

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 km<sup>2</sup>, 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

POLITECNICO DI MILANO



Milano, 26<sup>th</sup> 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



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

Azzellino Arianna<sup>a</sup>, Colombo Loris<sup>a</sup>, Lombi Silvia<sup>a</sup>, Marchesi Valeria<sup>b</sup>, Piana Alessandra<sup>b</sup>, Andrea Merri<sup>b</sup>, Alberti Luca<sup>a,c</sup>

<sup>a</sup> Politecnico di Milano, DICA, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy

<sup>b</sup> 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](mailto:arianna.azzellino@polimi.it)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

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

3  
4 Azzellino Arianna<sup>a</sup>, Colombo Loris<sup>a</sup>, Lombi Silvia<sup>a</sup>, Marchesi Valeria<sup>b</sup>, Piana Alessandra<sup>b</sup>, Andrea  
5 Merri<sup>b</sup>, Alberti Luca<sup>a,c</sup>

6  
7 <sup>a</sup> Politecnico di Milano, DICA, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy

8 | <sup>b</sup> ARPA Lombardia, Via Rosellini n° 17, -Milano, 20124, Italy.

9 <sup>c</sup> Scientific responsible of the research project  
10

11 **Abstract (250-300 words)**

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

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

35  
36  
37 **KEYWORDS** max 6:

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

40  
41

42

Formatted: Line spacing: Multiple  
1.15 li

43 **Introduction**

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

56 Urban sprawl phenomenon is a source of many groundwater pollutants coming from industrial  
57 and domestic sources; among those, the ~~well known~~ chlorinated solvents, tetrachloroethylene or  
58 perchloroethylene (PCE), trichloroethylene (TCE) and trichloromethane (TCM), are examples of the  
59 multi-source diffuse pollution (Balderacchi et al., 2013). Besides the environmental behavior,  
60 these compounds have also similar toxicity: TCM is suspected of being a human carcinogen under  
61 unfavorable environmental conditions (Barrlo-Lage et al., 1987; Vogel and McCarty, 1985).  
62 ~~excellent indicators of the groundwater pollution by multi-source diffuse type urban pollution~~  
63 ~~because their environmental behavior and toxicity have been thoroughly studied (Balderacchi et~~  
64 ~~al., 2013). Chlorinated solvents have been in fact widely used in several industries all over Europe,~~  
65 ~~and, not surprisingly, are in fact known as the most prevalent organic contaminants found in~~  
66 ~~groundwater (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroe et al., 2003);~~PCE and TCE

67 may undergo to reductive microbial dechlorination to dichloroethene (DCE) isomers and  
68 vinylchloride (VC), which are highly carcinogenic compounds. Chlorinated solvents have been  
69 widely used in several industries all over Europe, and, being capable to infiltrate rapidly into the  
70 subsurface, causing soil and groundwater pollution (Cortés et al., 2011; Kueper et al., 2003);

71 ~~These compounds have been also found with very high concentrations (from hundreds to~~  
72 ~~thousands of  $\mu\text{g L}^{-1}$ ) in urban groundwater, very often are the most prevalent contaminants in~~  
73 ~~groundwater (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroo et al., 2003). affected by~~  
74 ~~leaching of ponds filled with industrial residues and in groundwater contaminated by~~  
75 ~~petrochemical activities (Pecoraino et al., 2008).~~

76 There are several thousand of PCE/TCE and TCM impacted sites throughout North America,  
77 continental Europe and other industrialized areas of the world (Carter et al., 2012; Cortés et al.,  
78 2011; Hunkeler and Aravena, 2000; Kueper et al., 2003; Parker et al., 2004). Many of these sites  
79 are affected by releases that took places in the first half of the 20<sup>th</sup> century (Kueper et al., 2003).

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

86 Besides the industrial origin, TCM may also be generated as byproduct by chlorination of waters  
87 and wastewaters. All these compounds share features that favor their persistence and areal  
88 diffusion in groundwater. They are typically mobile and recalcitrant (Guilbeault et al., 2005; Rivett  
89 and Feenstra, 2005) and they originate, at the source, as immiscible liquids: at many PCE, TCE and  
90 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)



91 referred to as dense non-aqueous-phase-liquids (DNAPLs), within pore spaces or fractures (Kao  
92 and Lei, 2000; Munholland et al., 2016; Rivett and Feenstra, 2005; U.S.EPA, 1992). Groundwater  
93 flowing through the DNAPL zones dissolves them, generating plumes that commonly achieve  
94 exceptionally large sizes (Mackay and Cherry, 1989; Schwillie, 1984, 1988). ~~Moreover, TCM is also~~  
95 ~~suspected of being a human carcinogen,~~  
96 ~~Solvent DNAPLs may cross low permeability geological layers and under unfavorable~~  
97 ~~environmental conditions (Barrio Lage et al., 1987; Vogel and McCarty, 1985), PCE and TCE may~~  
98 ~~undergo to reductive microbial dechlorination to dichloroethene (DCE) isomers and vinylchloride~~  
99 ~~(VC), which are highly carcinogenic compounds. These compounds are also known to cross low~~  
100 ~~permeability geological layers. Solvent DNAPLs can in fact~~ penetrate into or through most types of  
101 aquitards, even those with very low bulk hydraulic conductivity, due to naturally occurring  
102 preferential pathways (Parker et al., 2004). So, not only phreatic aquifers but also the confined  
103 ones are highly vulnerable toward this type of contamination. DNAPLs can closely interact with  
104 low-permeability deposits, especially those having a significant organic matter content, where  
105 they can be trapped and subsequently released into the aquifers (Chapman et al., 2012) through a  
106 back-flow diffusion process ~~so that these.~~ These deposits may become a secondary source of  
107 pollution, persisting with low concentrations for times estimated up to hundreds of years  
108 (Chapman and Parker, 2005; Parker et al., 2004). Besides chlorinated ~~solvents~~solvents, chromium  
109 being one of the most abundant element in the earth's crust (Emsley, 2011) is also a very frequent  
110 ~~contaminant.~~ urban sprawl contaminant and, as Cr(VI), may cause toxic and genotoxic effects on  
111 the human health (WHO, 1988).  
112 ~~Although Cr(VI) may have a geogenic origin (e.g. Nriagu and Nieboer, 1988; Reinmann and De~~  
113 ~~Caritat, 1998; Izbicki et al., 2008),~~ Although Cr(VI) may have a geogenic origin (e.g. Izbicki et al.,  
114 2008; Nriagu and Nieboer, 1988; Reinmann and De Caritat, 1998), in most cases, the presence of

115 Cr(VI) is anthropogenic (~~Nriagu and Nieboer, 1988; Paine, 2001~~), and the (~~Izbicki et al., 2008;~~  
116 ~~Nriagu and Nieboer, 1988; Paine, 2001~~). The level of Cr(VI) species in soil depends on the pH  
117 ( $\text{CrO}_4^{2-}$  at pH 6.5–14;  $\text{HCrO}_4^-$  and  $\text{Cr}_2\text{O}_7^{2-}$  at pH 0.7–6.5), the redox potential, and the presence of  
118 natural oxidants (e.g., manganese oxide) or reducing agents (e.g., Fe(II), phosphate, sulfide, and  
119 organic matter). ~~The chemical form of chromium in the environment~~ Chromium speciation  
120 determines its mobility and bioavailability. It can be present as a solid mineral in association with  
121 several cations, leading to different chemical species with a large solubility range. The most  
122 insoluble compounds are those containing Pb, Ca, and Ba, whereas dichromates are highly soluble  
123 in soil-water systems (Unceta et al., 2010). ~~Cr(VI) causes toxic and genotoxic effects on the human~~  
124 ~~health (WHO, 1988).~~  
125 ~~The occurrence and the persistence of these pollutants also support the hypothesis that they may~~  
126 ~~be considered as representative of MPS.~~ Remediation of these pollutants can be very difficult ~~to~~  
127 ~~conduct~~. Experience from the past 20 years has in fact demonstrated that sites contaminated by  
128 these pollutants are difficult to investigate and challenging to remediate (Kueper et al., 2003;  
129 Wanner et al., 2012). In spite of innovative technologies being continuously developed to  
130 overcome the technical impracticability of source treatment (e.g. (~~Kueper et al., 2003~~)), ~~very~~  
131 ~~frequently it is not possible to locate and remove the residual pollutant concentrations, so~~  
132 ~~remediation (e.g. Pumping&Treat) is often applied as preventing~~ Kueper et al., 2003), ~~very~~  
133 ~~frequently it is not possible to locate the sources or to remove the residual concentrations. In~~  
134 ~~these situations, remediation (e.g. Pumping&Treat) is often applied only to prevent~~ further  
135 migration of dissolved contamination (Kao and Lei, 2000). ~~Conventional groundwater treatment~~  
136 technologies, such as pump and treat or containment, are in fact able to control contaminant  
137 plumes emanating from DNAPL or other contaminants source zones, ~~however they~~

Formatted: English (United Kingdom)

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

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

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

152 ~~Although clearly stated~~

153 ~~Legislative voids and needs concerning the groundwater diffuse pollution~~

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

160 ~~At National level~~In Italy, the Legislative decree (D.lgs. 152/06 which enforce the WFD) defines the  
161 anthropogenic ~~Diffuse Pollution~~diffuse pollution as the “chemical, physical and biological

162 alteration of environmental matrixes and contaminations determined by diffuse sources and not  
163 linked to a point source”, and it designates Regional authorities to recognize and to enact actions  
164 when such diffuse contamination is identified. ~~Due to this~~Such legislative demand, ~~there is~~ creates  
165 the need of scientific-based tools asto support ~~for recognizing~~the identification of areas affected  
166 by anthropogenic diffuse pollution ~~and identifying, defining~~ proper Diffuse Pollution Background  
167 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  
168 Background Level (NBL) of undesirable elements is proven to be higher than the specific  
169 Groundwater Quality Standard (GQS), NBLs should be assumed as GQSs. No clear procedure is yet  
170 defined instead on how to manage when Diffuse Pollution Background Levels (DPBLs) are higher  
171 than GQSs and make challenging if not unfeasible any remediation strategy, starting from the  
172 DPBLs definition itself. In this respect, it should be reminded that, if several and well established  
173 methodological approaches exist on the identification of NBLs (Matschullat et al., 2000) of  
174 groundwater contaminants: some parametric (e.g. (Carral et al., 1995; Reimann and Filzmoser,  
175 2000; Wendland et al., 2005) and some non-parametric (Hinsby et al., 2008; Molinari et al., 2014,  
176 2012; Muller et al., 2006; Wendland et al., 2008); consolidated methods are still lacking for the  
177 identification of the DPBLs. Aim of this paper is to discuss a multivariate statistical approach to  
178 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  
180 specific Groundwater Quality Standard (GQS), WFD and GWD accept that NBLs are assumed as  
181 GQSs. No clear procedure is yet defined instead on how to manage when Diffuse Pollution  
182 Background Levels (DPBLs) are higher than GQSs and make challenging if not unfeasible any  
183 remediation strategy, starting from the DPBLs definition itself. In this respect, if several and well-  
184 established methodological approaches exist on the identification of NBLs (Matschullat et al.,  
185

186 [2000](#) of groundwater contaminants: some parametric (e.g. Carral et al., 1995; Reimann and  
187 Filzmoser, 2000; Wendland et al., 2005) and some non-parametric (Hinsby et al., 2008; Molinari et  
188 al., 2014, 2012; Muller et al., 2006; Wendland et al., 2008); consolidated methods are still lacking  
189 for the identification of the DPBLs. Aim of this paper is to discuss a multivariate statistical  
190 approach to address the issue of identification of Diffuse Pollution Background Levels.

191

## 192 **Materials and methods**

193

### 194 *Study Area*

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

198

### FIGURE 1

199 The area is about 1600 km<sup>2</sup> wide, and it lies at the center of one of the most urbanized and  
200 industrialized areas in Europe, hosting about 5.6% of the whole Italian population (over 3400000  
201 people, (ISTAT, 2017)). The climate is continental, and the mean annual precipitation is about 960  
202 mm/year. The valleys of the main rivers (Ticino, Adda, Lambro) are deeply incised due to erosion  
203 of post-glacial deposits, leading to several orders of fluvial terraces, with the river bed lower than  
204 the regional groundwater level. For this reason, the main rivers play a dominant draining action  
205 permitting only small natural oscillations in time (Alberti et al., 2016; Alberti et al., 2018b; Giudici  
206 et al., 2007). The Milan – Po plain aquifer system see Fig.2 is composed of three main aquifers  
207 made up of Pliocene - Pleistocene continental sediments overlying marine depositional sequences  
208 (Carcano and Piccin, 2002; Perego et al., 2014). The unconfined (A) aquifer consists of coarse  
209 lithology, mainly gravel with a sandy matrix (gravel sand unit in Fig. 2c). The aquifer, 20–100 m

210 thick, overlays a clayey-silty aquitard. This shows a good continuity in the southern portion of the  
211 study area (South of Milan), whereas (Fig. 2) the aquitard becomes discontinuous and then  
212 disappear moving northward.

213 **FIGURE 2**

214 The underlying semi-confined (B) aquifer is 50 to 150 meter thick and consists of a sequence of  
215 gravel and medium-coarse sand in a sandy matrix with discontinuous confining layers of clay and  
216 silt. The base of the semi-confined aquifer consists of clay and silt layers and locally of  
217 conglomeratic units. The deep confined aquifer (C) consists of sandy lenses within clay and silt  
218 units representing the lower Pleistocene continental-marine transition facies (Colombo et al.,  
219 2018; Francani and Beretta, 1995; Pedretti et al., 2013). Concerning the same area, (Bonomi,  
220 2009) provided a 3D detailed definition of subsoil parameters proposing a data process method  
221 aimed to increase the value of the stratigraphic well logs. De Caro et al., (2017) more recently  
222 have provided a geochemical characterization of the area based on mapping of naturally  
223 controlled species, providing also NBLs for the area as a detailed mapping of contaminant trends  
224 and patterns. According to these authors, CrVI in this area cannot be attributed to geogenic  
225 sources due to the absence of ophiolites and serpentinites which are the main sources of natural  
226 chromium in other Italian regions (e.g. ~~De Giusti, 2003; Rotiroti et al., 2015~~ De Giusti et al., 2003;  
227 Rotiroti et al., 2015).

228  
229 *NBLs vs DPBLs determination*

230 The NBL or Baseline level is defined as “the range of concentration of a given element, isotope or  
231 chemical compound in solution, derived entirely from natural, geological, biological or  
232 atmospheric sources, under conditions not perturbed by anthropogenic activity” (Edmunds and  
233 Shand, 2009). Groundwaters from aquifers that are part of the active water cycle are influenced by

Formatted: Italian (Italy)

234 human activities. Water changes chemistry from the moment it enters the system through rainfall  
235 infiltration, exchanges with surface water bodies or through other sources, until it leaves through  
236 runoff, evaporation or withdrawal (Vázquez-Suñé et al., 2005). Shallow aquifers in fact rarely  
237 reflect true natural concentration levels, whereas deep aquifers may be more likely free from  
238 anthropogenic impacts (Muller et al., 2006). Groundwater status in highly urbanized and farm  
239 areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from  
240 many sources. This makes challenging to determine whether the observed groundwater  
241 conditions are the result of a natural chemical status according to the WFD directive or not  
242 (Wendland et al., 2005). Most of the methods used to determine NBLs rely on the separation of  
243 the natural from anthropogenic components based on some indicator chemical species, such as  
244 NO<sub>3</sub>, Cl or SO<sub>4</sub> (Hinsby et al., 2008; Matschullat et al., 2000; Muller et al., 2006) and the  
245 identification of the NBL as a fixed percentile value (e.g. 90<sup>th</sup> or 97<sup>th</sup>) of the observed distribution  
246 of the indicator concentrations. Another common characteristic of these methods is the univariate  
247 approach since, even when considering a set of parameters, the NBLs are evaluated on the basis of  
248 univariate frequency distributions. Moreover, even though many of the studies dealing with NBLs  
249 determination have also come up with a mapping of the contamination patterns (De Caro et al.,  
250 2017), only very few studies (Busico et al., 2018; Hasenmueller and Criss, 2013; Hwang et al.,  
251 2015) have attempted to determine some kind of anthropogenic DPBLs. One of the most  
252 innovative aspects of this study is the use of a multivariate approach to determine DPBLs.

253

254

255

256

257 *Used Data and pre-processing*

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

### 266 FIGURE 3

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

### 277 Table 1

278 |  
279 Moreover, the study area's lithology and precipitation regime were considered. Particularly the  
280 lithology layer, extracted from the Regione Lombardia geoportal was used  
281 (<http://www.geoportale.regione.lombardia.it/>), which contains the following information: the genetic  
282 classification of surface deposits; the classification of the rocky substratum based on the



283 characteristics of petrographic composition, structure and texture; the main structural features  
284 and layers as the depths of the soils reported in classes (e.g. 0-50 cm; 50-100 cm; 100-200 cm).  
285 Moreover, precipitation and temperature data available at 15 meteorological stations included in  
286 the FUA, were used to estimate the rainfall depurated by evapotranspiration based on the  
287 Thornthwaite method (Thornthwaite and Mather, 1955).  
288 Then the net precipitation monthly values have been interpolated through a kriging technique and  
289 used for the analysis. Moreover, a grid of 1x1 km<sup>2</sup> was used to associate both lithology and  
290 precipitation to spatial units and to clusters. Particularly lithology has been evaluated either in  
291 terms of areal fraction of the different lithological types, either as overall similarity of the cluster.  
292 As similarity metric the Percentage Similarity Index, (i.e. PSC, after Brock, (1977) see Eq.1) was  
293 used:

$$294 \quad PSC = 100 - 0.5 \sum_{i=1}^k |a - b| \quad \text{Eq.1}$$

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

296

297

## 298 *Statistical Analysis*

299 All the statistical computations were made using the statistical package IBM SPSS Statistics 24.0.  
300 Principal Component and Factor Analysis (hereinafter FA, cfr. Afifi et al., (2003)) were performed  
301 based on the correlation matrix of the concentration measurements. Particularly, Factor Analysis  
302 was obtained through the preliminary Principal Components Analysis (hereinafter PCA) which  
303 extracted the eigenvalues and eigenvectors from the covariance matrix of the original variances.  
304 Factor analysis was chosen to reduce the contribution of the less significant parameters within

305 each component, by extracting a new set of varifactors through rotating the axes defined by the  
306 PCA extraction. The Varimax rotation criterion was used to rotate the PCA axes allowing to  
307 maintain the axes orthogonality. The number of factors to be retained was chosen on the basis of  
308 the “*eigenvalue higher than 1*” criterion (i.e. all the factors that explained less than the variance of  
309 one of the original variables were discarded). That allowed to select few factors able to describe  
310 the whole dataset with minimum loss of original information. Moreover, K-means Cluster Analysis  
311 (hereinafter CA, cfr. Afifi et al., (2003)) was used to analyze the similarities among the water  
312 quality profiles at the different monitoring stations, using the Euclidean Distance as distance  
313 metric (see Eq. 2).

$$314 \quad d(x_1, x_2) = \sqrt{\sum_{k=1}^p (x_{ik} - y_{jk})^2} \quad \text{Eq.2}$$

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

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

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

325

## 326 **Results**

327 *Preliminary analysis of multivariate dataset*

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

334 The application of the listwise deletion criterion reduced the size of multivariate dataset to about  
335 60% (N: 25800) of the mean size of whole dataset concerning the investigated parameters (N:  
336 about 42300).

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

340 TABLE 2

341 As it can be observed, in fact, the distribution quartiles of the four pollutants of main interest for  
342 this study are quite comparable in the two situations; mean and standard deviation, being strongly  
343 affected by outliers that are only present in the whole dataset, do not correspond. Outliers in the  
344 whole dataset mostly refer to extremely high values measured at some specific pollution hotspot  
345 sites. In these sites, very often, only the information about pollutants was available, often lacking  
346 data about the other parameters of the multivariate dataset. In consideration of the fact that  
347 multivariate dataset had to be used to determine DPBLs, the absence of extreme outliers in the  
348 multivariate dataset was considered beneficial for the analysis. Moreover, in order to assess  
349 whether the spatial coverage of the multivariate dataset was about the same of the whole dataset  
350 considering the single parameters independently, the centroids of the measurement site  
351 coordinates, weighted by the corresponding frequency of measurements, obtained from either

352 the multivariate or the whole dataset were compared. As it can be observed in Figure 4, the two  
353 centroids are quite close, being the distance between them less than 500 m, so it can be  
354 concluded that the multivariate dataset has about the same spatial coverage of the whole dataset.

