	-0.215**	-0.106**	0.084**	-0.181**	0.060	0.359**
	5	1	-0.185**	-0.448**	0.243**	0.104**	0.087**	0.162**
	6	1	-0.263**	0.019	-0.039*	0.084**	0.037*	-
	7	1	-0.087**	-0.210**	-0.064*	0.270**	0.142**	0.420**
	13	1	0.079**	-0.201**	-0.118**	-0.256**	0.188**	0.225**
fac2_3	3	-0.215**	1	0.336**	-0.061	-0.053	-0.116**	-0.139**
	5	-0.185**	1	-0.054**	-0.060**	0.077**	-0.061**	-0.031**
	6	-0.263**	1	0.037*	0.075**	-0.092**	-0.075**	0.132**
	7	-0.087**	1	0.038	-0.187**	-.129**	-0.150**	-0.225**
	13	0.079**	1	-0.056**	0.010	.010	-0.129**	-0.169**
fac3_3	3	-0.106**	0.336**	1	0.307**	-0.108**	-0.010	-0.021
	5	-0.448**	-0.054**	1	-0.108**	-0.008	0.110**	0.217**
	6	0.019	0.037*	1	0.004	0.284**	0.061**	0.116**
	7	-0.210**	0.038	1	0.043	0.226**	0.047	-0.066*
	13	-0.201**	-0.056**	1	0.158**	0.294**	0.124**	0.163**
fac4_3	3	0.084**	-0.061	0.307**	1	-0.048	0.040	-0.118**
	5	0.243**	-0.060**	-0.108**	1	0.414**	0.030**	-0.191**
	6	-0.039*	0.075**	0.004	1	0.362**	0.076**	-0.020
	7	-0.064*	-0.187**	0.043	1	0.039	-0.059*	-0.070*
	13	-0.118**	0.010	0.158**	1	0.367**	0.031**	-0.237**
fac5_3	3	-0.181**	-0.053	-0.108**	-0.048	1	0.008	-0.053
	5	0.104**	0.077**	-0.008	0.414**	1	-0.005	-0.168**
	6	0.084**	-0.092**	0.284**	0.362**	1	0.088**	0.052**
	7	0.270**	-0.129**	0.226**	0.039	1	0.095**	0.052
	13	-0.256**	0.010	0.294**	0.367**	1	-0.028**	-0.248**
precip	3	0.060	-0.116**	-0.010	0.040	0.008	1	0.084**
	5	0.087**	-0.061**	0.110**	0.030**	-0.005	1	0.214**
	6	0.037*	-0.075**	0.061**	0.076**	0.088**	1	-0.020
	7	0.142**	-0.150**	0.047	-0.059*	0.095**	1	0.302**
	13	0.188**	-0.129**	0.124**	0.031**	-0.028**	1	0.273**
year	3	0.359**	-0.139**	-0.021	-0.118**	-0.053	0.084**	1
	5	0.162**	-0.031**	0.217**	-0.191**	-0.168**	0.214**	1
	6	-	0.132**	0.116**	-0.020	0.052**	-0.020	1
	7	0.420**	-0.225**	-0.066*	-0.070*	0.052	0.302**	1
	13	0.225**	-0.169**	0.163**	-0.237**	-0.248**	0.273**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 1

Click here to download Figure: Fig1AA.docx

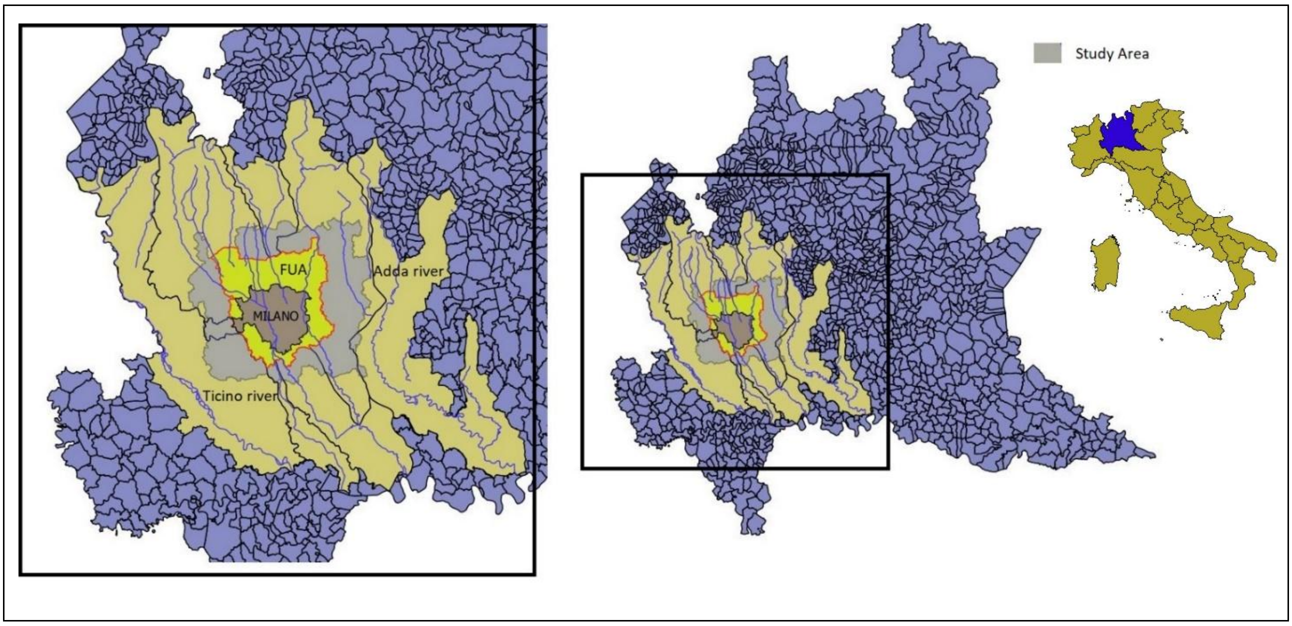


Fig. 1. Study Area: Milan FUA and surrounding municipalities.

