DOI: 10.7343/as-2017-301

Statistical methods and transport modeling to assess PCE hotspots and diffuse pollution in groundwater (Milan FUA)

Metodi statistici e modellistici per la determinazione delle componenti puntuali e diffuse di inquinamento da PCE nelle acque sotterranee di Milano

Loris Colombo

Riassunto: Il Decreto Legislativo 152/06 ha adottato i principi delle Direttive Europee nel campo delle acque sotterranee e ha delegato alle Regioni il compito di identificare le aree soggette ad un inquinamento diffuso. Nella Pianura Padana, la qualità delle acque sotterranee è principalmente condizionata dalla presenza di industrie e attività antropiche. Lo scopo di questo lavoro è stato quello di valutare l'inquinamento diffuso da tetracloroetilene (PCE) nell'area Allargata di Milano (Functional Urban Area, FUA) partendo dal corposo dataset di valori di concentrazione raccolti tra il 2003 e il 2014 da enti di controllo e gestori per il monitoraggio qualitativo delle acque sotterranee. Per questo, si è sviluppata una nuova metodologia che utilizza sia metodi statistici che deterministici. Dapprima, mediante una analisi cluster (CA) applicata al dataset di partenza, sono state identificate le sorgenti negli acquiferi superficiale e semiconfinato dell'area. E' stato poi implementato un modello numerico di trasporto per studiare l'estensione dei pennacchi e individuare i pozzi e i piezometri colpiti da inquinamenti provenienti da sorgenti puntuali. Alla luce dei risultati ottenuti, è stato costruito un nuovo dataset contenente esclusivamente i dati imputabili alla componente diffusa dell'inquinamento e dunque non correlabili ai pennacchi simulati. Tramite "*kriging*", metodologia usata per il trattamento geostatistico delle variabili regionalizzate, sono state prodotte mappe di concentrazione per identificare le aree oltre la concentrazione soglia di contaminazione (CSC - D.Lgs 152/06 – 1.1 μg/l). La combinazione delle mappe di contaminazione con le statistiche descrittive dei principali cluster, ottenute da un'analisi statistica multivariata del dataset di partenza, ha fornito un valido strumento di analisi della distribuzione della contaminazione diffusa da PCE (con valore mediano intorno a 10 μg/l) nell'area di studio. Lo sviluppo infine di una metodologia stocastica di modellazione, ha permesso di considerare le incertezze legate alle sorgenti multiple (rilascio di massa) e all'eterogeneità delle variabili idrogeologiche (conducibilità idraulica), al fine di identificare le aree dove il rilascio di massa contaminante negli acquiferi è più consistente e dove dunque con maggior probabilità è possibile individuare le sorgenti puntuali non note.

Keywords: *urban groundwater, numerical modeling, solute transport, diffuse contamination, inverse iterative modeling.*

Parole chiave: idrogeologia urbana, modellazione numerica, trasporto di soluti, contaminazione diffusa, modellazione inversa.

Loris COLOMBO
Politecnico di Milano
DICA - Civil and Infrastructure Engineering Department
P.zza Leonardo da Vinci, 12 Milano, Italy
loris.colombo@polimi.it

Ricevuto/*Received*: 29 September 2017-Accettato/*Accepted*: 09 November 2017
Pubblicato online /*Published online*: 12 December 2017