355 **FIGURE 4**

356 *Principal Component and Factor Analysis*

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

360 ~~TABLE 3~~

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

363 TABLE 3

364 ~~TABLE 4~~

365 The two sets of loadings are very similar, although in the rotated solution some parameters have  
366 “migrated” from a component to another. PCA solution had in fact a first component constituted  
367 by all the main hydrochemical parameters (i.e. conductivity and the main ions) explaining alone  
368 40% of the total variance, a second component strongly linked to PCE and TCE explaining about  
369 13% of the total variance, and the following components, each explaining less than 10% of the  
370 total variance. Moreover, in the PCA solution there are some parameters, namely Na, pH, Total Cr  
371 and TCM which load on two different components making more difficult the component  
372 interpretation.

373 On the other hand, the Varimax rotated solution (see Table 43) makes the PCA loadings either  
374 large or small on single varifactors, facilitating the interpretation. Particularly, in the rotated  
375 solution the hydrochemical parameters, which in the PCA solution almost all loaded the first

Formatted: Indent: Left: 0.63 cm

376 component, are now loading both the first (i.e. explaining 25% of total variance) and the second  
377 varifactor (i.e. explaining about 24% of total variance); the pollutants Cr and TCM show also a  
378 cleaner distribution across the rotated varifactors, being total chromium and TCM both  
379 respectively loaded only on the fourth (i.e. 8.5% of the total variance) and on the fifth (i.e. 8% of  
380 the total variance) varifactor.

381 In summary the extracted varifactors are the following:

382 *Varifactor 1:* Accounts 25.1% of the total variance and is loaded by Conductivity, Cl, Na, K, SO<sub>4</sub>; this  
383 varifactor represents the conductivity component that is most influenced by anthropic activity  
384 particularly concerning the salinization effect as described later; the high correlation between SO<sub>4</sub>,  
385 Cl and Na has been related to the leakage of agricultural and municipal wastes (~~Sikora et al.~~  
386 ~~1976~~).(Sikora et al., 1976).

387 *Varifactor 2:* Accounts 23.8% of the total variance and is loaded by Conductivity, Ca, Mg, NO<sub>3</sub> and  
388 pH; this varifactor represents the conductivity-hardness component of the groundwater. The  
389 correlation between Ca and NO<sub>3</sub> can be associated with the use of fertilizers such as  
390 NH<sub>4</sub>NO<sub>3</sub>·CaCO<sub>3</sub> (22% N and 33% NCaCO<sub>3</sub>) which is very common in cultivated regions of the study  
391 area (~~Tisdale and Nelson 1975~~).(Tisdale and Nelsson, 1975).

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

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

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

399

400

401 *K-means Cluster Analysis*

402 K-means cluster analysis was applied to the extracted FA varifactors. K was set to 15 and the  
403 analysis was run twice, using the final cluster centroids obtained from the first analysis as initial  
404 seeds for the second. Figure 5.a) shows the 15 clusters' composition in terms of number of cases  
405 included. It can be observed that most of the clusters contain a very small amount of data (i.e. less  
406 than 1% of the total cases) while few clusters, namely cluster 3, 5, 6, 7 and 13, contain more than  
407 98% of the records including most of the background value measurements. Figure 5.b) shows the  
408 cluster characteristics in terms of varifactors.

409

FIGURE 5

410 It can be observed that clusters representing less than 1% of the total measurements are also the  
411 ones characterized by the higher values of the varifactors. Those clusters include in fact  
412 measurements that are outliers with respect of one or more of the considered parameters. Figure  
413 6 shows the characteristics of the five main clusters:

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

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

428 FIGURE 6

429

430 *Cluster interpretation and trend*

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

436 FIGURE 7

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

439 FIGURE 8

440 It can be observed that they present a comparable temporal trend. Such trends have been also  
441 tested through a Spearman's rho rank correlation analysis revealing that the PCE concentration  
442 profiles in the unconfined (A) and semiconfined (B) aquifers are correlated at a 0.05 and a 0.10  
443 significance level ~~(see Table 5)~~. The Spearman's correlation coefficients between A and B aquifers  
444 in the three bigger clusters (i.e. 5, 6 and 13) respectively are: 0.519 (P: 0.084; n: 12), 0.627 (P:  
445 0.039; n: 11) and 0.591 (P: 0.056; n: 11).

446

TABLE 5

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

- 451 - *Cluster 3*: presents an increasing annual trend of the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>; r:  
452 0.359 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO<sub>3</sub>, pH r: - 0.139  
453 P<0.01). No trend is shown by the PCE/TCE and TCM varifactors whereas the fourth  
454 varifactor shows a weak decreasing trend (i.e. Total Cr; r: - 0.118 P<0.01).
- 455 - *Cluster 5*: as far as the conductivity components are concerned, cluster 5 presents a weak  
456 increasing trend of the first (i.e. Cl, Na, K, SO<sub>4</sub>; r: 0.162 P<0.01) and a very weak decreasing  
457 trend of the second varifactor (i.e. Ca, Mg, N-NO<sub>3</sub>; r: -0.03 P<0.01). Concerning the  
458 pollutant components, cluster 5 presents a weak increasing trend of the PCE/TCE varifactor  
459 (r: 0.217 P<0.01) and decreasing trends for the Total chromium and the TCM varifactors  
460 (respectively: r: -0.191 P<0.01; r: -0.168 P<0.01).
- 461 - *Cluster 6*: cluster 6 presents no trend concerning the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>; r:  
462 0.162 P<0.01) and a weak increasing trend of the second varifactor (i.e. Ca, Mg, N-NO<sub>3</sub>; r:  
463 0.132 P<0.01). Moreover, cluster 6 presents weak increasing trends of both the PCE/TCE  
464 varifactor (r: 0.116 P<0.01) and TCM varifactors (r: 0.052 P<0.01). No significant trend is  
465 shown by the Total-chromium varifactor.
- 466 - *Cluster 7*: cluster 7 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>; r:  
467 0.420 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO<sub>3</sub>, pH r: - 0.225  
468 P<0.01). Concurrently, the trend is very weakly decreasing for the PCE/TCE and Total



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

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

#### 476 FIGURE 9

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

#### 479 FIGURE 10

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

#### 485 FIGURE 11

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

#### 489 FIGURE 12

490 Table ~~64~~ shows the multiple comparison test results.

491 TABLE 4

492 ~~TABLE 6~~

Formatted: Indent: Left: 0.63 cm

493 The correlation with net precipitation (i.e. depurated by evapotranspiration) was also  
494 considered. Table 75 shows the correlation analysis of the five varifactors with net  
495 precipitation, cluster by cluster. It can be observed that most of the varifactors presents a  
496 correlation with both the year and net precipitation.

497 TABLE 75  
498

#### 499 *Towards the DPBLs determination*

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

505 FIGURE 13

506 Particularly, clusters 3 and 7 (see Figure 10) have a more localized spatial distribution and show  
507 higher DPBLs in all the aquifers. Particularly, cluster 3 is characterized by solvents concentrations  
508 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  
509 cluster 7 has chromium and nitrates levels that respectively are almost double or very close to the  
510 corresponding GQS (respectively Cr-IV: 5 ug/l and NO<sub>3</sub>: 50 mg/l). On the other hand, clusters 5 and  
511 13 have a more diffuse spatial distribution and tend to have lower background values concerning  
512 all the pollutants. Cluster 6 is instead mostly concentrated within the municipality of Milan and it  
513 shows also well-defined characteristics: chromium levels that are higher than the ones of cluster 5  
514 and 13 but lower than the GQS, solvents and nitrates levels comparable with cluster 7 levels, being  
515 the PCE and TCM levels (TCE levels are higher than the GQS only in the semiconfined aquifer)  
516 higher than the corresponding GQS all over the area. Overall, it is relevant to observe, that both

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

523

## 524 **Discussion**

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

530 Remediation of diffuse pollution to GQS levels, which very often are much lower than the  
531 background levels, is very difficult if not ~~totally~~ unfeasible, and decision-makers, in many situations  
532 concerning Functional Urban Areas, are in fact urged to define new remediation goals, different  
533 from the GQS which are very often determined through a conservative risk assessment process.

534 However, such a procedure is generally conducted by each Regional authority on case-by-case  
535 basis process while a national and European legislation about the management of diffuse pollution  
536 is still lacking. There is no doubt that relying on site-specific risk analysis in a FUAs with large  
537 patterns of diffuse contamination, it substantially increases clean-up associated costs compared to  
538 a very small improvement of groundwater quality. On the other hand, assessing the type of  
539 pollution, whether diffuse or localized, is critical as Public Authorities need to face those  
540 contaminations applying completely different actions, with dramatically different associated costs.

541 If well-established methodological approaches exist for the identification of Natural Background  
542 Levels (Matschullat et al., 2000) of groundwater contaminants, no consolidated method either  
543 exist for identifying the anthropogenic groundwater diffuse pollution. The methodology proposed  
544 in this study is not intrinsically innovative, since multivariate techniques of the kind applied have  
545 been also used in several monitoring studies concerning both surface and groundwaters (Busico et  
546 al., 2018; Gourdol et al., 2013; Marcelli et al., 2010; Mendizabal et al., 2011; Olsen et al., 2012;  
547 Selle et al., 2013; Sheikhy Narany et al., 2014; Yidana et al., 2008; Yu et al., 2014); however, to our  
548 knowledge, this is the first time that these multivariate methods are employed to define DPBLs in  
549 FUA where MPS are present. The strength of this data driven approach is the possibility to identify  
550 hidden patterns in monitoring data (e.g. the hydrosomes as in Mendizabal et al., (2011)) directly  
551 from the dataset. The analysis showed clearly that the Milan FUA is characterized by different  
552 ~~hydro-chemical~~hydrochemical groundwater subsystems with well differentiated water quality  
553 characteristics, which vary in time and space. The five main clusters identified through the analysis  
554 were in fact found to have different temporal profiles and background concentration levels.  
555 Moreover, the clusters were found to be characterized by different lithological characteristics. The  
556 acquired evidence is not enough to support the hypothesis that these clusters are a sort of  
557 hydrosomes as the ones described by (Mendizabal et al., 2011) (e.g. *hydrosome* is defined as a  
558 coherent, three-dimensional unit of groundwater with a specific origin) but it certainly supports  
559 the hypothesis that the FUA should not be seen or managed as an homogeneous groundwater  
560 body. Concerning the hydrogeochemical constitutes, in four of the five main clusters, the analysis  
561 reveals the presence of an increasing trend of the first varifactor (i.e. Na, K, Cl, SO<sub>4</sub>) and a weak  
562 but highly significant (P<0.01) decreasing trend of the second varifactor (i.e. Ca, Mg, NO<sub>3</sub>, pH).  
563 While the increasing trend of the first varifactor is a quite widespread pattern that has been found  
564 in many urbanised and cultivated territories, which to date concerns more than one-third of the

565 world's irrigated land (Abbas et al., 2013; Heuperman et al., 2002; Houk et al., 2006; Li et al.,  
566 2015), the weak decreasing trend of the second varifactor might be due to cation exchange. As  
567 shown in Figure 14a, the dissolution of calcite, dolomite and gypsum seems to be the dominant  
568 reaction in the system as indicated in the plot of  $(Ca^{2+} + Mg^{2+})$  vs.  $(HCO_3^- + SO_4^{2-})$ , which is close to  
569 a 1:1 line.

570 The plot of  $Ca^{2+}$  and  $Mg^{2+}$  concentrations compared with those of  $HCO_3^-$  mostly in cluster 6 and 13  
571 may indicate the existence of an additional  $Ca^{2+}$  and  $Mg^{2+}$  source. Such a result seems to be  
572 consistent with De Caro et al., 2017 findings which revealed higher calcium concentrations in the  
573 northeastern sector of their study area which partially corresponds to ours. It is in fact worthwhile  
574 to observe that those high ratios between  $Ca^{2+}$  and  $Mg^{2+}$  and  $HCO_3^-$  concentrations may not be  
575 attributed to  $HCO_3^-$  depletion because of the existing neutral to slightly alkaline conditions (e.g.  
576 groundwater pH ranges from 6.85 to 8.10, with an average of 7.52 and a standard deviation of  
577 0.25) which do not favour the formation of carbonic acid ( $H_2CO_3$ , see Spears, (1986)). Moreover,  
578 according to Jankowski et al. (1998), if active cation exchange between  $Na^+$  and  $Ca^{2+} + Mg^{2+}$  is  
579 occurring in the aquifer, the slope of the bivariate plot between  $Cl^-$  corrected  $Na^+ + K^+$  and  $(Ca^{2+} +$   
580  $Mg^{2+})$  corrected  $HCO_3^- + SO_4^{2-}$  (see Figure 14c) would be  $-1$  (i.e.,  $y = -x$ ). The slope obtained for  
581 this plot in the study area ( $-1.72$ ) seems to indicate that cation exchange is acting as an important  
582 process.

583 **FIGURE 14**

584 It is also worthwhile to underline that correlations with precipitation are weak, being insignificant  
585 (i.e. cluster 3) or weakly direct concerning the first varifactor, and weakly indirect concerning the  
586 second varifactor. The correlation between  $Na^+$  and  $Cl^-$  in the five main clusters is also different,  
587 being always significant at the  $P < 0.01$  level, and higher than 0.6-0.7 in the smaller clusters (i.e.  
588 0.76, 0.63 and 0.73 respectively for clusters 3, 6 and 7 distributed mostly within the Milan

589 municipality) and lower than 0.5 for the bigger clusters (i.e. 0.38, and 0.55 respectively for clusters  
590 5, and 13). In this respect it has been suggested that the poor correlation ( $r^2 < 0.54$  that would  
591 correspond to an  $r < 0.74$ ) between  $\text{Na}^+$  and  $\text{Cl}^-$  might indicate a possible evaporative  
592 concentration of these ions (Rose, 2002).  $\text{Na}^+/\text{Cl}^-$  ratio has been used, especially in semiarid or arid  
593 regions, to identify the mechanism for acquiring salinity, since the  $\text{Cl}^-$  ion is not affected by soil  
594 retention while the  $\text{Na}^+$  ion is retained (Jalali and Khanlari, 2008; Tiwari and Singh, 2014);  
595 enrichment of  $\text{Na}^+$  may also result from reactions taking place in the clay mineral of the surface  
596 soil horizons. As shown in Figure 14d the clusters, especially the bigger ones, have a marked  
597 fingerprint with respect to the  $\text{Na}^+/\text{Cl}^-$  ratio vs  $\text{Cl}^-$  concentrations being cluster 5 the one of the  
598 highest  $\text{Na}^+$  enrichment and cluster 6 the one with the lowest.

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

607 Correlations with precipitation for the pollutant varifactors are all very weak although significant  
608 suggesting that meteo-climatic variability may also affect the observed patterns. As pointed out  
609 earlier, cluster 3 and 7 both have higher average concentrations of the studied contaminants:  
610 cluster 3 having PCE and TCM levels respectively 2 times and 5 times higher than the whole  
611 sample average, and cluster 7 showing PCE and Total-Cr levels 2 times higher the whole sample  
612 average. The wells belonging to these clusters are mostly inside the boundaries of Milan

613 municipality and are positioned downgradient to a couple of historical industrial district areas  
614 where many brownfields were remediated in the last 20 years. So, despite the effort done in the  
615 past to find the contamination hot spots (i.e. PS) and to remediate soil and groundwater, there are  
616 still traces of the past production activities whose impact can still be detected. This is because  
617 many small unidentifiable sources (i.e. MPS) are still present, but since they are releasing a low  
618 mass of contaminants there is no chance to locate their position and to apply a specific  
619 remediation action. Luckily concentrations in these two clusters are showing a decreasing trend in  
620 time, probably due to Natural Attenuation Processes that could be assessed, monitored and  
621 maybe enhanced to accelerate the achievement of the groundwater quality status. Territorial  
622 authorities need support in the process of identifying such trends, and their awareness of the  
623 existence of portions of the groundwater body that may have different quality characteristics and  
624 trends need to be arisen. For the same reason appropriate DPBLs should be identified for these  
625 portions since only through this process it would be possible to properly manage these areas  
626 applying adequate planning policies and actions (e.g. investigations and monitoring). We showed  
627 that diffuse pollution is present within the FUA with different water quality levels and different  
628 distribution: cluster 13 being the dominant water quality profile representing the 35% of the  
629 boreholes, followed by cluster 5 representing the 31% and by cluster 6 representing the 21%.  
630 Among the contaminants considered in this study only PCE and TCM systematically exceed their  
631 corresponding ~~GO&GQSs~~ with median and interquartile values respectively of 3, 1÷7 µg/l, and 0.5,  
632 0.5÷2 µg/l, while TCE and Cr-VI present only spots of higher contamination, with median and  
633 interquartile values respectively of 1, 0.5÷2.7 µg/l and 3.9, 2.3÷6.6 µg/l. Defining appropriate  
634 DPBLs for these contaminants is both a political and a scientific-technical issue and is beyond the  
635 scope of this paper, however, we the evidence provided by this study about the FUA not being a  
636 whole homogeneous groundwater body strongly support the hypothesis that different DBPLs

637 should be defined within FUA subareas. Moreover, as weak trends were found, revealing a slow  
638 temporal evolution of the contamination (i.e. in average less than 0.06 ug/l per year for the  
639 considered decade). Reliable decennial background levels should be drawn every 5-10 years with  
640 the caveat of periodically revising. Furthermore, being the clusters robust and stable in time, they  
641 allow to overcome the problem of spatial and temporal heterogeneity of measurements which  
642 particularly affects this area and offer a robust data series to define spatially the DPBLs through  
643 geostatistical approaches. Deterministic fate and transport models will also be useful to simulate  
644 the most important plumes deriving from a PS contamination in the FUA, and to identify the  
645 subareas affected by these plumes that should be excluded from the geostatistical interpolation.

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

650

## 651 **Conclusions**

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



661 their corresponding GQs. The clusters described in this study offer a robust knowledge basis for  
662 the drafting a diffuse pollution management plan of the area.