Figure 2
Click here to download Figure: Fig 2_rev.docx

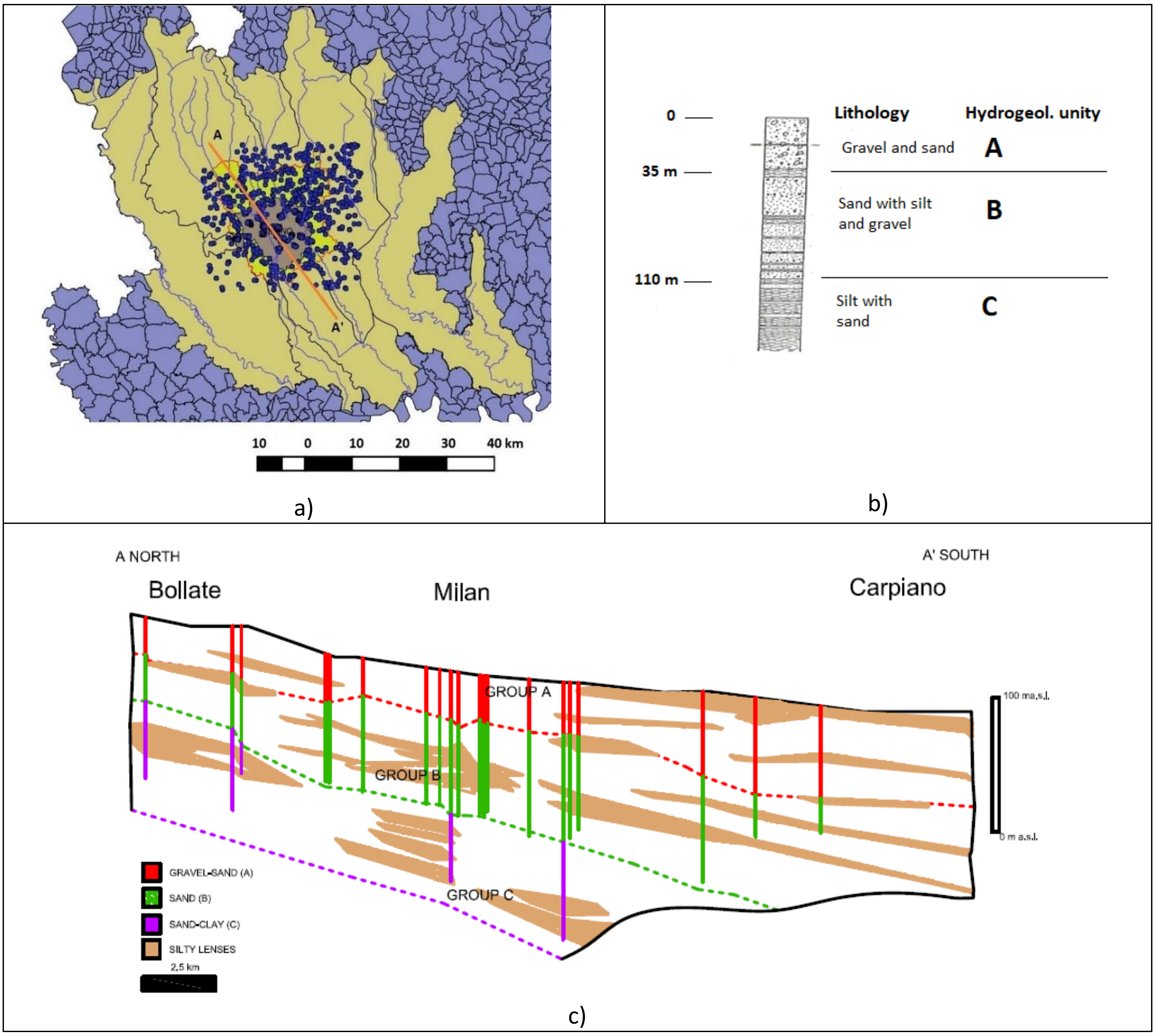


Fig. 2 – Hydrogeological settings of the study area according to Regione Lombardia and ENI (2002). (a) map of the borehole database and the cross section profile through the FUA of Milan; (b) hydrostratigraphic units. (extracted from Foglio Geologico Milano); (c) simplified cross section of the investigated aquifer bodies;