Abstract: *The Italian law 152/2006 adopted the EU Water Framework Directive principles and delegated to the Regions the task of identifying areas subject to groundwater diffuse pollution. In the Lombardy Plain, the qualitative groundwater conditions are affected mainly by the presence of industries and anthropic activities. The aim of this work was to assess tetrachloroethylene (PCE) diffuse pollution in the Milan Functional Urban Area (FUA), where chlorinated solvents are the main groundwater contaminants and the results of monitoring campaigns for the years 2003-2014 were collected in a dataset. For this purpose, a new methodology was implemented both in a deterministic and stochastic process. At first, hotspots were identified through Cluster Analysis (CA) applied to concentration values collected in unconfined/confined aquifers (2003-14). Then, a numerical transport model was implemented to study the hotspot plume extension in reason to identify monitoring wells not affected by diffuse pollution but related to specific hotspot sources. Consequently, it was possible to erase these data from the whole initial dataset in order to have a new one containing only diffuse concentrations. Interpolating them through ordinary kriging, PCE iso-concentrations maps identified areas where values are over the Maximum Contaminant Level (1.1 μg/l, Italian Law 152/06). Considering descriptive statistics and iso-PCE concentration maps, a median PCE value estimation (10 μg/l) was find as representative of PCE diffuse contamination Milan city. Moreover, a stochastic methodology was used in order to consider uncertainties due to unknown multiple-sources and environmental heterogeneity. The innovative approach gave some interesting solutions to point out areas where the contaminant mass release is higher and where high probability unknown sources can be found.*

Introduction

The problem of groundwater resources contamination in highly urbanized areas, during the last two decades, is one of the most important environmental issues at both European (European Union, 2006) and National level. In Italy, the Po plain, and in particular the Lombardy Region, is one of the most densely populated areas where human activities have caused a high impact on the groundwater quality. Nowadays, thanks to EU legislation, public authorities drive continue environmental monitoring activities that led to the creation of databases available for the study and resolution of environmental protection problems. In recent years, the new National and Regional regulations consider the necessity to develop plans for the remediation and management of the most industrialized areas. Here, groundwater is affected by contamination due both to point sources (PS, associated with medium dimension sources, i.e. hotspots) and multiple point sources (MPS, constituted by a series of unidentifiable small sources clustered in a large area and causing a diffuse contamination).

The latter category predominates in European Functional Urban Areas (FUA) and cannot be managed with remediation techniques such as those commonly used for large/medium contaminated sites, mainly because of the difficulty to identify many different source areas that release small contaminant mass. Consequently, the usual remediation procedures are not economically sustainable and often fail to provide results in an acceptable time frame. The Italian Law 152/2006 holds the definition of diffuse contamination as "a physical-chemical alteration of environmental medium due to diffuse sources and not linked directly to known and unique source". However, nor the National legislation nor the contents of the Regional plans provide useful methodologies to quantify this phenomenon, making impossible to distinguish the PS and MPS pollution existing in urban areas. Moreover, also in scientific literature there are few examples (Frumkin 2002; Nolan et al. 2002; Stevenazzi et al. 2017), mostly developed for the assessment of diffuse pollution spread linked to anthropogenic practices carried out over large areas, such as agricultural ones (e.g. Fertilizers, herbicides). On the contrary, in the case of urban areas, it is not possible to determine diffuse pollution sources because the contamination is due to multiple sources, too small to be precisely identified and removed.

The aim of this work is to give a contribution to find an integrated approach (statistical and deterministic mathematical modeling combined with statistical modeling) in order to:

- distinguish PS (hotspot) and MPS (diffuse) contamination in urban groundwater throughout multivariate and K-means statistical approach and transport numerical modelling with MODFLOW (Harbaugh and McDonald 1996;McDonald and Harbaugh 1988) and MT3DMS (Zheng and Wang 1999);
- estimate the contamination level for those areas mainly connected with a diffuse contamination by using kriging;
- identify the urban areas, which have the highest probability to be linked to MPS by stochastic particle back-tracking analysis or that can contribute to the contaminant mass inflow.