663  
664

**Formatted:** Font: Bold, English  
(United States)

## 665 **Acknowledgements**

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

671  
672  
673

## 674 **References**

- 675 | Abbas, A., Khan, S., Hussain, N., Hanjra, M.A., Akbar, S., 2013. Characterizing soil salinity in  
676 | irrigated agriculture using a remote sensing approach. *Phys. Chem. Earth*.  
677 | doi:10.1016/j.pce.2010.12.004
- 678 | Afifi, A., May, S., Clark, V.A., 2003. Cluster analysis, in: *Computer-Aided Multivariate Analysis*. pp.  
679 | 428–444.
- 680 | Alberti, L., Cantone, M., Colombo, L., Lombi, S., Piana, A., 2016. Numerical modeling of regional  
681 | groundwater flow in the Adda-Ticino Basin: Advances and new results. *Rend. Online Soc.*  
682 | *Geol. Ital.* 41. doi:10.3301/ROL.2016.80
- 683 | Alberti, L., Colombo, L., Formentin, G., 2018. Null-space Monte Carlo particle tracking to assess  
684 | groundwater PCE (Tetrachloroethene) diffuse pollution in north-eastern Milan functional  
685 | urban area. *Sci. Total Environ.* 621. doi:10.1016/j.scitotenv.2017.11.253
- 686 | Alberti, L., Colombo, L., La Licata, I., Cantone, M., 2018. Model calibration using the automatic  
687 | parameter estimation procedure (PEST) of the North-eastern zone of the Milan Functional  
688 | Urban Area (Italy). *Acque Sotter. - Ital. J. Groundw.* 7, 27–38. doi:10.7343/as-2018-336
- 689 | Balderacchi, M., Benoit, P., Cambier, P., Eklo, O.M., Gargini, A., Gemitzi, A., Gurel, M., Kløve, B.,  
690 | Nakic, Z., Predaa, E., Ruzicic, S., Wachniew, P., Trevisan, M., 2013. Groundwater pollution and  
691 | quality monitoring approaches at the European level. *Crit. Rev. Environ. Sci. Technol.*

**Formatted:** Justified

- 692 doi:10.1080/10643389.2011.604259
- 693 | Balderacchi, M., Filippini, M., Gemitzi, A., Kløve, B., Petitta, M., Trevisan, M., Wachniew, P.,  
694 Witczak, S., Gargini, A., 2014. Does groundwater protection in Europe require new EU-wide  
695 environmental quality standards? *Front. Chem.* doi:10.3389/fchem.2014.00032
- 696 | Barro-Lage, G., Parsons, F.Z., Nassar, R.S., 1987. Kinetics of the Depletion of Trichloroethene.  
697 *Environ. Sci. Technol.* 21, 366–370. doi:10.1021/es00158a005
- 698 | Bonomi, T., 2009. Database development and 3D modeling of textural variations in  
699 heterogeneous, unconsolidated aquifer media: Application to the Milan plain. *Comput.  
700 Geosci.* doi:10.1016/j.cageo.2007.09.006
- 701 | Brock, D.A., 1977. Comparison of community similarity indexes. *Water Pollut. Control Fed.* 49,  
702 2488–2494.
- 703 | Busico, G., Cuoco, E., Kazakis, N., Colombani, N., Mastrocicco, M., Tedesco, D., Voudouris, K., 2018.  
704 Multivariate statistical analysis to characterize/discriminate between anthropogenic and  
705 geogenic trace elements occurrence in the Campania Plain, Southern Italy. *Environ. Pollut.*  
706 doi:10.1016/j.envpol.2017.11.053
- 707 | Carcano, C., Piccin, A., 2002. *Padans aquifer geology of the Lombardy Plain.* Firenze.
- 708 | Carral, E., Puente, X., Villares, R., Carballeira, A., 1995. Background heavy metal levels in estuarine  
709 sediments and organisms in Galicia (northwest Spain) as determined by modal analysis. *Sci.  
710 Total Environ.* doi:10.1016/0048-9697(95)04788-3
- 711 | Carter, J.M., Moran, M.J., Zogorski, J.S., Price, C. V., 2012. Factors associated with sources,  
712 transport, and fate of chloroform and three other trihalomethanes in untreated groundwater  
713 used for drinking water. *Environ. Sci. Technol.* doi:10.1021/es301839p
- 714 | Chapman, S.W., Parker, B.L., 2005. Plume persistence due to aquitard back diffusion following  
715 dense nonaqueous phase liquid source removal or isolation. *Water Resour. Res.*  
716 doi:10.1029/2005WR004224
- 717 | Chapman, S.W., Parker, B.L., Sale, T.C., Doner, L.A., 2012. Testing high resolution numerical models  
718 for analysis of contaminant storage and release from low permeability zones. *J. Contam.  
719 Hydrol.* doi:10.1016/j.jconhyd.2012.04.006
- 720 | Colombo, L., Gattinoni, P., Scesi, L., 2018. Stochastic modelling of groundwater flow for hazard  
721 assessment along the underground infrastructures in Milan (northern Italy). *Tunn. Undergr.  
722 Sp. Technol.* 79. doi:10.1016/j.tust.2018.05.007
- 723 | Cortés, A., Puigserver, D., Carmona, J.M., Viladevall, M., 2011. Biological remediation approach  
724 involving soils and groundwaters polluted with chlorinated solvents in a Mediterranean  
725 context. *Recent Adv. Pharm. Sci.* 661, 223–246.
- 726 | De Caro, M., Crosta, G.B., Frattini, P., 2017. Hydrogeochemical characterization and Natural  
727 Background Levels in urbanized areas: Milan Metropolitan area (Northern Italy). *J. Hydrol.*  
728 doi:10.1016/j.jhydrol.2017.02.025
- 729 | De Giusti, F., Dal-Piaz, G.V.D., Massironi, M., Schiavo, A., 2003. Carta geotettonica della Valle

- 730 | ~~d'Aosta (Geotectonic map of d' Aosta Valley).~~ Mem. Sci. Geol. ~~55~~, 129–149.
- 731 | Edmunds, W.M., Shand, P., 2009. Natural Groundwater Quality, Natural Groundwater Quality.  
732 | doi:10.1002/9781444300345
- 733 |
- 734 | Emsley, J., 2011. Nature's Building Blocks: An A-Z Guide to the Elements, Oxford University Press.  
735 | doi:978-0-19-960563-7
- 736 | European Community, 2006. Groundwater directive 2006/118/CE. Directive of the European  
737 | Parliament and of the Council on the Protection of Groundwater Against Pollution and  
738 | Deterioration, ~~OJ L372, 27/12/2006 pp. 19–31. Off. J. Eur. Union 19, 19–31. doi:http://eur-~~  
739 | ~~lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006L0118~~
- 740 | ~~Emsley, J., 2011. Nature's Building Blocks: An A-Z Guide to the Elements, Oxford University Press.~~  
741 | ~~doi:978-0-19-960563-7~~
- 742 | Francani, V., Beretta, G. Pietro, 1995. Preliminary survey about sustainable use of deeper  
743 | groundwater in Milano province. Milan.
- 744 | Fruchter, J., 2002. In-Situ Treatment of Chromium-Contaminated Groundwater. Environ. Sci.  
745 | Technol. doi:10.1021/es022466i
- 746 | Giudici, M., Ponzini, G., Romano, E., Vassena, C., 2007. Some Lessons from Modeling Ground  
747 | Water Flow in the Metropolitan Area of Milano at Different Scales. Mem. Descr. della Cart.  
748 | Geol. d'Italia LXXVI, 207–218.
- 749 | Gourdol, L., Hissler, C., Hoffmann, L., Pfister, L., 2013. On the potential for the Partial Triadic  
750 | Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry. Appl.  
751 | Geochemistry. doi:10.1016/j.apgeochem.2013.10.002
- 752 | Guilbeault, M.A., Parker, B.L., Cherry, J.A., 2005. Mass and flux distributions from DNAPL zones in  
753 | sandy aquifers. Ground Water. doi:10.1111/j.1745-6584.2005.tb02287.x
- 754 | Hasenmueller, E.A., Criss, R.E., 2013. Multiple sources of boron in urban surface waters and  
755 | groundwaters. Sci. Total Environ. doi:10.1016/j.scitotenv.2013.01.001
- 756 | Heuperman, A.F., Kapoor, A.S., Denecke, H.W., 2002. Biodrainage - Principles, Experiences and  
757 | Applications. Rome.
- 758 | Hinsby, K., Condesso de Melo, M.T., Dahl, M., 2008. European case studies supporting the  
759 | derivation of natural background levels and groundwater threshold values for the protection  
760 | of dependent ecosystems and human health. Sci. Total Environ.  
761 | doi:10.1016/j.scitotenv.2008.03.018
- 762 | Houk, E., Frasier, M., Schuck, E., 2006. The agricultural impacts of irrigation induced waterlogging  
763 | and soil salinity in the Arkansas Basin. Agric. Water Manag. 85, 175–183.  
764 | doi:10.1016/j.agwat.2006.04.007
- 765 | Hunkeler, D., Aravena, R., 2000. Determination of compound-specific carbon isotope ratios of  
766 | chlorinated methanes, ethanes, and ethenes in aqueous samples. Environ. Sci. Technol.  
767 | doi:10.1021/es991178s

Formatted: Justified

- 768 | Hwang, H.-H., Panno, S. V., Hackley, K.C., 2015. Sources and Changes in Groundwater Quality with  
769 | Increasing Urbanization, Northeastern Illinois. Environ. Eng. Geosci.  
770 | doi:10.2113/gseegeosci.21.2.75
- 771 | ISTAT, 2017. Annual survey on movement and calculation of the resident population. Rome.
- 772 | Izbicki, J.A., Ball, J.W., Bullen, T.D., Sutley, S.J., 2008. Chromium, chromium isotopes and selected  
773 | trace elements, western Mojave Desert, USA. Appl. Geochemistry.  
774 | doi:10.1016/j.apgeochem.2007.11.015
- 775 | Jalali, M., Khanlari, Z. V., 2008. Major ion chemistry of groundwaters in the Damagh area,  
776 | Hamadan, western Iran. Environ. Geol. doi:10.1007/s00254-007-0795-6
- 777 | Kao, C.M., Lei, S.E., 2000. Using a peat biobarrier to remediate PCE/TCE contaminated aquifers.  
778 | Water Res. doi:10.1016/S0043-1354(99)00213-4
- 779 | Kueper, B.H., Wealhall, G.P., Smith, J.W.N., Leharne, S. a, Lerner, D.N., 2003. An illustrated  
780 | handbook of DNAPL transport and fate in the subsurface, R&D Publication 133.
- 781 | Li, W., Wang, M.-Y., Liu, L.-Y., Yan, Y., 2015. Assessment of Long-Term Evolution of Groundwater  
782 | Hydrochemical Characteristics Using Multiple Approaches: A Case Study in Cangzhou,  
783 | Northern China. Water 7, 1109–1128. doi:10.3390/w7031109
- 784 | Mackay, D.M., Cherry, J.A., 1989. Groundwater contamination: Pump-and-treat remediation.  
785 | Environ. Sci. Technol. doi:10.1021/es00064a001
- 786 | Marcelli, R., Bartolucci, G., Papaioannu, G., De Angelis, G., Lucibello, A., Proietti, E., Margesin, B.,  
787 | Giacomozzi, F., Deborgies, F., 2010. Reliability of RF MEMS switches due to charging effects  
788 | and their circuital modelling, in: Microsystem Technologies. pp. 1111–1118.  
789 | doi:10.1007/s00542-009-1006-z
- 790 | Matschullat, J., Ottenstein, R., Reimann, C., 2000. Geochemical background - Can we calculate it?  
791 | Environ. Geol. doi:10.1007/s002549900084
- 792 | McGuire, T.M., McDade, J.M., Newell, C.J., 2006. Performance of DNAPL source depletion  
793 | technologies at 59 chlorinated solvent-impacted sites. Gr. Water Monit. Remediat.  
794 | doi:10.1111/j.1745-6592.2006.00054.x
- 795 | ~~Meffe, R., de Bustamante, I., 2014. Emerging organic contaminants in surface water and~~  
796 | ~~groundwater: A first overview of the situation in Italy. Sci. Total Environ.~~  
797 | ~~doi:10.1016/j.scitotenv.2014.02.053~~
- 798 | Mendizabal, I., Stuyfzand, P.J., Wiersma, A.P., 2011. Hydrochemical system analysis of public  
799 | supply well fields, to reveal water-quality patterns and define groundwater bodies: The  
800 | Netherlands. Hydrogeol. J. doi:10.1007/s10040-010-0614-0
- 801 | Molinari, A., Chidichimo, F., Straface, S., Guadagnini, A., 2014. Assessment of natural background  
802 | levels in potentially contaminated coastal aquifers. Sci. Total Environ.  
803 | doi:10.1016/j.scitotenv.2013.12.125
- 804 | Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2012. Natural background levels and  
805 | threshold values of chemical species in three large-scale groundwater bodies in Northern

Formatted: Justified

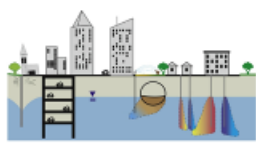
- 806 Italy. Sci. Total Environ. doi:10.1016/j.scitotenv.2012.03.015
- 807 | Muller, D., Blum, A., Hart, A., Hookey, J., Kunkel, R., Scheidleder, A., 2006. Final Report - BRIDGE (
- 808 | Background criteria for the identification of groundwater thresholds ) 13030.
- 809 | Munholland, J.L., Mumford, K.G., Kueper, B.H., 2016. Factors affecting gas migration and
- 810 | contaminant redistribution in heterogeneous porous media subject to electrical resistance
- 811 | heating. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2015.10.011
- 812 | Nriagu, J.O., Nieboer, E., 1988. Chromium in the Natural and Human Environments. John Wiley &
- 813 | Sons.
- 814 | [OECD, 2012. Redefining “Urban”: A New Way to Measure Metropolitan Areas. OECD Publ. 1–9.](#)
- 815 | [doi:10.1787/9789264174108-en](#)
- 816 | Olsen, R.L., Chappell, R.W., Loftis, J.C., 2012. Water quality sample collection, data treatment and
- 817 | results presentation for principal components analysis e literature review and Illinois River
- 818 | watershed case study. Water Res.
- 819 | Paine, A.J., 2001. Mechanisms of chromium toxicity, carcinogenicity and allergenicity: Review of
- 820 | the literature from 1985 to 2000. Hum. Exp. Toxicol. doi:10.1191/096032701682693062
- 821 | Parker, B.L., Cherry, J.A., Chapman, S.W., 2004. Field study of TCE diffusion profiles below DNAPL
- 822 | to assess aquitard integrity. J. Contam. Hydrol. doi:10.1016/j.jconhyd.2004.02.011
- 823 | Pecoraino, G., Scalici, L., Avellone, G., Ceraulo, L., Favara, R., Candela, E.G., Provenzano, M.C.,
- 824 | Scaletta, C., 2008. Distribution of volatile organic compounds in Sicilian groundwaters
- 825 | analysed by head space-solid phase micro extraction coupled with gas chromatography mass
- 826 | spectrometry (SPME/GC/MS). Water Res. doi:10.1016/j.watres.2008.07.022
- 827 | Pedretti, D., Masetti, M., Beretta, G.P., Vitiello, M., 2013. A revised conceptual model to reproduce
- 828 | the distribution of chlorinated solvents in the Rho Aquifer (Italy). Groundw. Monit. Remediat.
- 829 | doi:10.1111/gwmmr.12017
- 830 | Perego, R., Bonomi, T., Fumagalli, M.L., Benastini, V., Aghib, F., Rotiroti, M., Cavallin, A., 2014. 3D
- 831 | reconstruction of the multi-layer aquifer in a Po Plain area. Rend. Online Soc. Geol. Ital. 30,
- 832 | 41–44. doi:10.3301/ROL.2014.09
- 833 | Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: Death
- 834 | of a myth. Consequences for the statistical treatment of geochemical and environmental
- 835 | data. Environ. Geol. doi:10.1007/s002549900081
- 836 | Reinmann, C., De Caritat, P., 1998. Chemical Elements in the Environment. sPRINGER-VERLAG,
- 837 | Berlin.
- 838 | Rivett, M.O., Feenstra, S., 2005. Dissolution of an emplaced source of DNAPL in a natural aquifer
- 839 | setting. Environ. Sci. Technol. doi:10.1021/es040016f
- 840 | Rose, S., 2002. Comparative major ion geochemistry of Piedmont streams in the Atlanta, Georgia
- 841 | region: Possible effects of urbanization. Environ. Geol. doi:10.1007/s00254-002-0545-8
- 842 | Rotiroti, M., ~~DiMauro~~ [Di Mauro](#), B., [Bonomi, T.](#), Fumagalli, [M.L.](#), ~~Bonomi, T.~~ 2015. COMPSEC, a new
- 843 | tool to derive natural background levels by the component separation ~~approach~~:

Formatted: Justified

- 844 | ~~application approach~~ : Application in two different hydrogeological contexts in northern Italy-  
845 | Journal of Geochemical Exploration 158, 44–54 COMPSEC , a new tool to derive natural  
846 | background levels by the component separation. doi:10.1016/j.gexplo.2015.06.017
- 847 | Schwille, F. (1984), 1984. Migration of Organic Fluids Immiscible with Water in the unsaturated  
848 | zone 2–3.
- 849 | Schwille, F. (1988), 1988. Dense Chlorinated Solvents in Porous and Fractured Media: Model  
850 | Experiments, translated from the German by J. F. Pankow.
- 851 | Selle, B., Schwientek, M., Lischeid, G., 2013. Understanding processes governing water quality in  
852 | catchments using principal component scores. J. Hydrol. doi:10.1016/j.jhydrol.2013.01.030
- 853 | Sheikhy Narany, T., Ramli, M.F., Aris, A.Z., Sulaiman, W.N.A., Fakharian, K., 2014. Spatiotemporal  
854 | variation of groundwater quality using integrated multivariate statistical and geostatistical  
855 | approaches in Amol-Babol Plain, Iran. Environ. Monit. Assess. doi:10.1007/s10661-014-3820-  
856 | 8
- 857 | Sikora ~~LG, L., Bent MG, M., Corey RB, R., Keeney DR (, D., 1976).~~ Septic nitrogen and phosphorus  
858 | removal test system. Groundwater 14/5, 309–314.
- 859 | Spears, D.A., 1986. Mineralogical control of the chemical evolution of groundwter. Solute Process.  
860 | 512.
- 861 | Stevenazzi, S., Masetti, M., Nghiem, S. V., Sorichetta, A., 2015. Groundwater vulnerability maps  
862 | derived from a time-dependent method using satellite scatterometer data. Hydrogeol. J. 23,  
863 | 631–647. doi:10.1007/s10040-015-1236-3
- 864 | Stroo, H.F., Unger, M., Ward, C.H., Kavanaugh, M.C., Vogel, C., Leeson, A., Marqusee, J.A., Smith,  
865 | B.P., 2003. Peer Reviewed: Remediating Chlorinated Solvent Source Zones. Environ. Sci.  
866 | Technol. 37, 224A–230A. doi:10.1021/es032488k
- 867 | Thornthwaite, C.W., Mather, J.R., 1955. The water balance. Publ. Climatol. 8, 1–104.
- 868 | ~~Tisdal~~Tisdale, S., Nelsson, W ~~(, 1975).~~ Soil fertility and fertilizers. Macmillan Publishing Co Inc,  
869 | New ~~York, p 240~~Jork.
- 870 | Tiwari, A.K., Singh, A.K., 2014. Hydrogeochemical investigation and groundwater quality  
871 | assessment of Pratapgarh district, Uttar Pradesh. J. Geol. Soc. India. doi:10.1007/s12594-014-  
872 | 0045-y
- 873 | U.S.EPA, 1992. Dense Non Aqueous Phase Liquids - A Workshop Summary. Dallas.
- 874 | Unceta, N., Séby, F., Malherbe, J., Donard, O.F.X., 2010. Chromium speciation in solid matrices and  
875 | regulation: A review. Anal. Bioanal. Chem. doi:10.1007/s00216-009-3417-1
- 876 | Vázquez-Suñé, E., Sánchez-Vila, X., Carrera, J., 2005. Introductory review of specific factors  
877 | influencing urban groundwater, an emerging branch of hydrogeology, with reference to  
878 | Barcelona, Spain. Hydrogeol. J. doi:10.1007/s10040-004-0360-2
- 879 | Vogel, T.M., McCarty, P.L., 1985. Biotransformation of tetrachloroethylene to trichloroethylene,  
880 | dichloroethylene, vinyl chloride, and carbon dioxide under methanogenic conditions. Appl.  
881 | Environ. Microbiol.