Figure 3

Click here to download Figure: Fig 3.docx

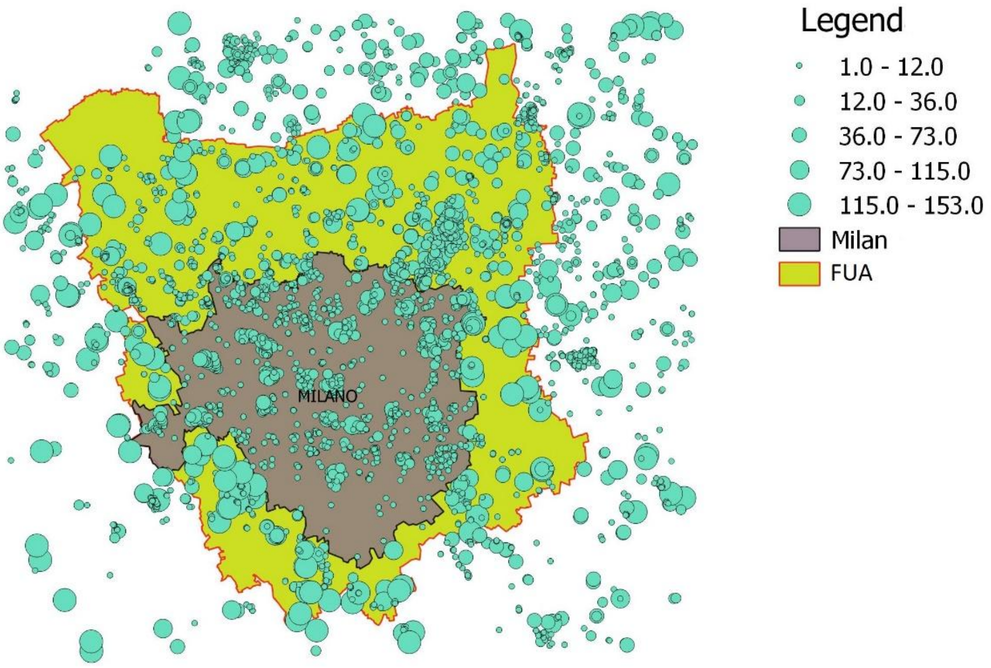


Figure 3: spatial distribution of sampling wells across the study area and the data availability at each sampling well.

Figure 4

Click here to download Figure: Figure 4.docx

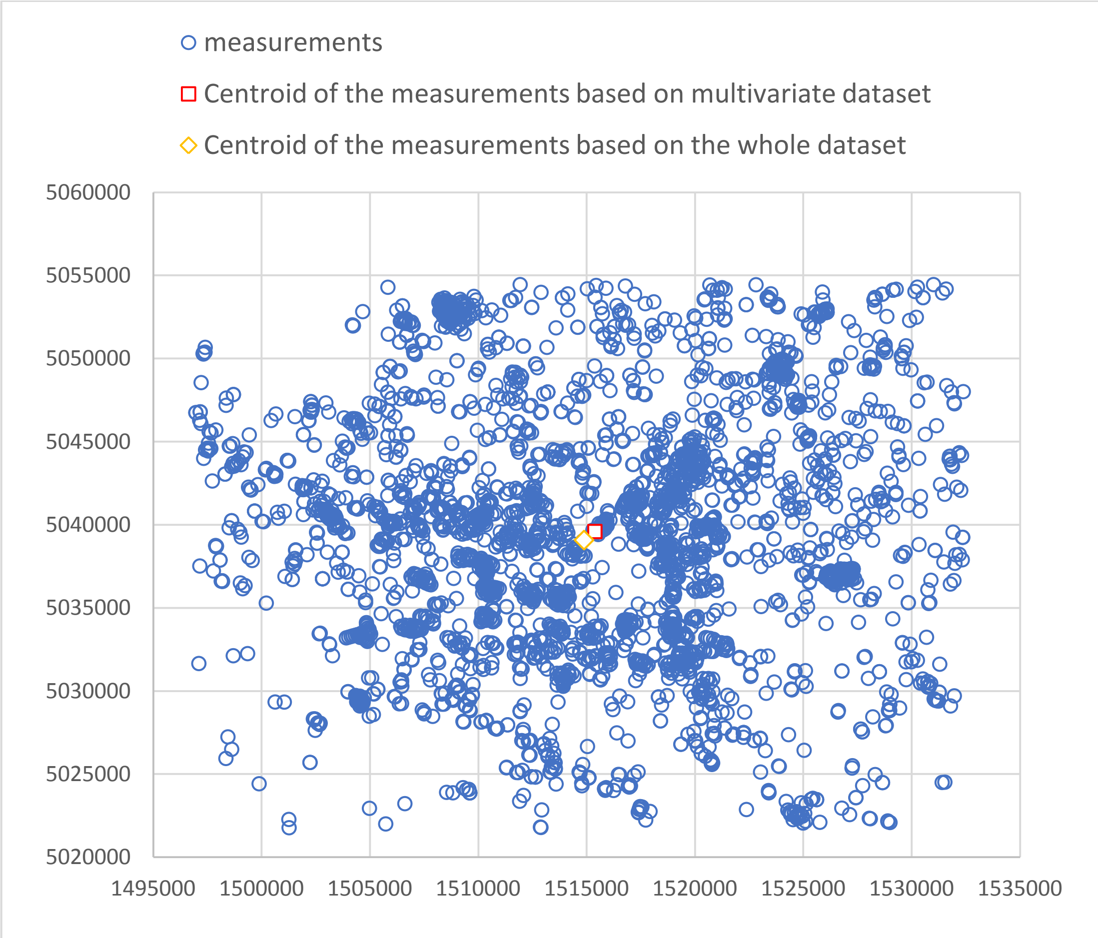


Figure 4 – Map showing the coordinates of the measurements concerning the whole and the multivariate dataset. The centroids weighted on the frequency of measurements are also shown.