Study area

Milan FUA is one of the most densely populated areas in Lombardy (north Italy). In particular, in the FUA (composed by the city of Milan and some neighbouring municipalities such as Monza and Sesto San Giovanni) about 4.000.000 inhabitants live within 1120 km^2 (Fig. 1). The northern area of FUA is characterised by a dense agglomeration of companies where, especially near the Milan City, (from '50) many industries such as automotive, refineries, chemical plants, still and tires production are historically located (Provincia di Milano 1992). Because of the high hydraulic conductivity (10^{-4}-10^{-3} m/s) and the high groundwater withdrawal rate (18 m^3/s, Gattinoni and Scesi 2017), Milan represents a drainage area of groundwater and many pollutants flow into municipality. The geological build up delineates 2 main aquifers (Group A and Group B) in the study area divided by an aquitard (thickness of 10 meters of 10^{-8} m/s low permeability) (ARPA Lombardia 2015; Carcano and Piccin 2002; Francani and Beretta 1995). The shallow aquifer (named Group A) is composed by a gravel sand material with a relative high hydraulic conductivity (10^{-3} m/s) whereas the semiconfined aquifer (named Group B) is constituted by fine sand and gravel (10^{-4} m/s). The separation of two aquifers becomes more and more discontinuous in the norther part of Milan (in Fig. 1b the limit of aquitard is the black line). Because of this hydrogeological conformation, mainly in the northern area of Milan, since '50 groundwater quality has been strongly compromised by development of several plumes that sometimes overlap each other contributing to the spread of diffuse contamination with concentration values that exceed the national threshold values of PCE (1.1 μg/l). PCE is a manufactured chemical and does not occur naturally in the environment. Slow natural biodegradation of PCE may occur under anaerobic conditions when microorganisms are acclimated. However, the biodegradation process degrades PCE to TCE and eventually to vinyl chloride, which are also considered human carcinogens.

Materials and methods

The main goal of this paper is to propose a methodology useful to explore big datasets using statistical tools and to couple the dataset analysis with a numerical transport model. The applied methodology consists of a combination of different steps:

a. data collection and preparation (a large database must contain concentration of pollutants, characteristics of monitoring network and a documental research on contaminated sites) in order to have a robust statistical treatment of data (outlier detection, errors and missing values);