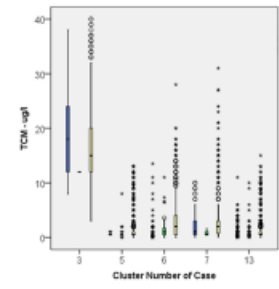
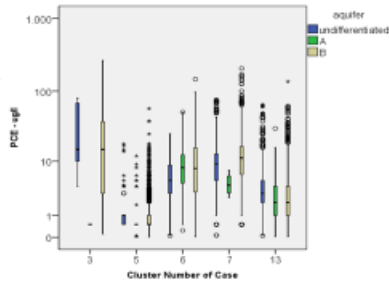
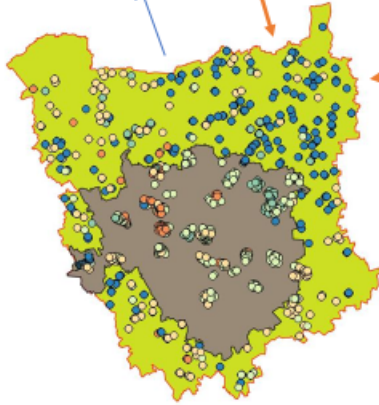
- 882 | Wanner, C., Zink, S., Eggenberger, U., Mäder, U., 2012. Assessing the Cr(VI) reduction efficiency of  
883 a permeable reactive barrier using Cr isotope measurements and 2D reactive transport  
884 modeling. *J. Contam. Hydrol.* doi:10.1016/j.jconhyd.2012.01.007
- 885 | Wendland, F., Blum, A., Coetsiers, M., Gorova, R., Griffioen, J., Grima, J., Hinsby, K., Kunkel, R.,  
886 Marandi, A., Melo, T., Panagopoulos, A., Pauwels, H., Ruisi, M., Traversa, P., Vermooten,  
887 J.S.A., Walraevens, K., 2008. European aquifer typology: A practical framework for an  
888 overview of major groundwater composition at European scale. *Environ. Geol.*  
889 doi:10.1007/s00254-007-0966-5
- 890 | Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H.J., Wolter, R., 2005. A procedure to  
891 define natural groundwater conditions of groundwater bodies in Germany. *Water Sci.*  
892 *Technol.*
- 893 | WHO, 1988. *Environmental Health Criteria*. Geneve.
- 894 | Yidana, S.M., Ophori, D., Banoeng-Yakubo, B., 2008. A multivariate statistical analysis of surface  
895 water chemistry data-The Ankobra Basin, Ghana. *J. Environ. Manage.* 86, 80–87.  
896 doi:10.1016/j.jenvman.2006.11.023
- 897 | Yu, Y., Song, X., Zhang, Y., Zheng, F., Liang, J., Liu, L., 2014. Identifying spatio-temporal variation  
898 and controlling factors of chemistry in groundwater and river water recharged by reclaimed  
899 water at Huai River, North China. *Stoch. Environ. Res. Risk Assess.* doi:10.1007/s00477-013-  
900 0803-1
- 901
- 902

# Graphical abstract

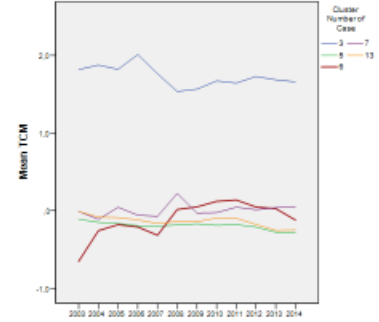
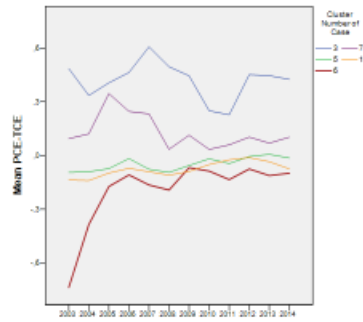


**Spatial Clusters with different Diffuse Pollution Background levels**

Milan FUA



**Clusters with different temporal trend profiles**





## 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

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

3

4 Azzellino Arianna<sup>a</sup>, Colombo Loris<sup>a</sup>, Lombi Silvia<sup>a</sup>, Marchesi Valeria<sup>b</sup>, Piana Alessandra<sup>b</sup>, Andrea  
5 Merri<sup>b</sup>, Alberti Luca<sup>a, c</sup>

6

7 <sup>a</sup> Politecnico di Milano, DICA, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy

8 <sup>b</sup> ARPA Lombardia, Via Rosellini n° 17, Milano, 20124, Italy.

9 <sup>c</sup> Scientific responsible of the research project

10

11 **Abstract (250-300 words)**

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

35

36

37 **KEYWORDS** max 6:

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

40

41

42

## 43 **Introduction**

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

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

67 Feenstra, 2005; Stroo et al., 2003). There are several thousand of PCE/TCE and TCM impacted sites  
68 throughout North America, continental Europe and other industrialized areas of the world (Carter  
69 et al., 2012; Cortés et al., 2011; Hunkeler and Aravena, 2000; Kueper et al., 2003; Parker et al.,  
70 2004). Many of these sites are affected by releases that took places in the first half of the 20<sup>th</sup>  
71 century (Kueper et al., 2003). Very high concentrations (from hundreds to thousands of  $\mu\text{g L}^{-1}$ ) of  
72 these compounds have been found in groundwaters affected by leaching of ponds filled with  
73 industrial residues and in groundwater contaminated by petrochemical activities (Pecoraino et al.,  
74 2008). Besides the industrial origin, TCM may also be generated as byproduct by chlorination of  
75 waters and wastewaters. All these compounds share features that favor their persistence and  
76 areal diffusion in groundwater. They are typically mobile and recalcitrant (Guilbeault et al., 2005;  
77 Rivett and Feenstra, 2005) and they originate, at the source, as immiscible liquids: at many PCE,  
78 TCE and TCM spill sites, residual amounts of these compounds persist in a pure liquid phase,  
79 commonly referred to as dense non-aqueous-phase-liquids (DNAPLs), within pore spaces or  
80 fractures (Kao and Lei, 2000; Munholland et al., 2016; Rivett and Feenstra, 2005; U.S.EPA, 1992).  
81 Groundwater flowing through the DNAPL zones dissolves them, generating plumes that commonly  
82 achieve exceptionally large sizes (Mackay and Cherry, 1989; Schwille, 1984, 1988).  
83 Solvent DNAPLs may cross low permeability geological layers and penetrate into or through most  
84 types of aquitards, even those with very low bulk hydraulic conductivity, due to naturally occurring  
85 preferential pathways (Parker et al., 2004). So, not only phreatic aquifers but also the confined  
86 ones are highly vulnerable toward this type of contamination. DNAPLs can closely interact with  
87 low-permeability deposits, especially those having a significant organic matter content, where  
88 they can be trapped and subsequently released into the aquifers (Chapman et al., 2012) through a  
89 back-flow diffusion process. These deposits may become a secondary source of pollution,  
90 persisting with low concentrations for times estimated up to hundreds of years (Chapman and

91 Parker, 2005; Parker et al., 2004). Besides chlorinated solvents, chromium being one of the most  
92 abundant element in the earth's crust (Emsley, 2011) is also a very frequent urban sprawl  
93 contaminant and, as Cr(VI), may cause toxic and genotoxic effects on the human health (WHO,  
94 1988).

95 Although Cr(VI) may have a geogenic origin (e.g. Izbicki et al., 2008; Nriagu and Nieboer, 1988;  
96 Reinmann and De Caritat, 1998), in most cases, the presence of Cr(VI) is anthropogenic (Izbicki et  
97 al., 2008; Nriagu and Nieboer, 1988; Paine, 2001). The level of Cr(VI) species in soil depends on the  
98 pH ( $\text{CrO}_4^{2-}$  at pH 6.5–14;  $\text{HCrO}_4^-$  and  $\text{Cr}_2\text{O}_7^{2-}$  at pH 0.7–6.5), the redox potential, and the presence  
99 of natural oxidants (e.g., manganese oxide) or reducing agents (e.g., Fe(II), phosphate, sulfide, and  
100 organic matter). Chromium speciation determines its mobility and bioavailability. It can be present  
101 as a solid mineral in association with several cations, leading to different chemical species with a  
102 large solubility range. The most insoluble compounds are those containing Pb, Ca, and Ba, whereas  
103 dichromates are highly soluble in soil-water systems (Unceta et al., 2010).

104 Remediation of these pollutants can be very difficult. Experience from the past 20 years has in fact  
105 demonstrated that sites contaminated by these pollutants are difficult to investigate and  
106 challenging to remediate (Kueper et al., 2003; Wanner et al., 2012). In spite of innovative  
107 technologies being continuously developed to overcome the technical impracticability of source  
108 treatment (e.g. Kueper et al., 2003), very frequently it is not possible to locate the sources or to  
109 remove the residual concentrations. In these situations, remediation (e.g. Pumping&Treat) is often  
110 applied only to prevent further migration of dissolved contamination (Kao and Lei, 2000).

111 Conventional groundwater treatment technologies, such as pump and treat or containment, are in  
112 fact able to control contaminant plumes emanating from DNAPL or other contaminants source  
113 zones, nevertheless involving extended operating periods (decades) and potentially high life cycle  
114 costs (Fruchter, 2002; McGuire et al., 2006).

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

121

### 122 *Legislative voids and needs concerning the groundwater diffuse pollution*

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

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

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

139 Background Levels (DPBLs) are higher than GQs and make challenging if not unfeasible any  
140 remediation strategy, starting from the DPBLs definition itself. In this respect, if several and well-  
141 established methodological approaches exist on the identification of NBLs (Matschullat et al.,  
142 2000) of groundwater contaminants: some parametric (e.g. Carral et al., 1995; Reimann and  
143 Filzmoser, 2000; Wendland et al., 2005) and some non-parametric (Hinsby et al., 2008; Molinari et  
144 al., 2014, 2012; Muller et al., 2006; Wendland et al., 2008); consolidated methods are still lacking  
145 for the identification of the DPBLs. Aim of this paper is to discuss a multivariate statistical  
146 approach to address the issue of identification of Diffuse Pollution Background Levels.

147

## 148 **Materials and methods**

149

### 150 *Study Area*

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

154

#### FIGURE 1

155 The area is about 1600 km<sup>2</sup> wide, and it lies at the center of one of the most urbanized and  
156 industrialized areas in Europe, hosting about 5.6% of the whole Italian population (over 3400000  
157 people, (ISTAT, 2017). The climate is continental, and the mean annual precipitation is about 960  
158 mm/year. The valleys of the main rivers (Ticino, Adda, Lambro) are deeply incised due to erosion  
159 of post-glacial deposits, leading to several orders of fluvial terraces, with the river bed lower than  
160 the regional groundwater level. For this reason, the main rivers play a dominant draining action  
161 permitting only small natural oscillations in time (Alberti et al., 2016; Alberti et al., 2018b; Giudici  
162 et al., 2007). The Milan – Po plain aquifer system see Fig.2 is composed of three main aquifers

163 made up of Pliocene - Pleistocene continental sediments overlying marine depositional sequences  
164 (Carcano and Piccin, 2002; Perego et al., 2014). The unconfined (A) aquifer consists of coarse  
165 lithology, mainly gravel with a sandy matrix (gravel sand unit in Fig. 2c). The aquifer, 20–100 m  
166 thick, overlays a clayey-silty aquitard. This shows a good continuity in the southern portion of the  
167 study area (South of Milan), whereas (Fig. 2) the aquitard becomes discontinuous and then  
168 disappear moving northward.

169

## FIGURE 2

170 The underlying semi-confined (B) aquifer is 50 to 150 meter thick and consists of a sequence of  
171 gravel and medium-coarse sand in a sandy matrix with discontinuous confining layers of clay and  
172 silt. The base of the semi-confined aquifer consists of clay and silt layers and locally of  
173 conglomeratic units. The deep confined aquifer (C) consists of sandy lenses within clay and silt  
174 units representing the lower Pleistocene continental-marine transition facies (Colombo et al.,  
175 2018; Francani and Beretta, 1995; Pedretti et al., 2013). Concerning the same area, (Bonomi,  
176 2009) provided a 3D detailed definition of subsoil parameters proposing a data process method  
177 aimed to increase the value of the stratigraphic well logs. De Caro et al., (2017) more recently have  
178 provided a geochemical characterization of the area based on mapping of naturally controlled  
179 species, providing also NBLs for the area as a detailed mapping of contaminant trends and  
180 patterns. According to these authors, CrVI in this area cannot be attributed to geogenic sources  
181 due to the absence of ophiolites and serpentinites which are the main sources of natural  
182 chromium in other Italian regions (e.g. De Giusti et al., 2003; Rotiroti et al., 2015).

183

### 184 *NBLs vs DPBLs determination*

185 The NBL or Baseline level is defined as “the range of concentration of a given element, isotope or  
186 chemical compound in solution, derived entirely from natural, geological, biological or



187 atmospheric sources, under conditions not perturbed by anthropogenic activity” (Edmunds and  
188 Shand, 2009). Groundwaters from aquifers that are part of the active water cycle are influenced by  
189 human activities. Water changes chemistry from the moment it enters the system through rainfall  
190 infiltration, exchanges with surface water bodies or through other sources, until it leaves through  
191 runoff, evaporation or withdrawal (Vázquez-Suñé et al., 2005). Shallow aquifers in fact rarely  
192 reflect true natural concentration levels, whereas deep aquifers may be more likely free from  
193 anthropogenic impacts (Muller et al., 2006). Groundwater status in highly urbanized and farm  
194 areas is particularly affected by anthropogenic influence due to diffuse pollution deriving from  
195 many sources. This makes challenging to determine whether the observed groundwater  
196 conditions are the result of a natural chemical status according to the WFD directive or not  
197 (Wendland et al., 2005). Most of the methods used to determine NBLs rely on the separation of  
198 the natural from anthropogenic components based on some indicator chemical species, such as  
199  $\text{NO}_3$ ,  $\text{Cl}$  or  $\text{SO}_4$  (Hinsby et al., 2008; Matschullat et al., 2000; Muller et al., 2006) and the  
200 identification of the NBL as a fixed percentile value (e.g. 90<sup>th</sup> or 97<sup>th</sup>) of the observed distribution  
201 of the indicator concentrations. Another common characteristic of these methods is the univariate  
202 approach since, even when considering a set of parameters, the NBLs are evaluated on the basis of  
203 univariate frequency distributions. Moreover, even though many of the studies dealing with NBLs  
204 determination have also come up with a mapping of the contamination patterns (De Caro et al.,  
205 2017), only very few studies (Busico et al., 2018; Hasenmueller and Criss, 2013; Hwang et al.,  
206 2015) have attempted to determine some kind of anthropogenic DPBLs. One of the most  
207 innovative aspects of this study is the use of a multivariate approach to determine DPBLs.

208

209

210

211

212 *Used Data and pre-processing*

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

221

FIGURE 3

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

232

Table 1

233 Moreover, the study area's lithology and precipitation regime were considered. Particularly the  
234 lithology layer, extracted from the Regione Lombardia geoportal was used

235 (<http://www.geoportale.regione.lombardia.it/>), which contains the following information: the genetic  
236 classification of surface deposits; the classification of the rocky substratum based on the  
237 characteristics of petrographic composition, structure and texture; the main structural features  
238 and layers as the depths of the soils reported in classes (e.g. 0-50 cm; 50-100 cm; 100-200 cm).  
239 Moreover, precipitation and temperature data available at 15 meteorological stations included in  
240 the FUA, were used to estimate the rainfall depurated by evapotranspiration based on the  
241 Thornthwaite method (Thornthwaite and Mather, 1955).  
242 Then the net precipitation monthly values have been interpolated through a kriging technique and  
243 used for the analysis. Moreover, a grid of 1x1 km<sup>2</sup> was used to associate both lithology and  
244 precipitation to spatial units and to clusters. Particularly lithology has been evaluated either in  
245 terms of areal fraction of the different lithological types, either as overall similarity of the cluster.  
246 As similarity metric the Percentage Similarity Index, (i.e. PSC, after Brock, (1977) see Eq.1) was  
247 used:

$$248 \quad PSC = 100 - 0.5 \sum_{i=1}^k |a - b| \quad \text{Eq.1}$$

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

250

### 251 *Statistical Analysis*

252 All the statistical computations were made using the statistical package IBM SPSS Statistics 24.0.  
253 Principal Component and Factor Analysis (hereinafter FA, cfr. Afifi et al., (2003)) were performed  
254 based on the correlation matrix of the concentration measurements. Particularly, Factor Analysis  
255 was obtained through the preliminary Principal Components Analysis (hereinafter PCA) which  
256 extracted the eigenvalues and eigenvectors from the covariance matrix of the original variances.  
257 Factor analysis was chosen to reduce the contribution of the less significant parameters within

258 each component, by extracting a new set of varifactors through rotating the axes defined by the  
259 PCA extraction. The Varimax rotation criterion was used to rotate the PCA axes allowing to  
260 maintain the axes orthogonality. The number of factors to be retained was chosen on the basis of  
261 the “*eigenvalue higher than 1*” criterion (i.e. all the factors that explained less than the variance of  
262 one of the original variables were discarded). That allowed to select few factors able to describe  
263 the whole dataset with minimum loss of original information. Moreover, K-means Cluster Analysis  
264 (hereinafter CA, cfr. Afifi et al., (2003)) was used to analyze the similarities among the water  
265 quality profiles at the different monitoring stations, using the Euclidean Distance as distance  
266 metric (see Eq. 2).

$$267 \quad d(x_1, x_2) = \sqrt{\sum_{k=1}^p (x_{ik} - y_{jk})^2} \quad \text{Eq.2}$$

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

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

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

278

## 279 **Results**

280 *Preliminary analysis of multivariate dataset*

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

287 The application of the listwise deletion criterion reduced the size of multivariate dataset to about  
288 60% (N: 25800) of the mean size of whole dataset concerning the investigated parameters (N:  
289 about 42300).

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

#### 293 TABLE 2

294 As it can be observed, in fact, the distribution quartiles of the four pollutants of main interest for  
295 this study are quite comparable in the two situations; mean and standard deviation, being strongly  
296 affected by outliers that are only present in the whole dataset, do not correspond. Outliers in the  
297 whole dataset mostly refer to extremely high values measured at some specific pollution hotspot  
298 sites. In these sites, very often, only the information about pollutants was available, often lacking  
299 data about the other parameters of the multivariate dataset. In consideration of the fact that  
300 multivariate dataset had to be used to determine DPBLs, the absence of extreme outliers in the  
301 multivariate dataset was considered beneficial for the analysis. Moreover, in order to assess  
302 whether the spatial coverage of the multivariate dataset was about the same of the whole dataset  
303 considering the single parameters independently, the centroids of the measurement site  
304 coordinates, weighted by the corresponding frequency of measurements, obtained from either

305 the multivariate or the whole dataset were compared. As it can be observed in Figure 4, the two  
306 centroids are quite close, being the distance between them less than 500 m, so it can be  
307 concluded that the multivariate dataset has about the same spatial coverage of the whole dataset.

308 FIGURE 4

309 *Principal Component and Factor Analysis*

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

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

315 TABLE 3

316 The two sets of loadings are very similar, although in the rotated solution some parameters have  
317 “migrated” from a component to another. PCA solution had in fact a first component constituted  
318 by all the main hydrochemical parameters (i.e. conductivity and the main ions) explaining alone  
319 40% of the total variance, a second component strongly linked to PCE and TCE explaining about  
320 13% of the total variance, and the following components, each explaining less than 10% of the  
321 total variance. Moreover, in the PCA solution there are some parameters, namely Na, pH, Total Cr  
322 and TCM which load on two different components making more difficult the component  
323 interpretation.

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

329 across the rotated varifactors, being total chromium and TCM both respectively loaded only on the  
330 fourth (i.e. 8.5% of the total variance) and on the fifth (i.e. 8% of the total variance) varifactor.

331 In summary the extracted varifactors are the following:

332 *Varifactor 1:* Accounts 25.1% of the total variance and is loaded by Conductivity, Cl, Na, K, SO<sub>4</sub>; this  
333 varifactor represents the conductivity component that is most influenced by anthropic activity  
334 particularly concerning the salinization effect as described later; the high correlation between SO<sub>4</sub>,  
335 Cl and Na has been related to the leakage of agricultural and municipal wastes (Sikora et al., 1976).

336 *Varifactor 2:* Accounts 23.8% of the total variance and is loaded by Conductivity, Ca, Mg, NO<sub>3</sub> and  
337 pH; this varifactor represents the conductivity-hardness component of the groundwater. The  
338 correlation between Ca and NO<sub>3</sub> can be associated with the use of fertilizers such as  
339 NH<sub>4</sub>NO<sub>3</sub>·CaCO<sub>3</sub> (22% N and 33% NCaCO<sub>3</sub>) which is very common in cultivated regions of the study  
340 area (Tisdale and Nelsson, 1975).

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

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

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

#### 348 *K-means Cluster Analysis*

349 K-means cluster analysis was applied to the extracted FA varifactors. K was set to 15 and the  
350 analysis was run twice, using the final cluster centroids obtained from the first analysis as initial  
351 seeds for the second. Figure 5.a) shows the 15 clusters' composition in terms of number of cases  
352 included. It can be observed that most of the clusters contain a very small amount of data (i.e. less

353 than 1% of the total cases) while few clusters, namely cluster 3, 5, 6, 7 and 13, contain more than  
354 98% of the records including most of the background value measurements. Figure 5.b) shows the  
355 cluster characteristics in terms of varifactors.

356 **FIGURE 5**

357 It can be observed that clusters representing less than 1% of the total measurements are also the  
358 ones characterized by the higher values of the varifactors. Those clusters include in fact  
359 measurements that are outliers with respect of one or more of the considered parameters. Figure  
360 6 shows the characteristics of the five main clusters:

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

375 **FIGURE 6**

376



377 *Cluster interpretation and trend*

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

383 **FIGURE 7**

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

386 **FIGURE 8**

387 It can be observed that they present a comparable temporal trend. Such trends have been also  
388 tested through a Spearman's rho rank correlation analysis revealing that the PCE concentration  
389 profiles in the unconfined (A) and semiconfined (B) aquifers are correlated at a 0.05 and a 0.10  
390 significance level. The Spearman's correlation coefficients between A and B aquifers in the three  
391 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  
392 0.591 (P: 0.056; n: 11).