Figure 5
Click here to download Figure: Figure 5.docx

a)

Number of Cases in each Cluster

Cluster	1	218
	2	13
	3	1001
	4	1
	5	8613
	6	3902
	7	1358
	8	151
	9	19
	10	59
	11	3
	12	1
	13	10410
	14	2
	15	6
Valid		25757
Missing		34015

b)

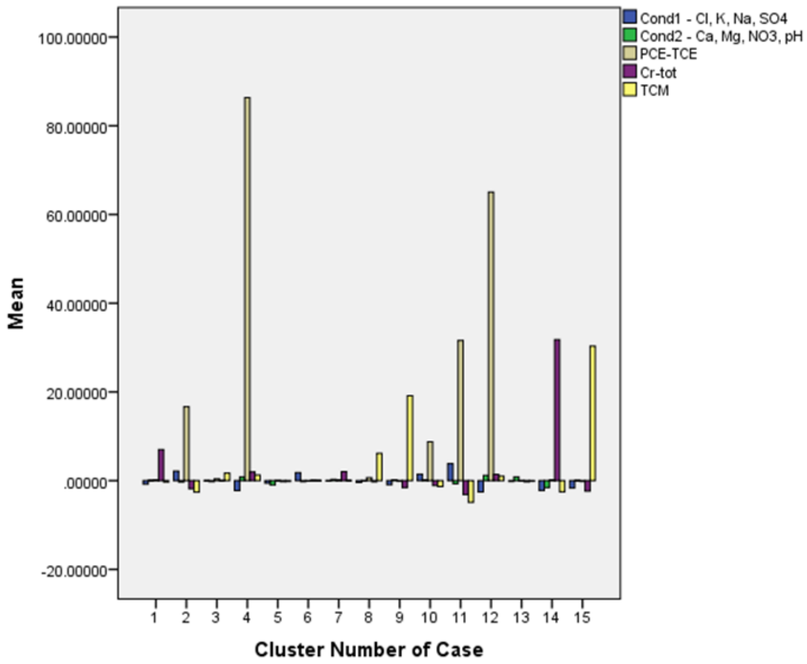


Figure 5 1- a) number of cases included in the K-means clusters; b) cluster characteristics in terms of standardised varifactors (i.e. average 0 and standard deviation 1): bars show the average value of the varifactors within each cluster.

Figure 6
Click here to download Figure: Figure 6.docx

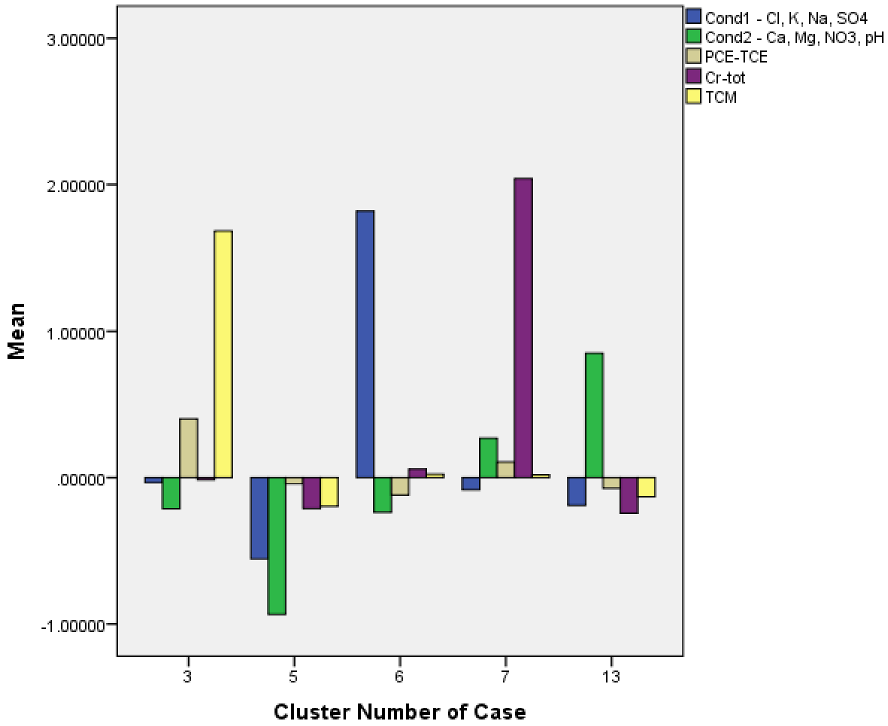


Figura 6 – Varifactor characteristics of the main clusters (> 1% of the total sample): bars show the mean varifactor values within the clusters, while 0 that is the total sample average.