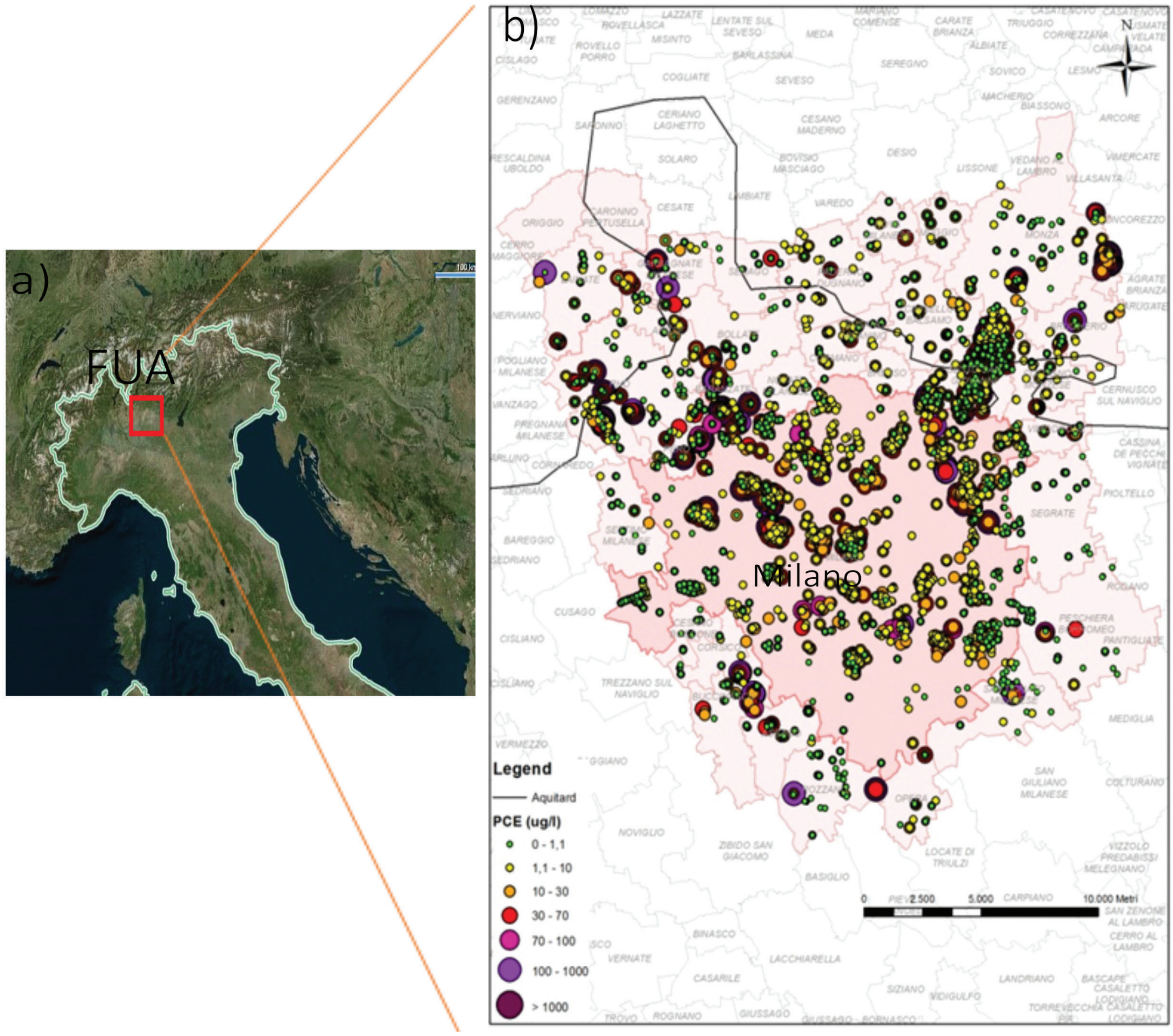

Fig. 1 - a) Location of the study area; b) FUA of Milan with PCE samplings (2003-2014). The pink area was considered by numerical modeling.

Fig. 1 - a) Ubicazione dell'area di studio; b) Area Funzionale Urbana di Milano con i dati misurati di PCE (2003-2014). Il colore rosa rappresenta l'area oggetto di modellazione.

b. application of multivariate analysis in order to evaluate dataset Principal Components (PCs) and different levels of diffuse contamination;
c. Cluster Analysis (CA) application to an univariate set of concentration PCE data in order to separate the hotspots (PS) from the diffuse pollution (MPS);
d. plume simulation (MODFLOW/MT3DMS) starting from the identified PS. The evaluation of the plume extensions allows to separate monitoring data respectively linked to PS and to MPS;
e. geostatistical analysis in order to assess diffuse contamination considering the dataset without concentration values related to the PS;
f. innovative stochastic numerical modelling for the simulation of diffuse pollution.

All these tools estimated the magnitude of diffuse pollution in Milan FUA both for shallow and semiconfined aquifer. The following sub-sections present the most important steps for the application of the methodology to PCE contaminant.

Data collection and preparation

Initial database consists of more than 44000 hydro-chemical data related to the period 2003-2014 and detected in 3458 wells/piezometers (see Fig. 1b). Moreover, the historical data for the period 1960-2000 were very useful to understand the history of the hot-spot sources. All data were provided by ARPA Lombardia (Regional Environmental Agency) and included in a large database with all available information (name of monitoring point, depth, screening positions, concentration values and other qualitative information about groundwater such as ions).

Cluster and multivariate analysis for the identification of hotspots and outliers

CA (for detail see Afifi et al. 2003; Fabbris and Gallo 1993), applied to the univariate PCE concentration dataset in Milan FUA (45602 samples during the period 2003-2014) was extremely useful to identify outliers and contamination hotspots. Though many methods are presented in literature

(Rosner 1975) for detecting outliers, in this work once the hotspots were identified, a detailed analysis for each monitoring point (i.e. piezometer) was done in reason to delete the true outliers (singular concentration values 10 times higher than the time series average concentration) from the dataset. Furthermore, the hot spots obtained with CA were compared with potential contaminant sources located into FUA (database of contaminated sites is provided by ARPA Lombardia).