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

- 397 - *Cluster 3*: presents an increasing annual trend of the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>; r:  
398 0.359 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO<sub>3</sub>, pH r: - 0.139  
399 P<0.01). No trend is shown by the PCE/TCE and TCM varifactors whereas the fourth  
400 varifactor shows a weak decreasing trend (i.e. Total Cr; r: - 0.118 P<0.01).

- 401 - *Cluster 5*: as far as the conductivity components are concerned, cluster 5 presents a weak  
402 increasing trend of the first (i.e. Cl, Na, K, SO<sub>4</sub>; r: 0.162 P<0.01) and a very weak decreasing  
403 trend of the second varifactor (i.e. Ca, Mg, N-NO<sub>3</sub>; r: -0.03 P<0.01). Concerning the  
404 pollutant components, cluster 5 presents a weak increasing trend of the PCE/TCE varifactor  
405 (r: 0.217 P<0.01) and decreasing trends for the Total chromium and the TCM varifactors  
406 (respectively: r: -0.191 P<0.01; r: -0.168 P<0.01).
- 407 - *Cluster 6*: cluster 6 presents no trend concerning the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>; r:  
408 0.162 P<0.01) and a weak increasing trend of the second varifactor (i.e. Ca, Mg, N-NO<sub>3</sub>; r:  
409 0.132 P<0.01). Moreover, cluster 6 presents weak increasing trends of both the PCE/TCE  
410 varifactor (r: 0.116 P<0.01) and TCM varifactors (r: 0.052 P<0.01). No significant trend is  
411 shown by the Total-chromium varifactor.
- 412 - *Cluster 7*: cluster 7 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>; r:  
413 0.420 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO<sub>3</sub>, pH r: - 0.225  
414 P<0.01). Concurrently, the trend is very weakly decreasing for the PCE/TCE and Total  
415 chromium varifactors (respectively: r: -0.066 P<0.01; r: -0.07 P<0.01), while the TCM  
416 varifactor does not show any significant trend (P>0.05).
- 417 - *Cluster 13*: cluster 13 presents an increasing trend of the first varifactor (i.e. Cl, Na, K, SO<sub>4</sub>;  
418 r: 0.225 P<0.01) and a decreasing trend of the second (i.e. Ca, Mg, N-NO<sub>3</sub>, pH r: - 0.169  
419 P<0.01). Moreover, while the trend is increasing for the PCE/TCE varifactor (r: 0.163  
420 P<0.01), Total chromium and TCM varifactor trends are significantly decreasing  
421 (respectively: r: -0.237 P<0.01; r: -0.248 P<0.01).

#### 422 FIGURE 9

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

425

#### FIGURE 10

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

431

#### FIGURE 11

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

435

#### FIGURE 12

436 Table 4 shows the multiple comparison test results.

437

#### TABLE 4

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

442

#### TABLE 5

443

#### 444 *Towards the DPBLs determination*

445 The 5 identified clusters, representing the 98% of the sample, offer a robust support for the  
446 definition of DPBLs for the study area. As shown in Figure 13 in fact the five main clusters are very  
447 different also in terms of pollutant background values (see Supplementary tables S1, S2, S3) and

448 their own spatial distribution suggest that in the study area coexist different water quality profiles  
449 and different pollution background levels.

450 **FIGURE 13**

451 Particularly, clusters 3 and 7 (see Figure 10) have a more localized spatial distribution and show  
452 higher DPBLs in all the aquifers. Particularly, cluster 3 is characterized by solvents concentrations  
453 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  
454 cluster 7 has chromium and nitrates levels that respectively are almost double or very close to the  
455 corresponding GQS (respectively Cr-IV: 5 ug/l and NO<sub>3</sub>: 50 mg/l). On the other hand, clusters 5 and  
456 13 have a more diffuse spatial distribution and tend to have lower background values concerning  
457 all the pollutants. Cluster 6 is instead mostly concentrated within the municipality of Milan and it  
458 shows also well-defined characteristics: chromium levels that are higher than the ones of cluster 5  
459 and 13 but lower than the GQS, solvents and nitrates levels comparable with cluster 7 levels, being  
460 the PCE and TCM levels (TCE levels are higher than the GQS only in the semiconfined aquifer)  
461 higher than the corresponding GQS all over the area. Overall, it is relevant to observe, that both  
462 PCE and TCE present a wide logarithmic variability in their concentration values, concerning all the  
463 clusters, being respectively 3, 7 and 6 the cluster with the higher levels, which on some spots are  
464 many orders of magnitude higher than the GQS. Overall, background pollution levels in the  
465 different aquifers follow the same pattern of the cluster, being generally the background levels in  
466 the semiconfined (B) and in the undifferentiated aquifer higher than the corresponding levels in  
467 the unconfined aquifer (A).

468

## 469 **Discussion**

470 A widespread pattern of pollution occurrence in groundwaters has been documented in many  
471 urban areas all around Europe (Kao and Lei, 2000; Rivett and Feenstra, 2005; Stroo et al., 2003). In

472 time plumes coming from single point-sources (PS), mix with each other and end into a diffuse  
473 pattern of pollution linked to unidentifiable Multi-Point Sources, which, in most cases, started  
474 decades ago (Cortés et al., 2011).

475 Remediation of diffuse pollution to GQS levels, which very often are much lower than the  
476 background levels, is very difficult if not very unfeasible, and decision-makers, in many situations  
477 concerning Functional Urban Areas, are in fact urged to define new remediation goals, different  
478 from the GQS which are very often determined through a conservative risk assessment process.  
479 However, such a procedure is generally conducted by each Regional authority on case-by-case  
480 basis process while a national and European legislation about the management of diffuse pollution  
481 is still lacking. There is no doubt that relying on site-specific risk analysis in a FUAs with large  
482 patterns of diffuse contamination, it substantially increases clean-up associated costs compared to  
483 a very small improvement of groundwater quality. On the other hand, assessing the type of  
484 pollution, whether diffuse or localized, is critical as Public Authorities need to face those  
485 contaminations applying completely different actions, with dramatically different associated costs.  
486 If well-established methodological approaches exist for the identification of Natural Background  
487 Levels (Matschullat et al., 2000) of groundwater contaminants, no consolidated method either  
488 exist for identifying the anthropogenic groundwater diffuse pollution. The methodology proposed  
489 in this study is not intrinsically innovative, since multivariate techniques of the kind applied have  
490 been also used in several monitoring studies concerning both surface and groundwaters (Busico et  
491 al., 2018; Gourdol et al., 2013; Marcelli et al., 2010; Mendizabal et al., 2011; Olsen et al., 2012;  
492 Selle et al., 2013; Sheikhy Narany et al., 2014; Yidana et al., 2008; Yu et al., 2014); however, to our  
493 knowledge, this is the first time that these multivariate methods are employed to define DPBLs in  
494 FUA where MPS are present. The strength of this data driven approach is the possibility to identify  
495 hidden patterns in monitoring data (e.g. the hydrosomes as in Mendizabal et al., (2011)) directly

496 from the dataset. The analysis showed clearly that the Milan FUA is characterized by different  
497 hydrochemical groundwater subsystems with well differentiated water quality characteristics,  
498 which vary in time and space. The five main clusters identified through the analysis were in fact  
499 found to have different temporal profiles and background concentration levels. Moreover, the  
500 clusters were found to be characterized by different lithological characteristics. The acquired  
501 evidence is not enough to support the hypothesis that these clusters are a sort of hydrosomes as  
502 the ones described by (Mendizabal et al., 2011) (e.g. *hydrosome* is defined as a coherent, three-  
503 dimensional unit of groundwater with a specific origin) but it certainly supports the hypothesis  
504 that the FUA should not be seen or managed as an homogeneous groundwater body. Concerning  
505 the hydrogeochemical constitutes, in four of the five main clusters, the analysis reveals the  
506 presence of an increasing trend of the first varifactor (i.e. Na, K, Cl, SO<sub>4</sub>) and a weak but highly  
507 significant (P<0.01) decreasing trend of the second varifactor (i.e. Ca, Mg, NO<sub>3</sub>, pH). While the  
508 increasing trend of the first varifactor is a quite widespread pattern that has been found in many  
509 urbanised and cultivated territories, which to date concerns more than one-third of the world's  
510 irrigated land (Abbas et al., 2013; Heuperman et al., 2002; Houk et al., 2006; Li et al., 2015), the  
511 weak decreasing trend of the second varifactor might be due to cation exchange. As shown in  
512 Figure 14a, the dissolution of calcite, dolomite and gypsum seems to be the dominant reaction in  
513 the system as indicated in the plot of (Ca<sup>2+</sup>+ Mg<sup>2+</sup>) vs. (HCO<sub>3</sub><sup>-</sup> + SO<sub>4</sub><sup>2-</sup>), which is close to a 1:1 line.  
514 The plot of Ca<sup>2+</sup> and Mg<sup>2+</sup> concentrations compared with those of HCO<sub>3</sub><sup>-</sup> mostly in cluster 6 and 13  
515 may indicate the existence of an additional Ca<sup>2+</sup> and Mg<sup>2+</sup> source. Such a result seems to be  
516 consistent with De Caro et al., 2017 findings which revealed higher calcium concentrations in the  
517 northeastern sector of their study area which partially corresponds to ours. It is in fact worthwhile  
518 to observe that those high ratios between Ca<sup>2+</sup> and Mg<sup>2+</sup> and HCO<sub>3</sub><sup>-</sup> concentrations may not be  
519 attributed to HCO<sub>3</sub> depletion because of the existing neutral to slightly alkaline conditions (e.g.

520 groundwater pH ranges from 6.85 to 8.10, with an average of 7.52 and a standard deviation of  
521 0.25) which do not favour the formation of carbonic acid ( $\text{H}_2\text{CO}_3$ , see Spears, (1986)). Moreover,  
522 according to Jankowski et al. (1998), if active cation exchange between  $\text{Na}^+$  and  $\text{Ca}^{2+}$   $\text{Mg}^{2+}$  is  
523 occurring in the aquifer, the slope of the bivariate plot between  $\text{Cl}^-$  corrected  $\text{Na}^+ + \text{K}^+$  and ( $\text{Ca}^{2+} +$   
524  $\text{Mg}^{2+}$ ) corrected  $\text{HCO}_3^- + \text{SO}_4^{2-}$  (see Figure 14c) would be  $-1$  (i.e.,  $y = -x$ ). The slope obtained for  
525 this plot in the study area ( $-1.72$ ) seems to indicate that cation exchange is acting as an important  
526 process.

#### 527 FIGURE 14

528 It is also worthwhile to underline that correlations with precipitation are weak, being insignificant  
529 (i.e. cluster 3) or weakly direct concerning the first varifactor, and weakly indirect concerning the  
530 second varifactor. The correlation between  $\text{Na}^+$  and  $\text{Cl}^-$  in the five main clusters is also different,  
531 being always significant at the  $P < 0.01$  level, and higher than 0.6-0.7 in the smaller clusters (i.e.  
532 0.76, 0.63 and 0.73 respectively for clusters 3, 6 and 7 distributed mostly within the Milan  
533 municipality) and lower than 0.5 for the bigger clusters (i.e. 0.38, and 0.55 respectively for clusters  
534 5, and 13). In this respect it has been suggested that the poor correlation ( $r^2 < 0.54$  that would  
535 correspond to an  $r < 0.74$ ) between  $\text{Na}^+$  and  $\text{Cl}^-$  might indicate a possible evaporative  
536 concentration of these ions (Rose, 2002).  $\text{Na}^+/\text{Cl}^-$  ratio has been used, especially in semiarid or arid  
537 regions, to identify the mechanism for acquiring salinity, since the  $\text{Cl}^-$  ion is not affected by soil  
538 retention while the  $\text{Na}^+$  ion is retained (Jalali and Khanlari, 2008; Tiwari and Singh, 2014);  
539 enrichment of  $\text{Na}^+$  may also result from reactions taking place in the clay mineral of the surface  
540 soil horizons. As shown in Figure 14d the clusters, especially the bigger ones, have a marked  
541 fingerprint with respect to the  $\text{Na}^+/\text{Cl}^-$  ratio vs  $\text{Cl}^-$  concentrations being cluster 5 the one of the  
542 highest  $\text{Na}^+$  enrichment and cluster 6 the one with the lowest.

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

551 Correlations with precipitation for the pollutant varifactors are all very weak although significant  
552 suggesting that meteo-climatic variability may also affect the observed patterns. As pointed out  
553 earlier, cluster 3 and 7 both have higher average concentrations of the studied contaminants:  
554 cluster 3 having PCE and TCM levels respectively 2 times and 5 times higher than the whole  
555 sample average, and cluster 7 showing PCE and Total-Cr levels 2 times higher the whole sample  
556 average. The wells belonging to these clusters are mostly inside the boundaries of Milan  
557 municipality and are positioned downgradient to a couple of historical industrial district areas  
558 where many brownfields were remediated in the last 20 years. So, despite the effort done in the  
559 past to find the contamination hot spots (i.e. PS) and to remediate soil and groundwater, there are  
560 still traces of the past production activities whose impact can still be detected. This is because  
561 many small unidentifiable sources (i.e. MPS) are still present, but since they are releasing a low  
562 mass of contaminants there is no chance to locate their position and to apply a specific  
563 remediation action. Luckily concentrations in these two clusters are showing a decreasing trend in  
564 time, probably due to Natural Attenuation Processes that could be assessed, monitored and  
565 maybe enhanced to accelerate the achievement of the groundwater quality status. Territorial  
566 authorities need support in the process of identifying such trends, and their awareness of the



567 existence of portions of the groundwater body that may have different quality characteristics and  
568 trends need to be arisen. For the same reason appropriate DPBLs should be identified for these  
569 portions since only through this process it would be possible to properly manage these areas  
570 applying adequate planning policies and actions (e.g. investigations and monitoring). We showed  
571 that diffuse pollution is present within the FUA with different water quality levels and different  
572 distribution: cluster 13 being the dominant water quality profile representing the 35% of the  
573 boreholes, followed by cluster 5 representing the 31% and by cluster 6 representing the 21%.  
574 Among the contaminants considered in this study only PCE and TCM systematically exceed their  
575 corresponding GQs with median and interquartile values respectively of 3, 1÷7 µg/l, and 0.5,  
576 0.5÷2 µg/l, while TCE and Cr-VI present only spots of higher contamination, with median and  
577 interquartile values respectively of 1, 0.5÷2.7 µg/l and 3.9, 2.3÷6.6 µg/l. Defining appropriate  
578 DPBLs for these contaminants is both a political and a scientific-technical issue and is beyond the  
579 scope of this paper, however, we the evidence provided by this study about the FUA not being a  
580 whole homogeneous groundwater body strongly support the hypothesis that different DBPLs  
581 should be defined within FUA subareas. Moreover, as weak trends were found, revealing a slow  
582 temporal evolution of the contamination (i.e. in average less than 0.06 ug/l per year for the  
583 considered decade). Reliable decennial background levels should be drawn every 5-10 years with  
584 the caveat of periodically revising. Furthermore, being the clusters robust and stable in time, they  
585 allow to overcome the problem of spatial and temporal heterogeneity of measurements which  
586 particularly affects this area and offer a robust data series to define spatially the DPBLs through  
587 geostatistical approaches. Deterministic fate and transport models will also be useful to simulate  
588 the most important plumes deriving from a PS contamination in the FUA, and to identify the  
589 subareas affected by these plumes that should be excluded from the geostatistical interpolation.

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

594

## 595 **Conclusions**

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

607

## 608 **Acknowledgements**

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

614 **References**

- 615 Abbas, A., Khan, S., Hussain, N., Hanjra, M.A., Akbar, S., 2013. Characterizing soil salinity in  
616 irrigated agriculture using a remote sensing approach. *Phys. Chem. Earth.*  
617 doi:10.1016/j.pce.2010.12.004
- 618 Afifi, A., May, S., Clark, V.A., 2003. Cluster analysis, in: *Computer-Aided Multivariate Analysis*. pp.  
619 428–444.
- 620 Alberti, L., Cantone, M., Colombo, L., Lombi, S., Piana, A., 2016. Numerical modeling of regional  
621 groundwater flow in the Adda-Ticino Basin: Advances and new results. *Rend. Online Soc.*  
622 *Geol. Ital.* 41. doi:10.3301/ROL.2016.80
- 623 Alberti, L., Colombo, L., Formentin, G., 2018. Null-space Monte Carlo particle tracking to assess  
624 groundwater PCE (Tetrachloroethene) diffuse pollution in north-eastern Milan functional  
625 urban area. *Sci. Total Environ.* 621. doi:10.1016/j.scitotenv.2017.11.253
- 626 Alberti, L., Colombo, L., La Licata, I., Cantone, M., 2018. Model calibration using the automatic  
627 parameter estimation procedure (PEST) of the North-eastern zone of the Milan Functional  
628 Urban Area (Italy). *Acque Sotter. - Ital. J. Groundw.* 7, 27–38. doi:10.7343/as-2018-336
- 629 Balderacchi, M., Benoit, P., Cambier, P., Eklo, O.M., Gargini, A., Gemitzi, A., Gurel, M., Kløve, B.,  
630 Nakic, Z., Predaa, E., Ruzicic, S., Wachniew, P., Trevisan, M., 2013. Groundwater pollution and  
631 quality monitoring approaches at the European level. *Crit. Rev. Environ. Sci. Technol.*  
632 doi:10.1080/10643389.2011.604259
- 633 Balderacchi, M., Filippini, M., Gemitzi, A., Kløve, B., Petitta, M., Trevisan, M., Wachniew, P.,  
634 Witczak, S., Gargini, A., 2014. Does groundwater protection in Europe require new EU-wide  
635 environmental quality standards? *Front. Chem.* doi:10.3389/fchem.2014.00032
- 636 Barrlo-Lage, G., Parsons, F.Z., Nassar, R.S., 1987. Kinetics of the Depletion of Trichloroethene.  
637 *Environ. Sci. Technol.* 21, 366–370. doi:10.1021/es00158a005
- 638 Bonomi, T., 2009. Database development and 3D modeling of textural variations in  
639 heterogeneous, unconsolidated aquifer media: Application to the Milan plain. *Comput.*  
640 *Geosci.* doi:10.1016/j.cageo.2007.09.006
- 641 Brock, D.A., 1977. Comparison of community similarity indexes. *Water Pollut. Control Fed.* 49,  
642 2488–2494.
- 643 Busico, G., Cuoco, E., Kazakis, N., Colombani, N., Mastrocicco, M., Tedesco, D., Voudouris, K., 2018.  
644 Multivariate statistical analysis to characterize/discriminate between anthropogenic and  
645 geogenic trace elements occurrence in the Campania Plain, Southern Italy. *Environ. Pollut.*  
646 doi:10.1016/j.envpol.2017.11.053
- 647 Carcano, C., Piccin, A., 2002. Padans aquifer geology of the Lombardy Plain. Firenze.
- 648 Carral, E., Puente, X., Villares, R., Carballeira, A., 1995. Background heavy metal levels in estuarine  
649 sediments and organisms in Galicia (northwest Spain) as determined by modal analysis. *Sci.*  
650 *Total Environ.* doi:10.1016/0048-9697(95)04788-3
- 651 Carter, J.M., Moran, M.J., Zogorski, J.S., Price, C. V., 2012. Factors associated with sources,