Figure 7
Click here to download Figure: Figure 7_rev.docx

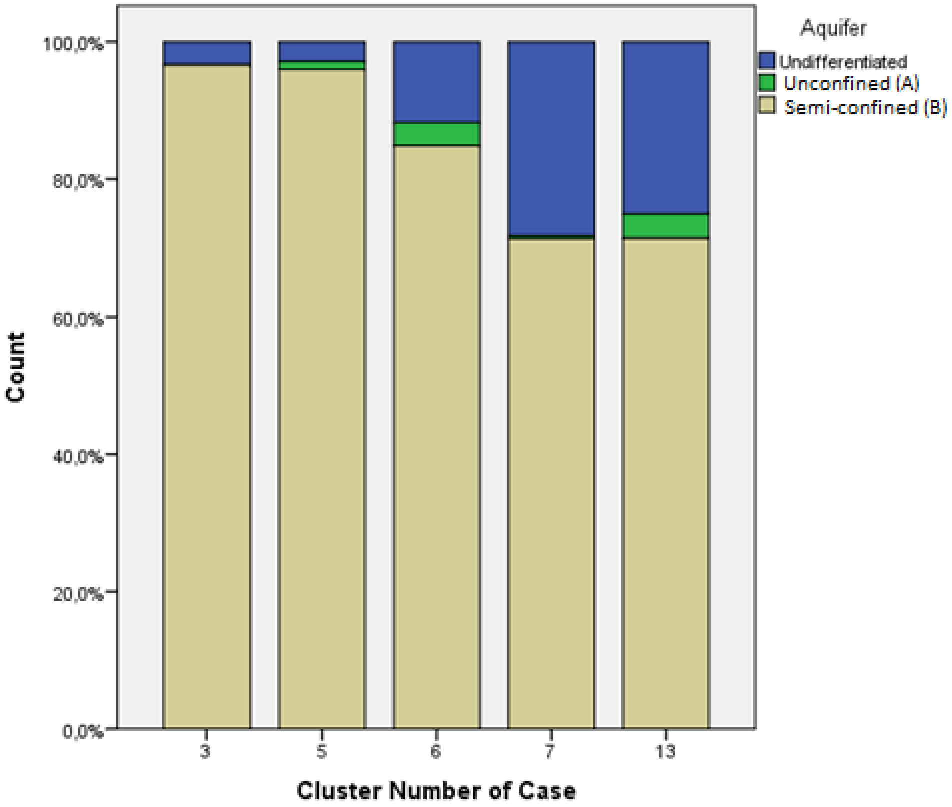
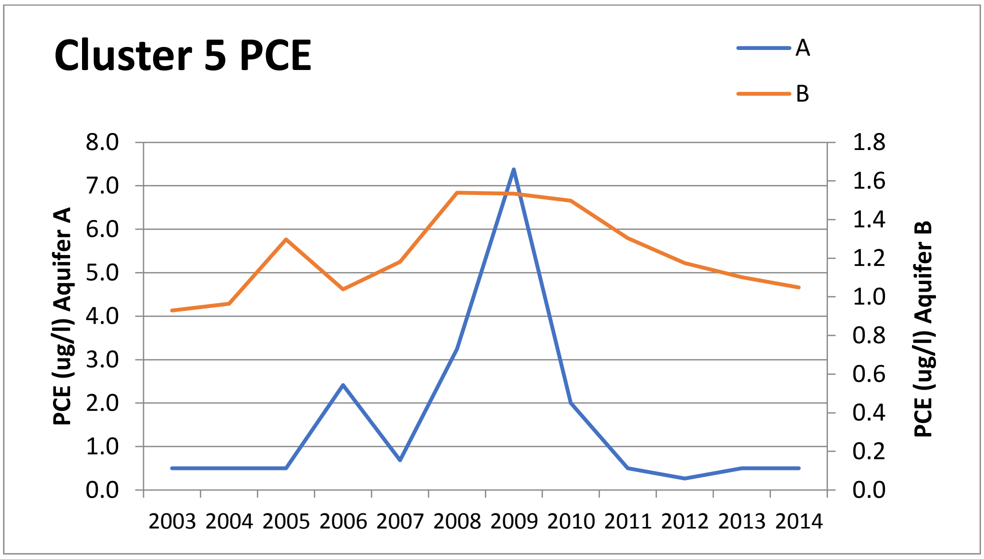
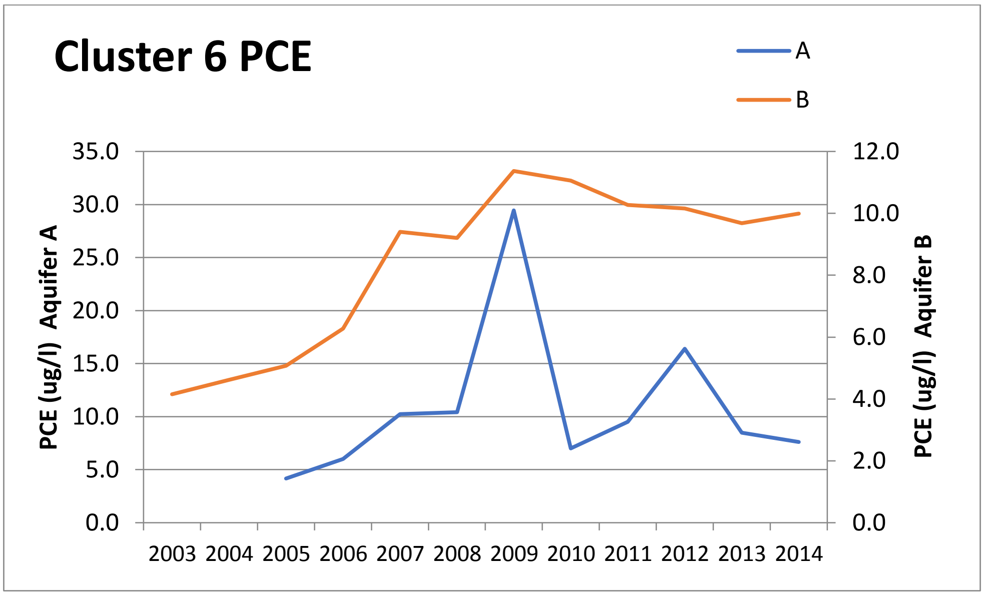


Fig. 7 – Aquifer composition of the five main clusters.

Figure 8

Click here to download Figure: Figure 8.docx





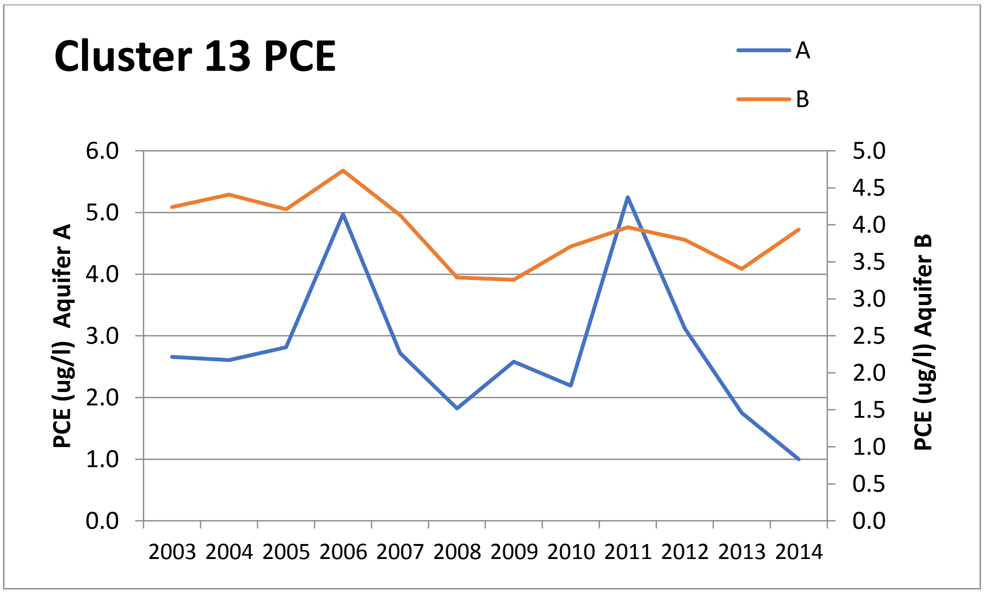


Fig. 8 – PCE water quality profiles in the three main clusters: the annual average concentration profiles are compared between the unconfined (A) and the semiconfined aquifer (B).

Figure 9

Click here to download Figure: Figure 9.docx

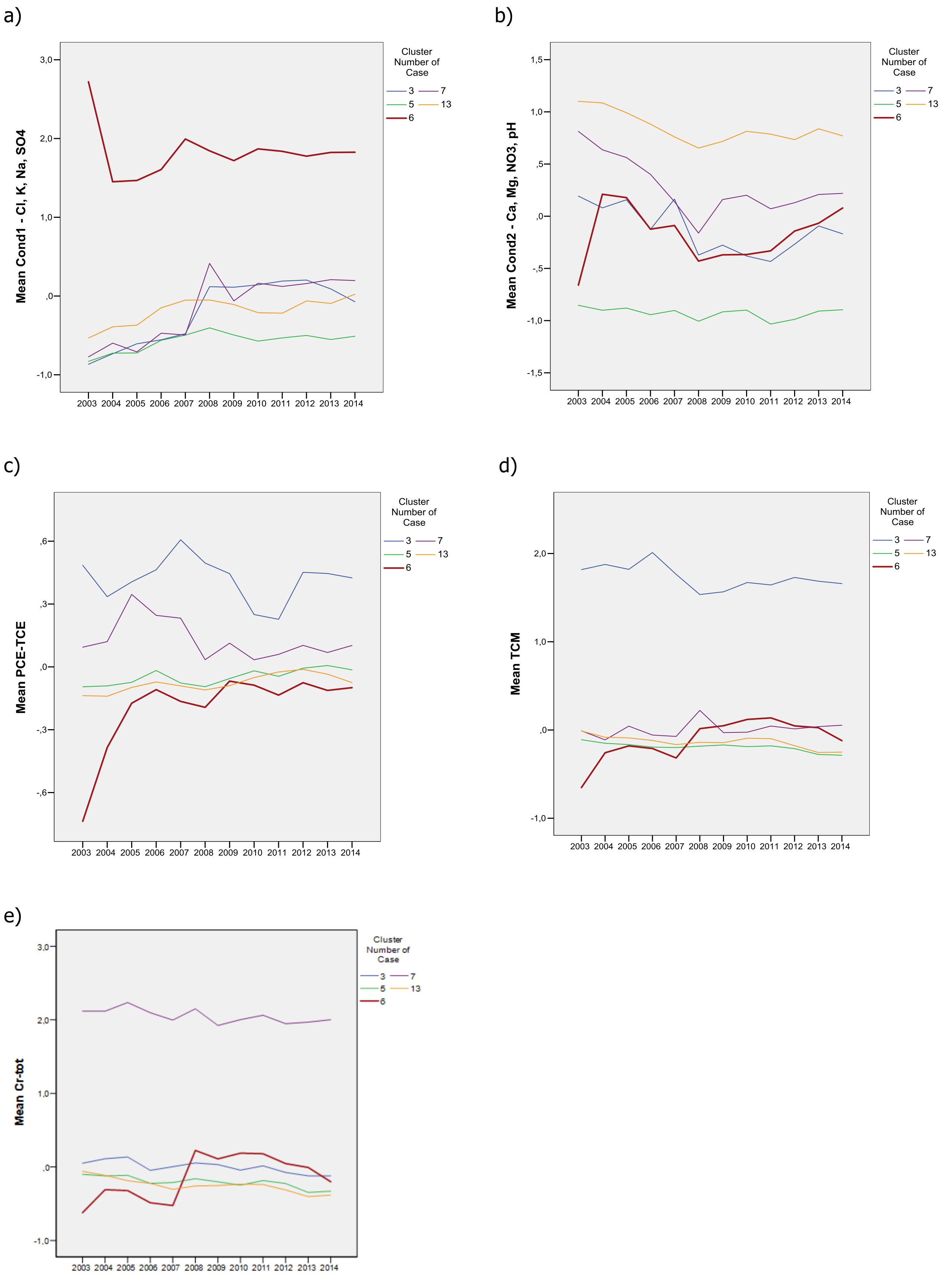


Fig. 9 – Water quality profiles of the five main clusters: a) Conductivity component 1 (Cl, K, Na, SO4); b) conductivity component 2 (Ca, Mg, NO3, pH); c) PCE/TCE component; d) TCM component; e) Total chromium.