- 652 transport, and fate of chloroform and three other trihalomethanes in untreated groundwater  
653 used for drinking water. *Environ. Sci. Technol.* doi:10.1021/es301839p
- 654 Chapman, S.W., Parker, B.L., 2005. Plume persistence due to aquitard back diffusion following  
655 dense nonaqueous phase liquid source removal or isolation. *Water Resour. Res.*  
656 doi:10.1029/2005WR004224
- 657 Chapman, S.W., Parker, B.L., Sale, T.C., Doner, L.A., 2012. Testing high resolution numerical models  
658 for analysis of contaminant storage and release from low permeability zones. *J. Contam.*  
659 *Hydrol.* doi:10.1016/j.jconhyd.2012.04.006
- 660 Colombo, L., Gattinoni, P., Scesi, L., 2018. Stochastic modelling of groundwater flow for hazard  
661 assessment along the underground infrastructures in Milan (northern Italy). *Tunn. Undergr.*  
662 *Sp. Technol.* 79. doi:10.1016/j.tust.2018.05.007
- 663 Cortés, A., Puigserver, D., Carmona, J.M., Viladevall, M., 2011. Biological remediation approach  
664 involving soils and groundwaters polluted with chlorinated solvents in a Mediterranean  
665 context. *Recent Adv. Pharm. Sci.* 661, 223–246.
- 666 De Caro, M., Crosta, G.B., Frattini, P., 2017. Hydrogeochemical characterization and Natural  
667 Background Levels in urbanized areas: Milan Metropolitan area (Northern Italy). *J. Hydrol.*  
668 doi:10.1016/j.jhydrol.2017.02.025
- 669 De Giusti, F., Piazz, D., Massironi, M., Schiavo, A., 2003. Carta geotettonica della Valle d ' Aosta.  
670 *Mem. Sci. Geol.* 129–149.
- 671 Edmunds, W.M., Shand, P., 2009. *Natural Groundwater Quality*, Natural Groundwater Quality.  
672 doi:10.1002/9781444300345
- 673 Emsley, J., 2011. *Nature's Building Blocks: An A-Z Guide to the Elements*, Oxford University Press.  
674 doi:978-0-19-960563-7
- 675 European Community, 2006. Groundwater directive 2006/118/CE. Directive of the European  
676 Parliament and of the Council on the Protection of Groundwater Against Pollution and  
677 Deterioration. *Off. J. Eur. Union* 19, 19–31. doi:http://eur-lex.europa.eu/legal-  
678 content/EN/TXT/?uri=CELEX:32006L0118
- 679 Francani, V., Beretta, G. Pietro, 1995. Preliminary survey about sustainable use of deeper  
680 groundwater in Milano province. Milan.
- 681 Fruchter, J., 2002. In-Situ Treatment of Chromium-Contaminated Groundwater. *Environ. Sci.*  
682 *Technol.* doi:10.1021/es022466i
- 683 Giudici, M., Ponzini, G., Romano, E., Vassena, C., 2007. Some Lessons from Modeling Ground  
684 Water Flow in the Metropolitan Area of Milano at Different Scales. *Mem. Descr. della Cart.*  
685 *Geol. d'Italia LXXVI*, 207–218.
- 686 Gourdol, L., Hissler, C., Hoffmann, L., Pfister, L., 2013. On the potential for the Partial Triadic  
687 Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry. *Appl.*  
688 *Geochemistry.* doi:10.1016/j.apgeochem.2013.10.002
- 689 Guilbeault, M.A., Parker, B.L., Cherry, J.A., 2005. Mass and flux distributions from DNAPL zones in

- 690 sandy aquifers. *Ground Water*. doi:10.1111/j.1745-6584.2005.tb02287.x
- 691 Hasenmueller, E.A., Criss, R.E., 2013. Multiple sources of boron in urban surface waters and  
692 groundwaters. *Sci. Total Environ*. doi:10.1016/j.scitotenv.2013.01.001
- 693 Heuperman, A.F., Kapoor, A.S., Denecke, H.W., 2002. *Biodrainage - Principles, Experiences and*  
694 *Applications*. Rome.
- 695 Hinsby, K., Condesso de Melo, M.T., Dahl, M., 2008. European case studies supporting the  
696 derivation of natural background levels and groundwater threshold values for the protection  
697 of dependent ecosystems and human health. *Sci. Total Environ*.  
698 doi:10.1016/j.scitotenv.2008.03.018
- 699 Houk, E., Frasier, M., Schuck, E., 2006. The agricultural impacts of irrigation induced waterlogging  
700 and soil salinity in the Arkansas Basin. *Agric. Water Manag.* 85, 175–183.  
701 doi:10.1016/j.agwat.2006.04.007
- 702 Hunkeler, D., Aravena, R., 2000. Determination of compound-specific carbon isotope ratios of  
703 chlorinated methanes, ethanes, and ethenes in aqueous samples. *Environ. Sci. Technol.*  
704 doi:10.1021/es991178s
- 705 Hwang, H.-H., Panno, S. V., Hackley, K.C., 2015. Sources and Changes in Groundwater Quality with  
706 Increasing Urbanization, Northeastern Illinois. *Environ. Eng. Geosci.*  
707 doi:10.2113/gseegeosci.21.2.75
- 708 ISTAT, 2017. Annual survey on movement and calculation of the resident population. Rome.
- 709 Izbicki, J.A., Ball, J.W., Bullen, T.D., Sutley, S.J., 2008. Chromium, chromium isotopes and selected  
710 trace elements, western Mojave Desert, USA. *Appl. Geochemistry*.  
711 doi:10.1016/j.apgeochem.2007.11.015
- 712 Jalali, M., Khanlari, Z. V., 2008. Major ion chemistry of groundwaters in the Damagh area,  
713 Hamadan, western Iran. *Environ. Geol.* doi:10.1007/s00254-007-0795-6
- 714 Kao, C.M., Lei, S.E., 2000. Using a peat biobarrier to remediate PCE/TCE contaminated aquifers.  
715 *Water Res.* doi:10.1016/S0043-1354(99)00213-4
- 716 Kueper, B.H., Wealthall, G.P., Smith, J.W.N., Leharne, S. a, Lerner, D.N., 2003. An illustrated  
717 handbook of DNAPL transport and fate in the subsurface, R&D Publication 133.
- 718 Li, W., Wang, M.-Y., Liu, L.-Y., Yan, Y., 2015. Assessment of Long-Term Evolution of Groundwater  
719 Hydrochemical Characteristics Using Multiple Approaches: A Case Study in Cangzhou,  
720 Northern China. *Water* 7, 1109–1128. doi:10.3390/w7031109
- 721 Mackay, D.M., Cherry, J.A., 1989. Groundwater contamination: Pump-and-treat remediation.  
722 *Environ. Sci. Technol.* doi:10.1021/es00064a001
- 723 Marcelli, R., Bartolucci, G., Papaioannu, G., De Angelis, G., Lucibello, A., Proietti, E., Margesin, B.,  
724 Giacomozzi, F., Deborgies, F., 2010. Reliability of RF MEMS switches due to charging effects  
725 and their circuitual modelling, in: *Microsystem Technologies*. pp. 1111–1118.  
726 doi:10.1007/s00542-009-1006-z
- 727 Matschullat, J., Ottenstein, R., Reimann, C., 2000. Geochemical background - Can we calculate it?

- 728 Environ. Geol. doi:10.1007/s002549900084
- 729 McGuire, T.M., McDade, J.M., Newell, C.J., 2006. Performance of DNAPL source depletion  
730 technologies at 59 chlorinated solvent-impacted sites. *Gr. Water Monit. Remediat.*  
731 doi:10.1111/j.1745-6592.2006.00054.x
- 732 Mendizabal, I., Stuyfzand, P.J., Wiersma, A.P., 2011. Hydrochemical system analysis of public  
733 supply well fields, to reveal water-quality patterns and define groundwater bodies: The  
734 Netherlands. *Hydrogeol. J.* doi:10.1007/s10040-010-0614-0
- 735 Molinari, A., Chidichimo, F., Straface, S., Guadagnini, A., 2014. Assessment of natural background  
736 levels in potentially contaminated coastal aquifers. *Sci. Total Environ.*  
737 doi:10.1016/j.scitotenv.2013.12.125
- 738 Molinari, A., Guadagnini, L., Marcaccio, M., Guadagnini, A., 2012. Natural background levels and  
739 threshold values of chemical species in three large-scale groundwater bodies in Northern  
740 Italy. *Sci. Total Environ.* doi:10.1016/j.scitotenv.2012.03.015
- 741 Muller, D., Blum, A., Hart, A., Hookey, J., Kunkel, R., Scheidleder, A., 2006. Final Report - BRIDGE (  
742 Background criteria for the identification of groundwater thresholds ) 13030.
- 743 Munholland, J.L., Mumford, K.G., Kueper, B.H., 2016. Factors affecting gas migration and  
744 contaminant redistribution in heterogeneous porous media subject to electrical resistance  
745 heating. *J. Contam. Hydrol.* doi:10.1016/j.jconhyd.2015.10.011
- 746 Nriagu, J.O., Nieboer, E., 1988. Chromium in the Natural and Human Environments. John Wiley &  
747 Sons.
- 748 OECD, 2012. Redefining "Urban": A New Way to Measure Metropolitan Areas. *OECD Publ.* 1–9.  
749 doi:10.1787/9789264174108-en
- 750 Olsen, R.L., Chappell, R.W., Loftis, J.C., 2012. Water quality sample collection, data treatment and  
751 results presentation for principal components analysis e literature review and Illinois River  
752 watershed case study. *Water Res.*
- 753 Paine, A.J., 2001. Mechanisms of chromium toxicity, carcinogenicity and allergenicity: Review of  
754 the literature from 1985 to 2000. *Hum. Exp. Toxicol.* doi:10.1191/096032701682693062
- 755 Parker, B.L., Cherry, J.A., Chapman, S.W., 2004. Field study of TCE diffusion profiles below DNAPL  
756 to assess aquitard integrity. *J. Contam. Hydrol.* doi:10.1016/j.jconhyd.2004.02.011
- 757 Pecoraino, G., Scalici, L., Avellone, G., Ceraulo, L., Favara, R., Candela, E.G., Provenzano, M.C.,  
758 Scaletta, C., 2008. Distribution of volatile organic compounds in Sicilian groundwaters  
759 analysed by head space-solid phase micro extraction coupled with gas chromatography mass  
760 spectrometry (SPME/GC/MS). *Water Res.* doi:10.1016/j.watres.2008.07.022
- 761 Pedretti, D., Masetti, M., Beretta, G.P., Vitiello, M., 2013. A revised conceptual model to reproduce  
762 the distribution of chlorinated solvents in the Rho Aquifer (Italy). *Groundw. Monit. Remediat.*  
763 doi:10.1111/gwmr.12017
- 764 Perego, R., Bonomi, T., Fumagalli, M.L., Benastini, V., Aghib, F., Rotiroti, M., Cavallin, A., 2014. 3D  
765 reconstruction of the multi-layer aquifer in a Po Plain area. *Rend. Online Soc. Geol. Ital.* 30,

- 766 41–44. doi:10.3301/ROL.2014.09
- 767 Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: Death  
768 of a myth. Consequences for the statistical treatment of geochemical and environmental  
769 data. *Environ. Geol.* doi:10.1007/s002549900081
- 770 Reinmann, C., De Caritat, P., 1998. *Chemical Elements in the Environment*. SPRINGER-VERLAG,  
771 Berlin.
- 772 Rivett, M.O., Feenstra, S., 2005. Dissolution of an emplaced source of DNAPL in a natural aquifer  
773 setting. *Environ. Sci. Technol.* doi:10.1021/es040016f
- 774 Rose, S., 2002. Comparative major ion geochemistry of Piedmont streams in the Atlanta, Georgia  
775 region: Possible effects of urbanization. *Environ. Geol.* doi:10.1007/s00254-002-0545-8
- 776 Rotiroti, M., Di Mauro, B., Bonomi, T., Fumagalli, M.L., 2015. COMPSEC , a new tool to derive  
777 natural background levels by the component separation approach : Application in two  
778 different hydrogeological contexts in northern Italy COMPSEC , a new tool to derive natural  
779 background levels by the component separation. doi:10.1016/j.gexplo.2015.06.017
- 780 Schwille, F. (1984), 1984. Migration of Organic Fluids Immiscible with Water in the unsaturated  
781 zone 2–3.
- 782 Schwille, F. (1988), 1988. *Dense Chlorinated Solvents in Porous and Fractured Media: Model  
783 Experiments*, translated from the German by J. F. Pankow.
- 784 Selle, B., Schwientek, M., Lischeid, G., 2013. Understanding processes governing water quality in  
785 catchments using principal component scores. *J. Hydrol.* doi:10.1016/j.jhydrol.2013.01.030
- 786 Sheikhy Narany, T., Ramli, M.F., Aris, A.Z., Sulaiman, W.N.A., Fakharian, K., 2014. Spatiotemporal  
787 variation of groundwater quality using integrated multivariate statistical and geostatistical  
788 approaches in Amol-Babol Plain, Iran. *Environ. Monit. Assess.* doi:10.1007/s10661-014-3820-  
789 8
- 790 Sikora, L., Bent, M., Corey, R., Keeney, D., 1976. Septic nitrogen and phosphorus removal test  
791 system. *Groundwater* 14, 309–314.
- 792 Spears, D.A., 1986. Mineralogical control of the chemical evolution of groundwter. *Solute Process.*  
793 512.
- 794 Stevenazzi, S., Masetti, M., Nghiem, S. V., Sorichetta, A., 2015. Groundwater vulnerability maps  
795 derived from a time-dependent method using satellite scatterometer data. *Hydrogeol. J.* 23,  
796 631–647. doi:10.1007/s10040-015-1236-3
- 797 Stroo, H.F., Unger, M., Ward, C.H., Kavanaugh, M.C., Vogel, C., Leeson, A., Marqusee, J.A., Smith,  
798 B.P., 2003. Peer Reviewed: Remediating Chlorinated Solvent Source Zones. *Environ. Sci.*  
799 *Technol.* 37, 224A–230A. doi:10.1021/es032488k
- 800 Thornthwaite, C.W., Mather, J.R., 1955. The water balance. *Publ. Climatol.* 8, 1–104.
- 801 Tisdale, S., Nelsson, W., 1975. *Soil fertility and fertilizers*. Macmillan Publishing Co Inc, New York.
- 802 Tiwari, A.K., Singh, A.K., 2014. Hydrogeochemical investigation and groundwater quality

803 assessment of Pratapgarh district, Uttar Pradesh. *J. Geol. Soc. India.* doi:10.1007/s12594-014-  
804 0045-y

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

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

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

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

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

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

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

825 WHO, 1988. *Environmental Health Criteria.* Geneve.

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

829 Yu, Y., Song, X., Zhang, Y., Zheng, F., Liang, J., Liu, L., 2014. Identifying spatio-temporal variation  
830 and controlling factors of chemistry in groundwater and river water recharged by reclaimed  
831 water at Huai River, North China. *Stoch. Environ. Res. Risk Assess.* doi:10.1007/s00477-013-  
832 0803-1

833

834





**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
<b>N</b>		45602	25757	45190	25757	44083	25757	46042	25757
<b>Mean</b>		15.316	6.193	13.816	13.818	3.542	2.535	9.114	3.854
<b>Median</b>		3.000	2.000	1.000	1.000	0.500	0.500	2.500	2.500
<b>Std. Deviation</b>		311.685	28.802	260.100	250.020	87.811	8.919	149.219	5.087
<b>Minimum</b>		0.003	0.010	0.003	0.010	0.005	0.010	0.005	0.200
<b>Maximum</b>		37800	3000	14000	14000	17214	341	12300	190
<b>Percentiles</b>	<b>25</b>	0.610	0.500	0.500	0.500	0.500	0.500	2.500	2.500
	<b>50</b>	3.000	2.000	1.000	1.000	0.500	0.500	2.500	2.500
	<b>75</b>	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						Rotated Component Matrix					
	Component						Component				
	1	2	3	4	5		1	2	3	4	5
Ca	<b>0.895</b>	-0.250	-0.194	-0.012	-0.053	Ca	0.462	<b>0.822</b>	-0.056	0.116	0.003
Cl	<b>0.839</b>	0.076	0.318	0.029	0.016	Cl	<b>0.821</b>	0.345	0.045	0.077	0.104
Conductivity	<b>0.960</b>	-0.128	0.020	-0.010	-0.017	Conductivity	<b>0.675</b>	<b>0.686</b>	-0.016	0.105	0.032
Total Cr	0.151	-0.022	-0.219	<b>0.689</b>	<b>0.606</b>	Total Cr	-0.005	0.043	0.021	<b>0.954</b>	-0.005
Mg	<b>0.691</b>	-0.318	-0.311	-0.052	-0.130	Mg	0.219	<b>0.798</b>	-0.096	0.035	-0.005
NO3	<b>0.714</b>	-0.050	-0.194	0.189	0.057	NO3	0.366	<b>0.589</b>	0.089	0.305	0.080
pH	-0.485	0.123	<b>0.550</b>	0.246	0.214	pH	0.051	<b>-0.774</b>	-0.166	0.147	0.089
K	<b>0.571</b>	-0.015	0.394	-0.146	0.102	K	<b>0.691</b>	0.142	-0.077	-0.042	-0.088
Na	<b>0.642</b>	0.350	<b>0.506</b>	-0.137	0.097	Na	<b>0.875</b>	0.002	0.223	-0.060	-0.005
Sulfates	<b>0.860</b>	0.119	0.100	0.090	0.014	Sulfates	<b>0.695</b>	0.472	0.160	0.164	0.120
PCE	0.203	<b>0.782</b>	-0.298	0.058	-0.015	PCE	0.071	0.057	<b>0.842</b>	0.095	0.133
TCE	0.143	<b>0.821</b>	-0.283	-0.203	0.054	TCE	0.075	0.001	<b>0.896</b>	-0.062	-0.085
TCM	0.037	0.195	0.158	<b>0.650</b>	<b>-0.693</b>	TCM	0.013	-0.030	0.038	0.000	<b>0.982</b>
<b>Eigenvalue</b>	5.212	1.664	1.239	1.090	0.939	<b>Eigenvalue</b>	3.262	3.098	1.645	1.102	1.037
<b>Explained variance (%)</b>	40.094	12.802	9.535	8.381	7.226	<b>Explained variance (%)</b>	25.096	23.833	12.652	8.478	7.979
<b>Cumulative (%) of variance</b>	40.094	52.896	62.430	70.811	78.037	<b>Cumulative (%) of variance</b>	25.096	48.929	61.581	70.059	78.037
Extraction Method: Principal Component Analysis.						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.

<i>Gravel</i>					
	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.	
<i>Silt</i>					
	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.	
<i>Sand</i>					
	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.	**	
<i>PSC</i>					
	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).

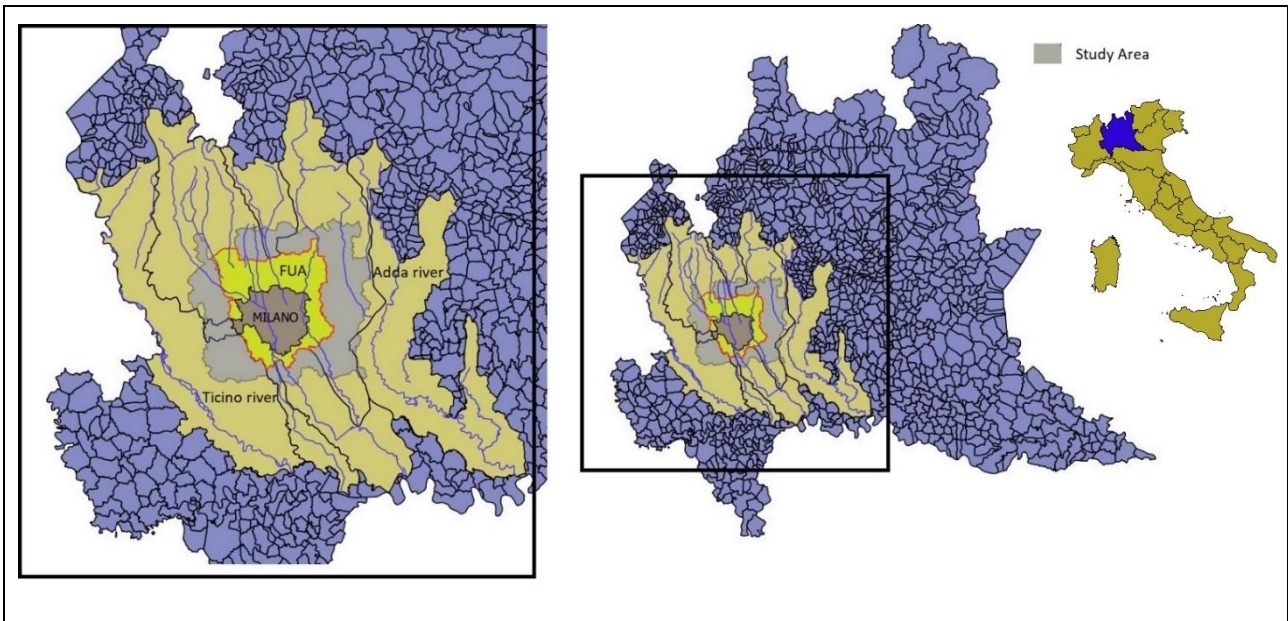


Fig. 1. Study Area: Milan FUA and surrounding municipalities.