Figure 10

Click here to download Figure: Figure 10.docx

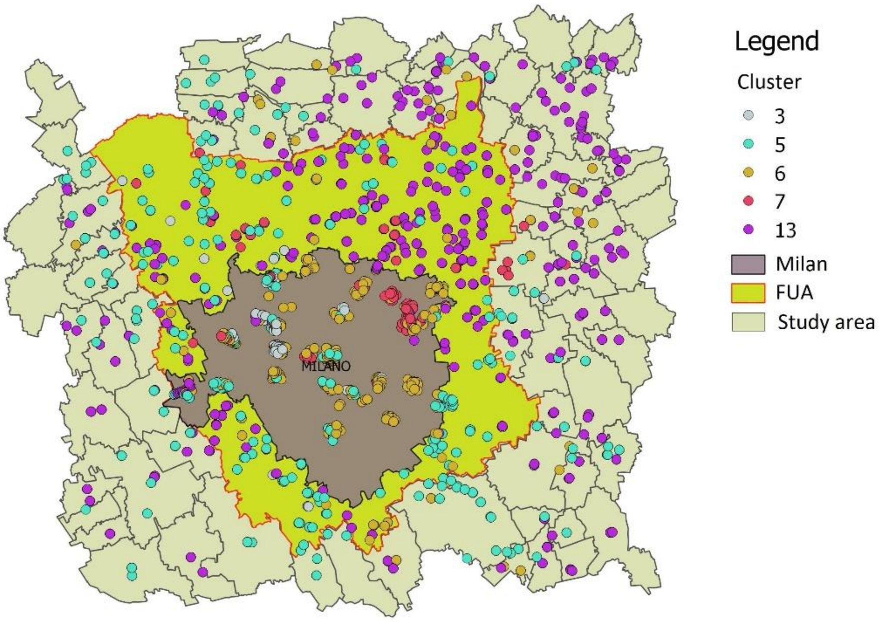


Fig. 10 – Map of the borehole database through the FUA of Milan; Different colours refer to different clusters

Figure 11

Click here to download Figure: Figure 11_rev.docx

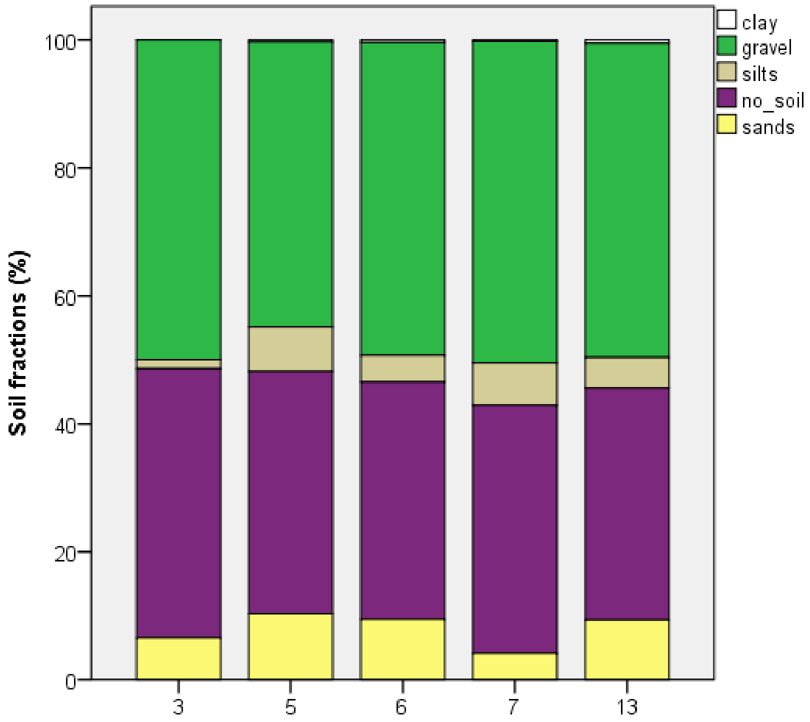


Fig.11 – Lithological composition (%) of the five main clusters.

Figure 12
Click here to download Figure: Figure 12.docx

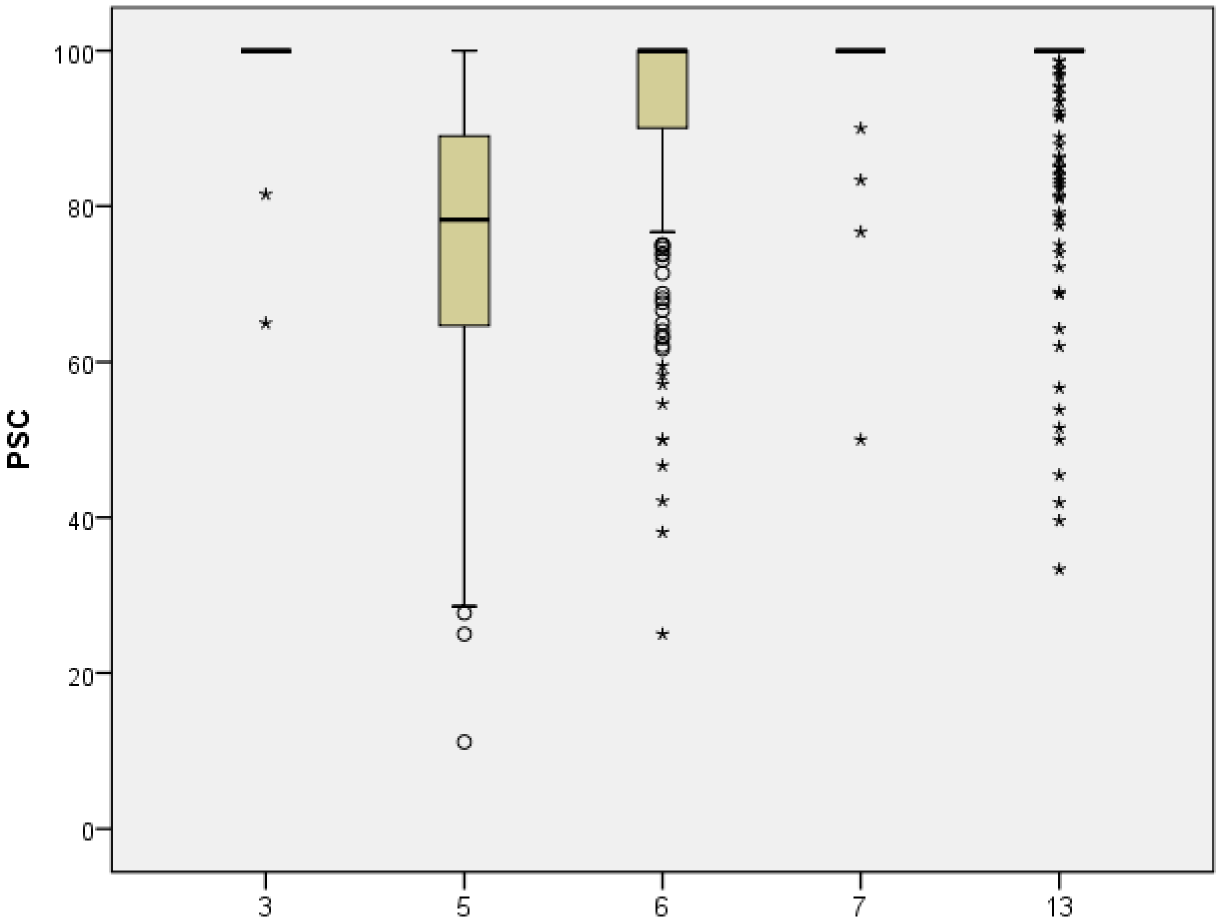


Fig. 12 – Percentage Similarity Index of the Lithological characteristics of five clusters.

Figure 13

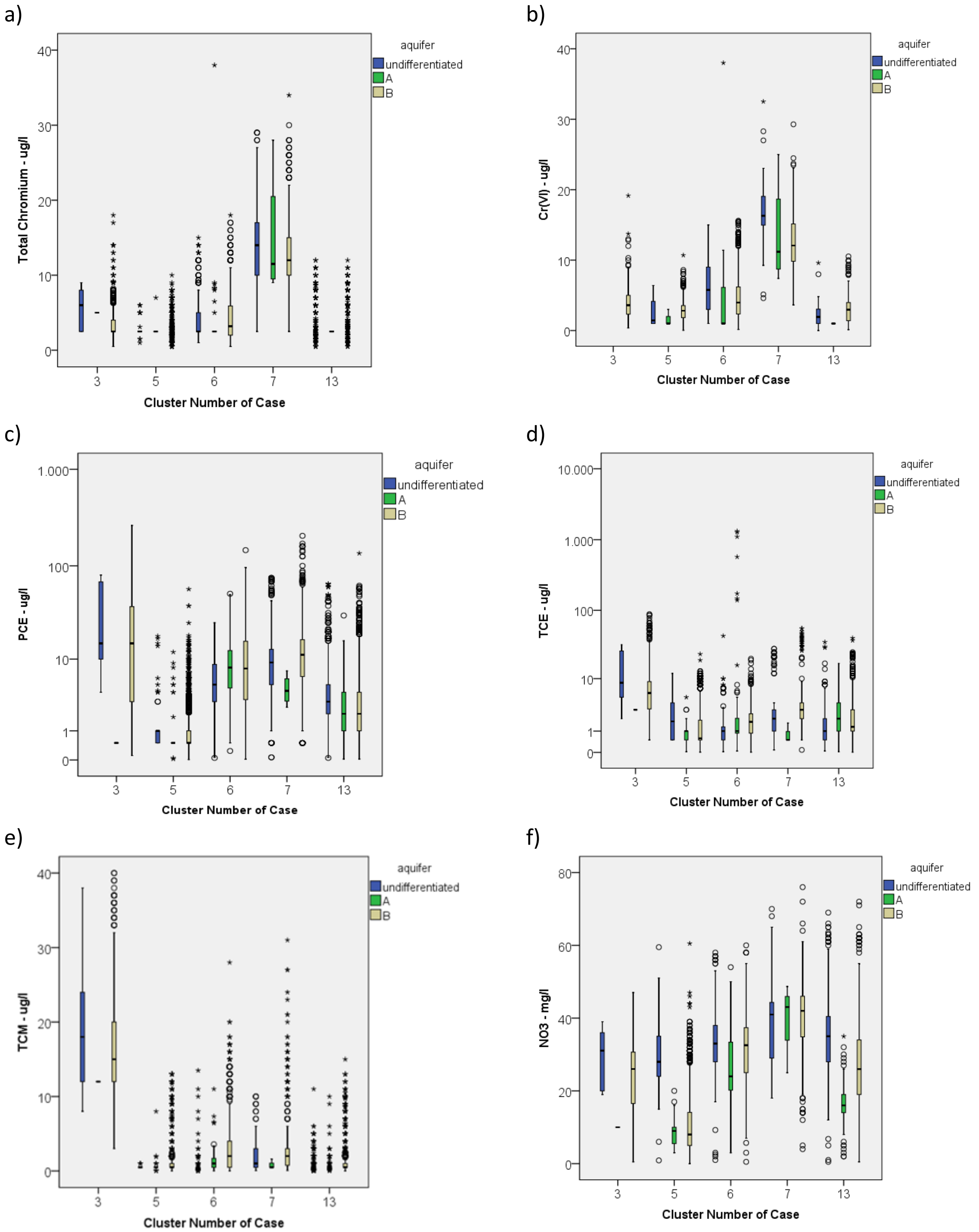


Fig. 13 – Comparison of diffuse pollutant background values among clusters: a) Total-Cr: b) Cr(VI); c) PCE d) TCE; e) TCM; f) NO3.

Figure 14

Click here to download Figure: Figure 14.docx