Figure 2

[Click here to download Figure: Fig 2\\_rev.docx](#)

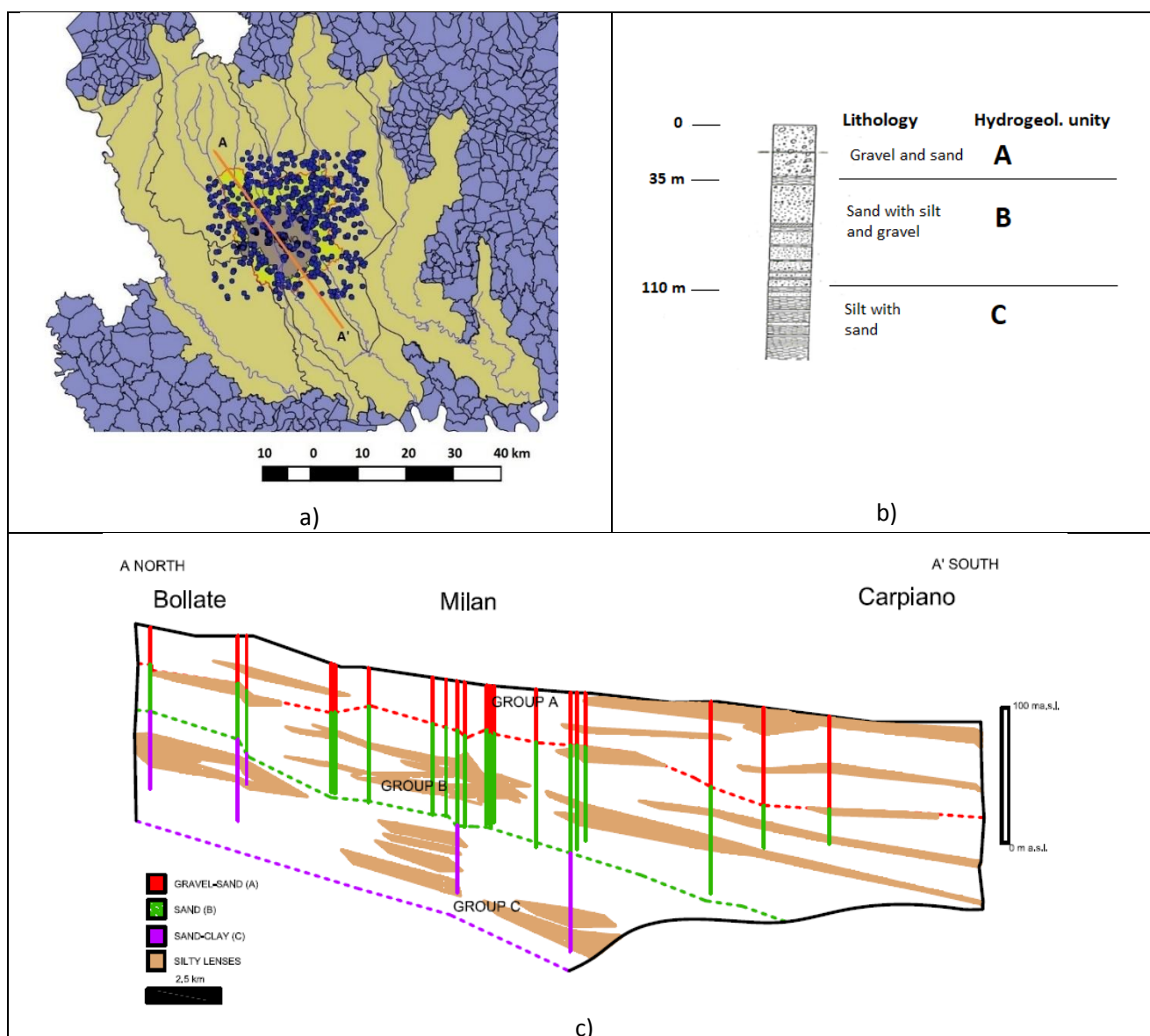


Fig. 2 – Hydrogeological settings of the study area according to Regione Lombardia and ENI (2002). (a) map of the borehole database and the cross section profile through the FUA of Milan; (b) hydrostratigraphic units. (extracted from Foglio Geologico Milano); (c) simplified cross section of the investigated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

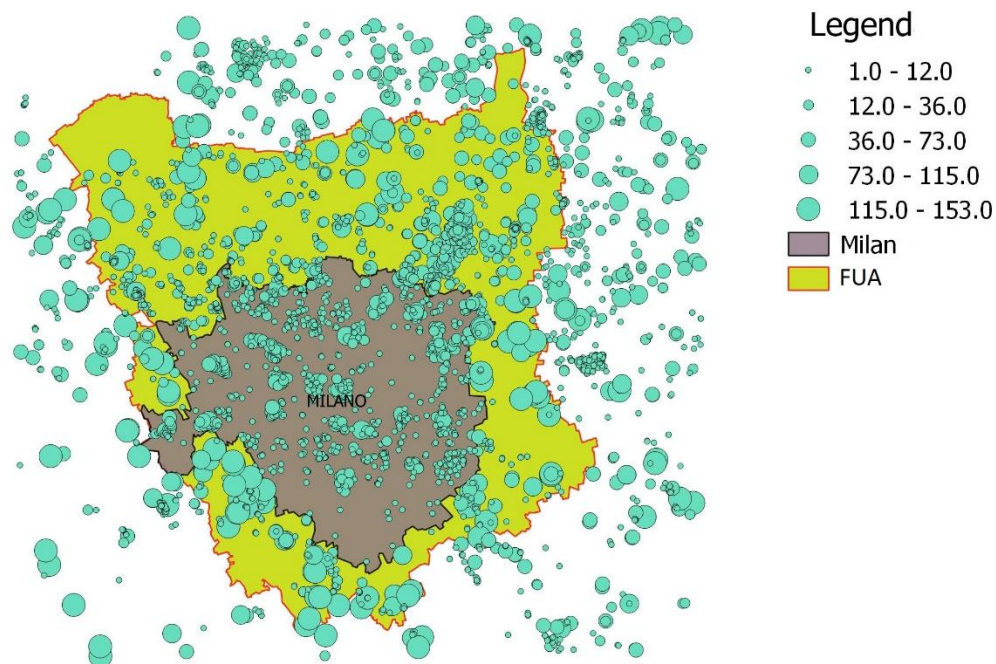


Figure 3: spatial distribution of sampling wells across the study area and the data availability at each sampling well.



## Figure 4

[Click here to download Figure: Figure 4.docx](#)

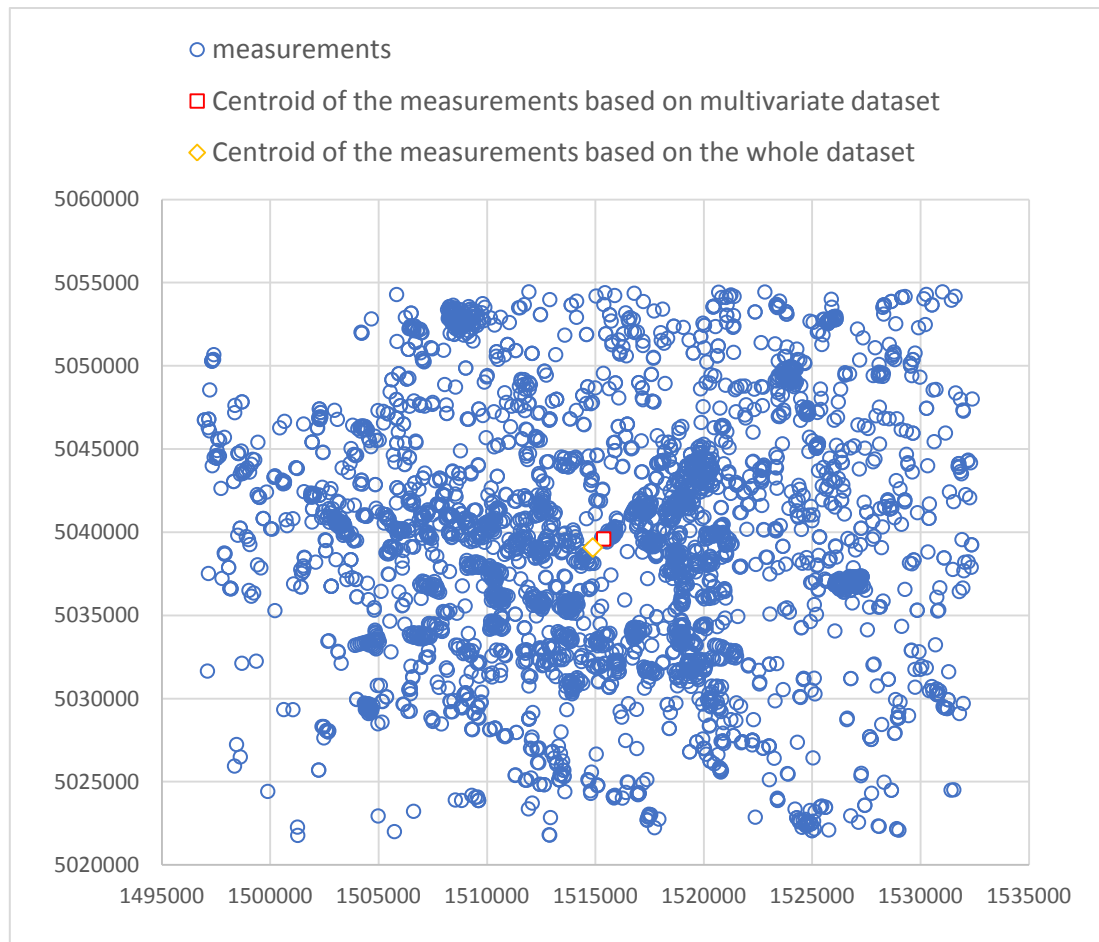


Figure 4 – Map showing the coordinates of the measurements concerning the whole and the multivariate dataset. The centroids weighted on the frequency of measurements are also shown.

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)

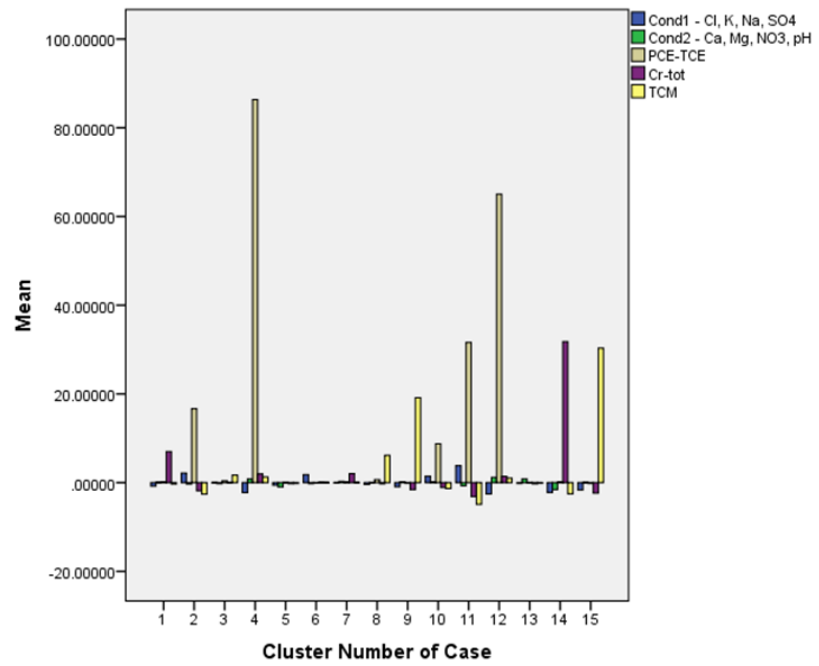


Figure 5 1- a) number of cases included in the K-means clusters; b) cluster characteristics in terms of standardised varifactors (i.e. average 0 and standard deviation 1): bars show the average value of the varifactors within each cluster.

# Figure 6

[Click here to download Figure: Figure 6.docx](#)

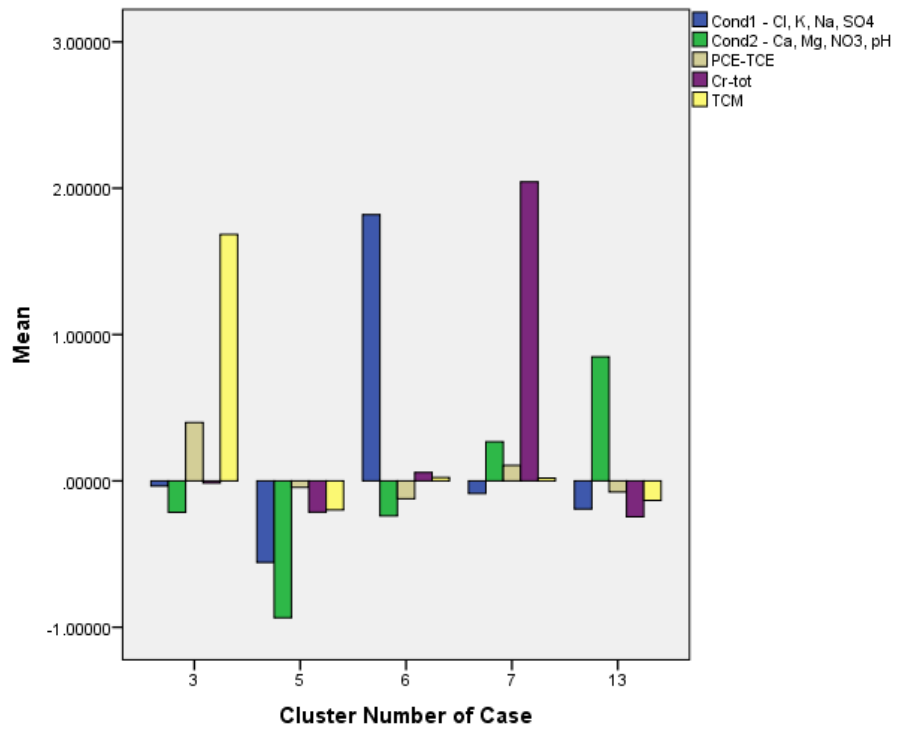


Figura 6 – Varifactor characteristics of the main clusters (> 1% of the total sample): bars show the mean varifactor values within the clusters, while 0 that is the total sample average.

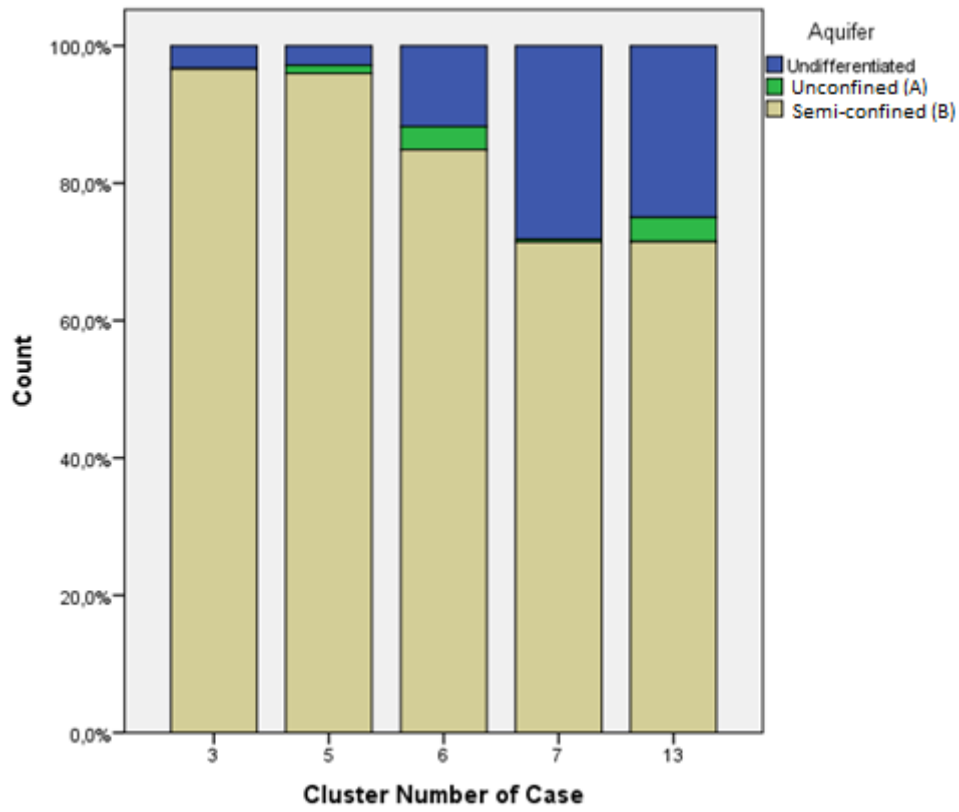


Fig. 7 – Aquifer composition of the five main clusters.

Figure 8

[Click here to download Figure: Figure 8.docx](#)

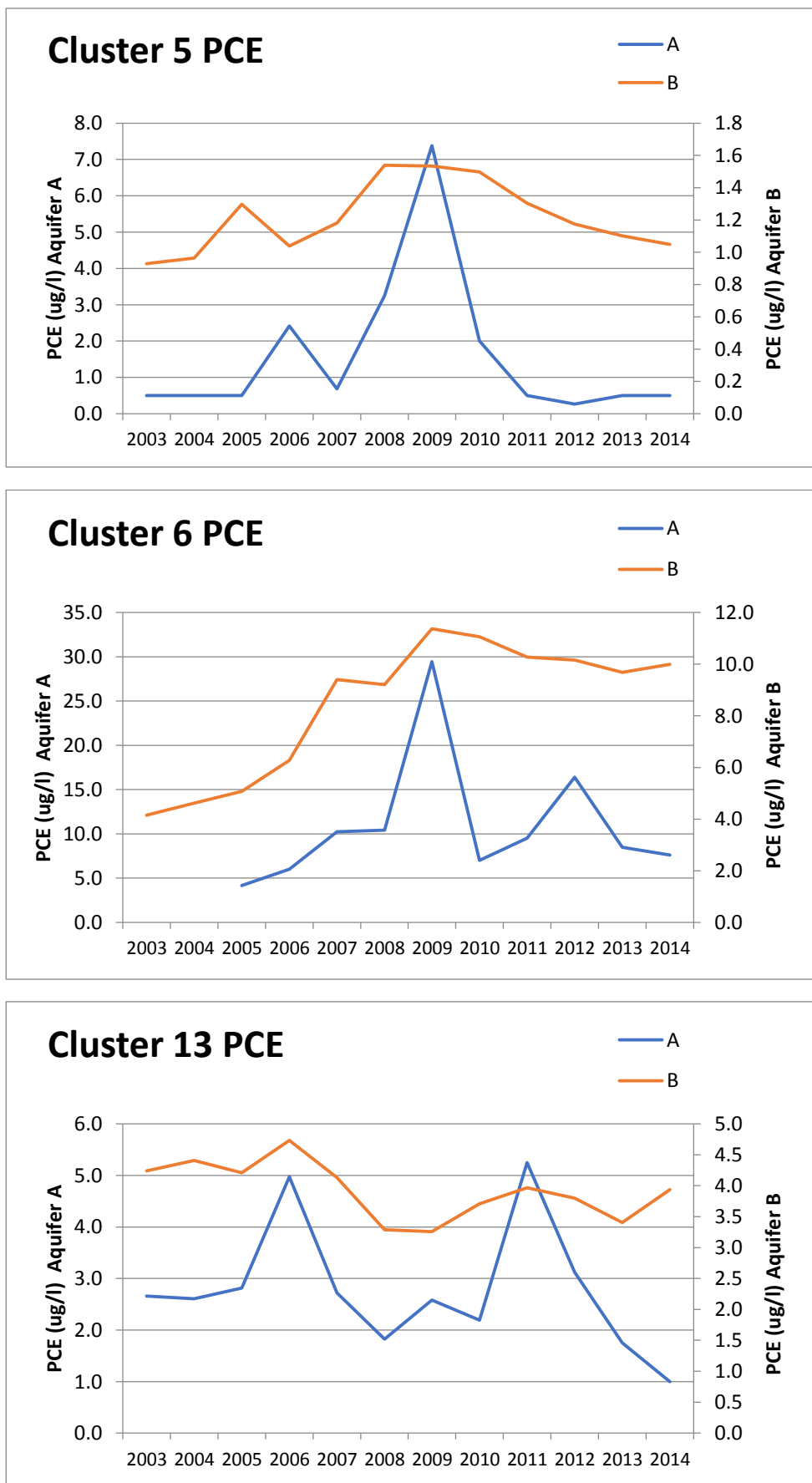


Fig. 8 – PCE water quality profiles in the three main clusters: the annual average concentration profiles are compared between the unconfined (A) and the semiconfined aquifer (B).

Figure 9

[Click here to download Figure: Figure 9.docx](#)

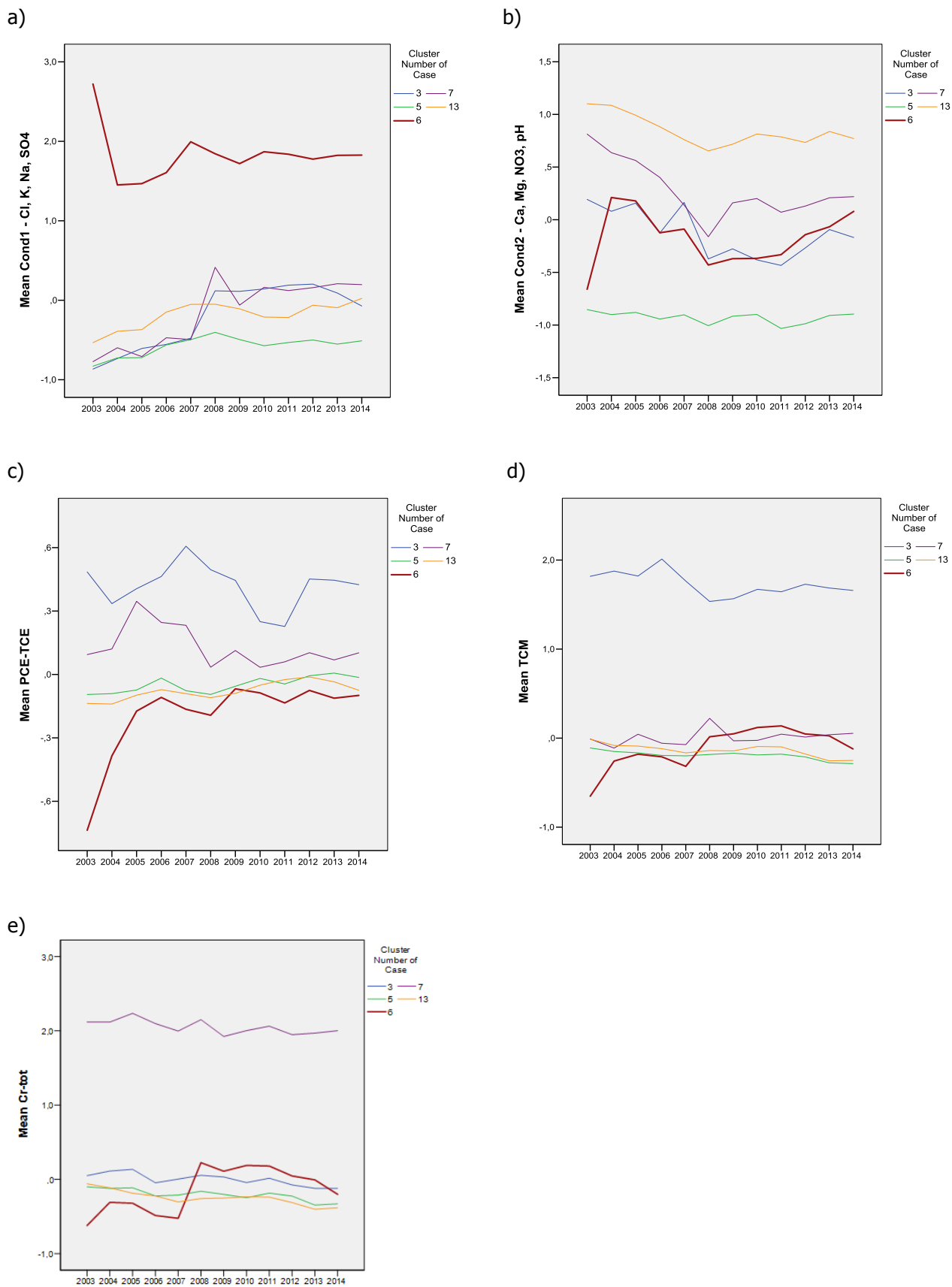


Fig. 9 – Water quality profiles of the five main clusters: a) Conductivity component 1 (Cl, K, Na, SO<sub>4</sub>); b) conductivity component 2 (Ca, Mg, NO<sub>3</sub>, pH); c) PCE/TCE component; d) TCM component; e) Total chromium.

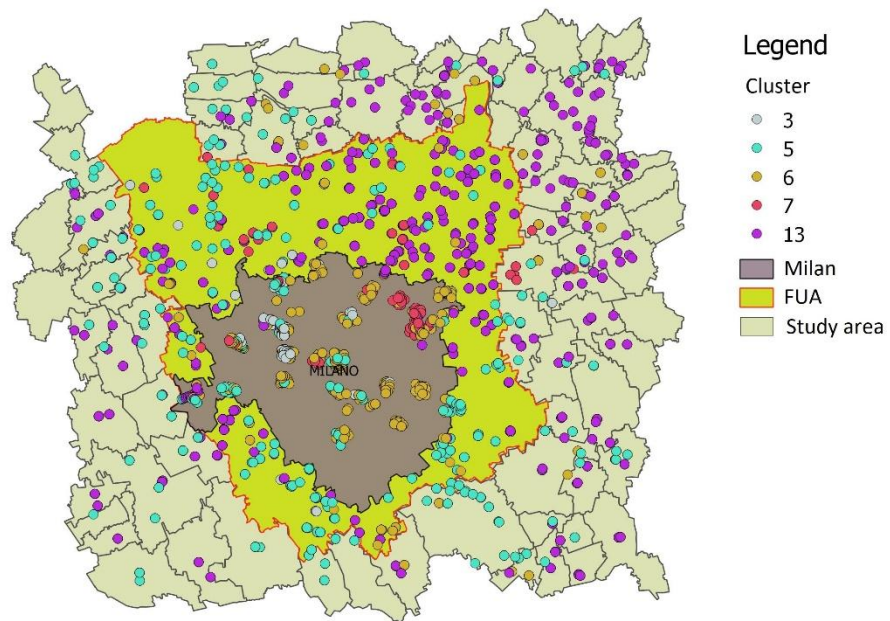


Fig. 10 – Map of the borehole database through the FUA of Milan; Different colours refer to different clusters

# Figure 11

[Click here to download Figure: Figure 11\\_rev.docx](#)

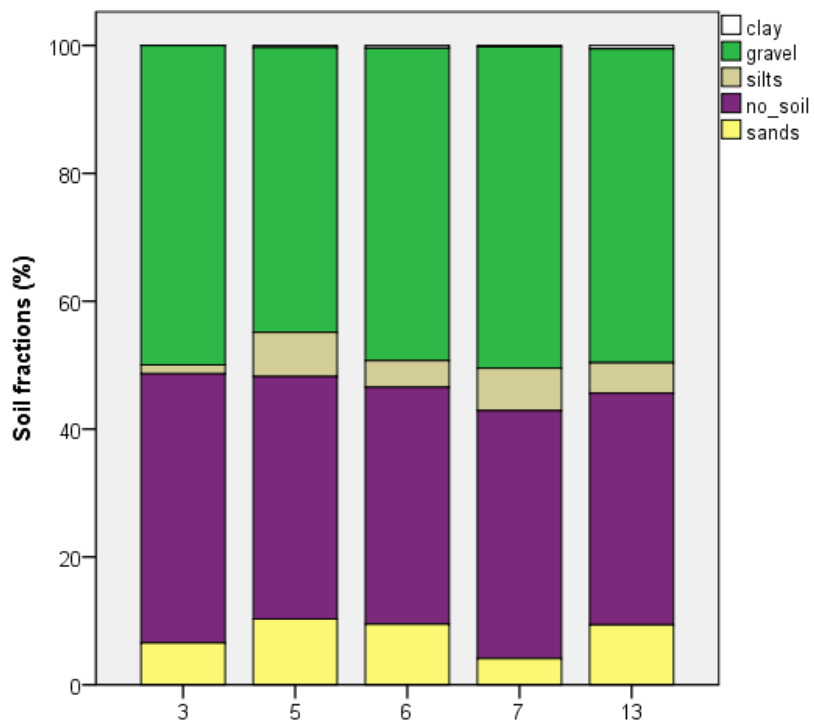


Fig.11 – Lithological composition (%) of the five main clusters.



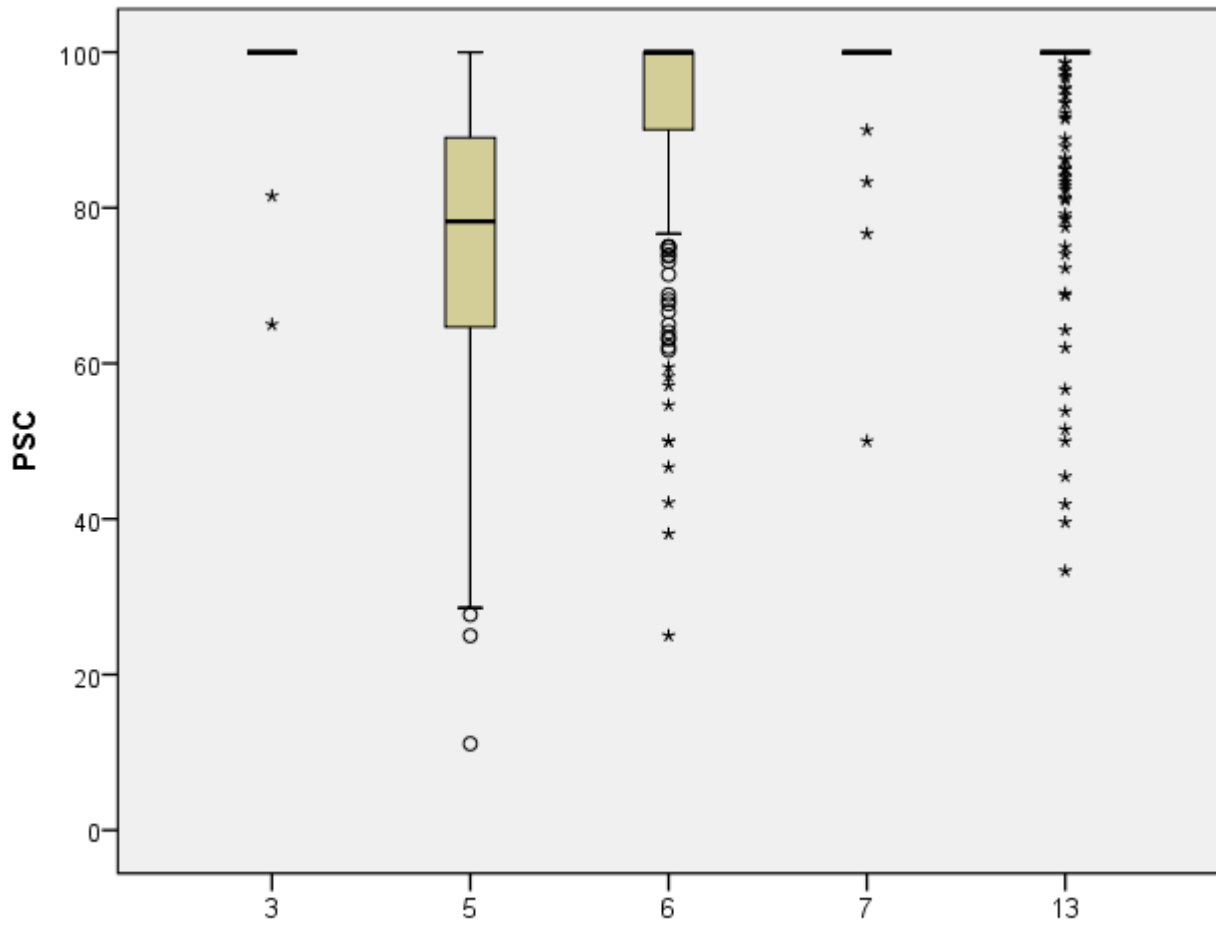


Fig. 12 – Percentage Similarity Index of the Lithological characteristics of five clusters.

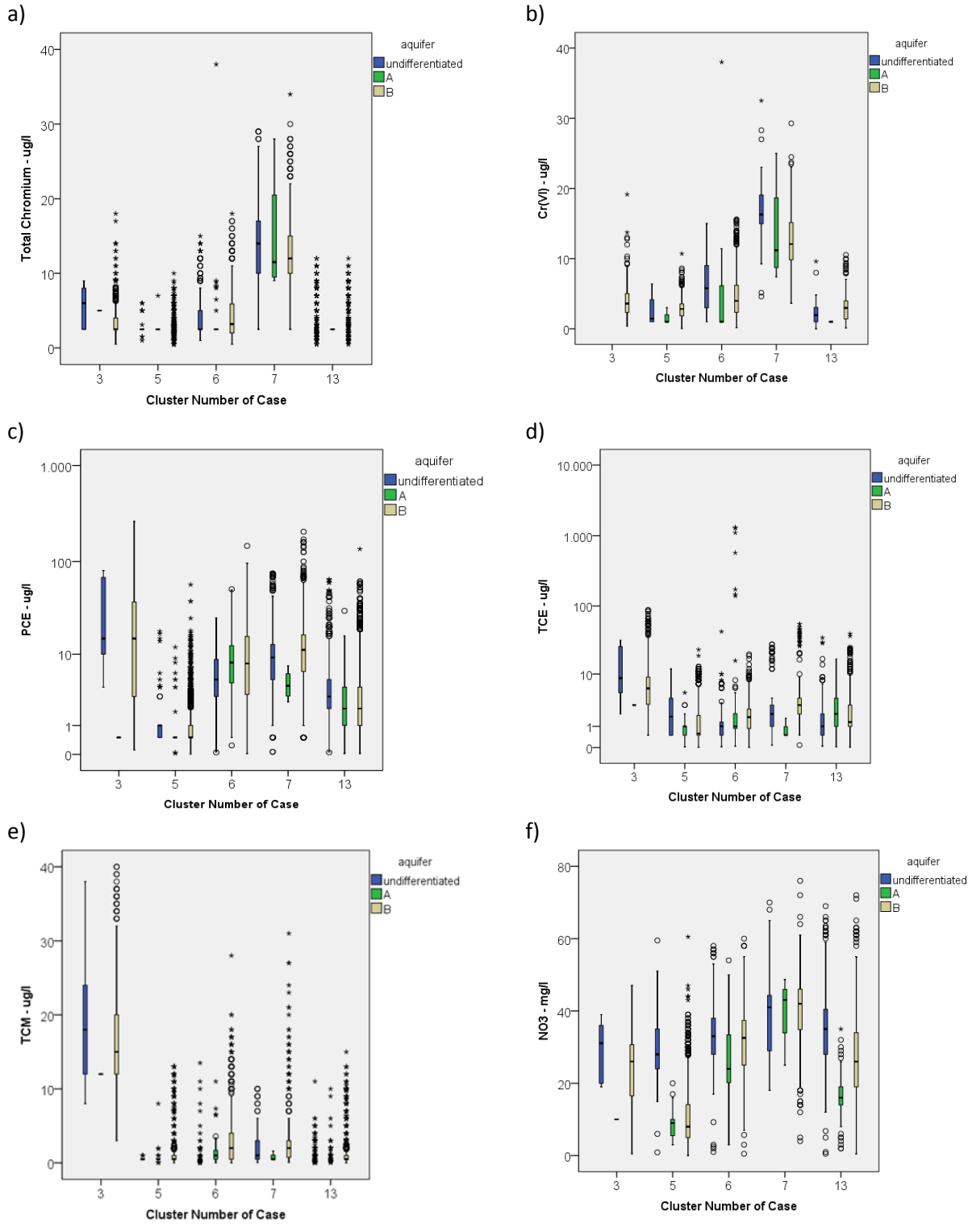
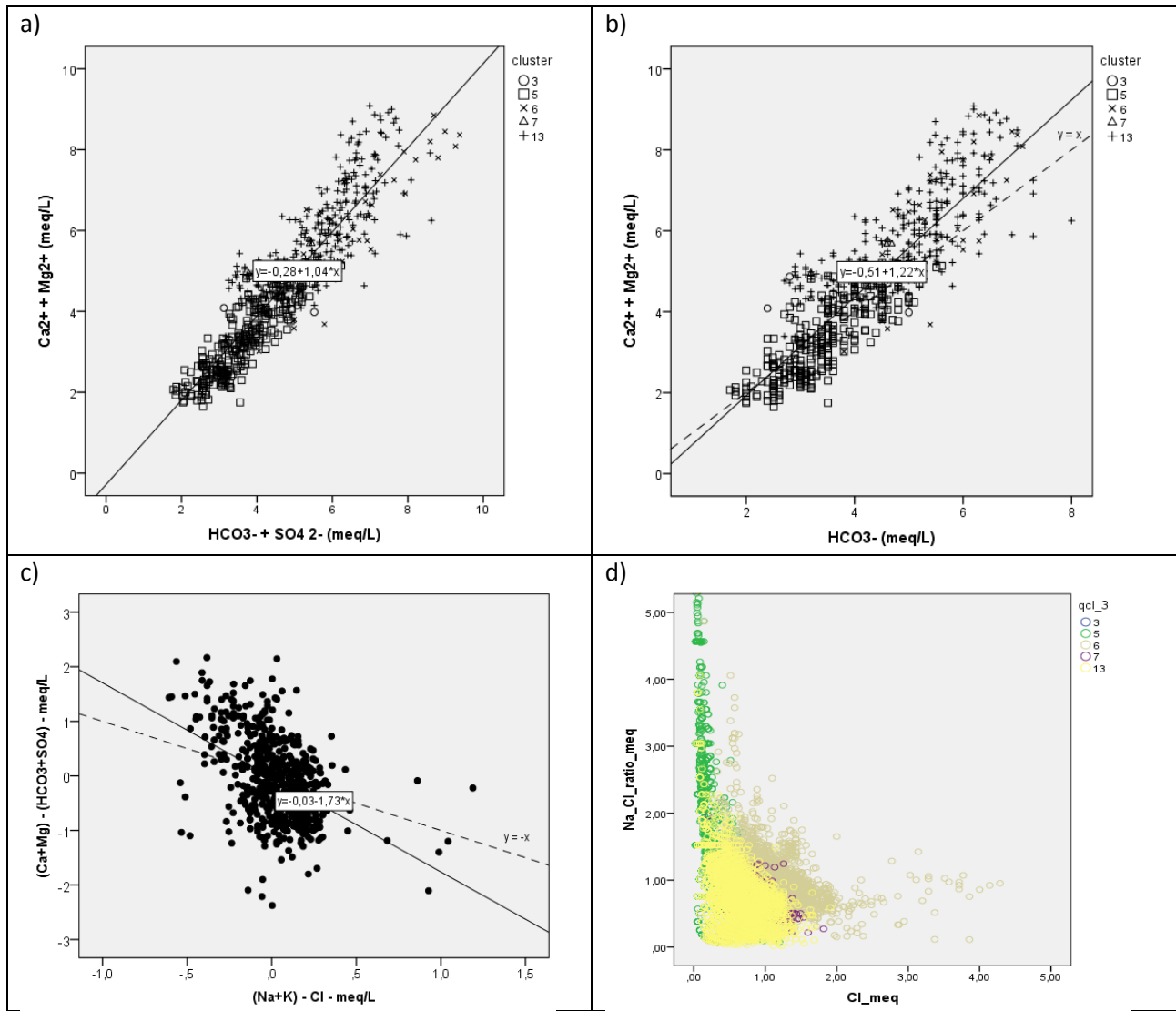


Fig. 13 – Comparison of diffuse pollutant background values among clusters: a) Total-Cr; b) Cr(VI); c) PCE d) TCE; e) TCM; f) NO3.

Figure 14  
Click here to download Figure: Figure 14.docx



**Supplementary material for on-line publication only**

[Click here to download Supplementary material for on-line publication only: Table S1 undifferentiated\\_rev.docx](#)

**Supplementary material for on-line publication only**

[Click here to download Supplementary material for on-line publication only: Table S2 unconfined\\_rev.docx](#)

**Supplementary material for on-line publication only**

[Click here to download Supplementary material for on-line publication only: Table S3 semiconfined\\_Rev.docx](#)