Multivariate data analysis (Arain and Pheng 2006; Debic et al. 2014; Mouron et al. 2006; Reghunath et al. 2002) may help to quantify the influence that both types of variabilities (explained-when phenomenon is understood and unexplained when phenomenon is unknown) have on a system so that it can be better understood.

PC Analysis, essentially, is a one-sample technique applied to data with no groupings or divide among the observations. When the number of independent variables is large respect to the number of observations or the independent variables are highly correlated making unstable the estimates of repression coefficients, principal components are used to reduce the number of dimensions of the dataset. In this study, because PCA is strongly affected by missing values, parameters characterized by a large number of missing observations are discarded from the further analysis. As consequences, PCA was applied only on 13 water quality constituents represented into the Tab. 1. (Ca, Cl, specific conductance, Cr, Mg, Nitrates, pH, K, Na, Sulphates, Tetrachloroethene (PCE), Trichloroethene (TCE), Trichloromethane (TCM)).

Numerical transport model

CA identified the hotspots and a documental researches supplied the information about contamination sources and site history (Fig. 3b). These are the fundamental elements for a development of a numerical transport model. For simplicity, 3 different scenarios (Fig. 2) have been considered as possible:

1. the hotspot and the contaminated site overlap (Fig. 2.1);
2. the hotspot is downgradient the contaminated site (Fig. 2.2);
3. no contaminated sites are associated with hotspot (Fig. 2.3).

In the first and second case, the plume simulation was possible by transport numerical model, whereas in the last case a particle tracking could help to identify areas for new investigations.

A deep documental research on contaminated sites and potential sources could supply many essential information to simulate the source in the numerical transport model, but it is not enough if the information is punctual and not related to the monitoring network nearby the source. A good monitoring network dataset, downgradient the source, is as much essential as the good site history knowledge to perform a realistic description of fate and transport of the contamination by numerical transport model.

The three-dimensional finite-difference groundwater model (MODFLOW) of the Milan FUA was implemented to simulate the PCE contamination transport related to the sources identified through CA. The model quantitatively estimated the extension of the most important plumes (MT3DMS) in the FUA under steady- state conditions,

Tab. 1 - Summary statistics of the groundwater quality constituents used for the PCA.

Tab. 1 - Statistiche descrittive dei principali elementi chimico-fisici della qualità delle acque sotterranee usati per la PCA.

	N		Mean	Median	Std. Deviation	Min	Max
	Valid	Missing					
Bicarbonate (μg/l)	1447	58328	262.66	256.00	78.84	2.50	662.00
Ca (μg/l)	40316	19459	73.34	72.00	48.70	0.50	4000.00
Cl (μg/l)	43773	16002	16.31	13.20	13.05	0.01	303.00
Specific conductance (Sm)	39468	20307	460.82	455.93	170.41	0.35	2896.00
Cr-VI (μg/l)	14794	44981	16.29	4.00	198.53	0.00	8022.90
Total-Cr (μg/l)	45552	14223	9.09	2.50	150.00	0.00	12300.00
Mg (μg/l)	39396	20379	15.85	16.00	5.72	0.50	249.00
NO_3^-(μg/l)	46716	13059	22.75	21.30	14.43	0.01	190.06
DO (μg/l)	176	59599	12.82	8.16	11.37	0.16	68.40
K (μg/l)	38667	21108	0.91	0.50	0.71	0.10	40.00
Dry residue	36352	23423	318.01	313.14	106.00	0.25	909.00
Na (μg/l)	39259	20516	8.74	6.50	7.42	0.50	214.00
SO_4^{-2} (μg/l)	43915	15860	30.99	30.00	27.49	0.01	1363.00
PCE (μg/l)	44912	14863	14.87	3.00	310.32	0.00	37800.00
TCE (μg/l)	44569	15206	13.97	1.00	261.90	0.00	14000.00
TCM (μg/l)	43541	16234	3.55	0.50	88.30	0.00	17214.00
pH (-)	39603	20172	7.66	7.70	0.27	5.56	9.56
Temperature (°C)	4041	55734	14.92	14.90	1.51	7.70	26.60

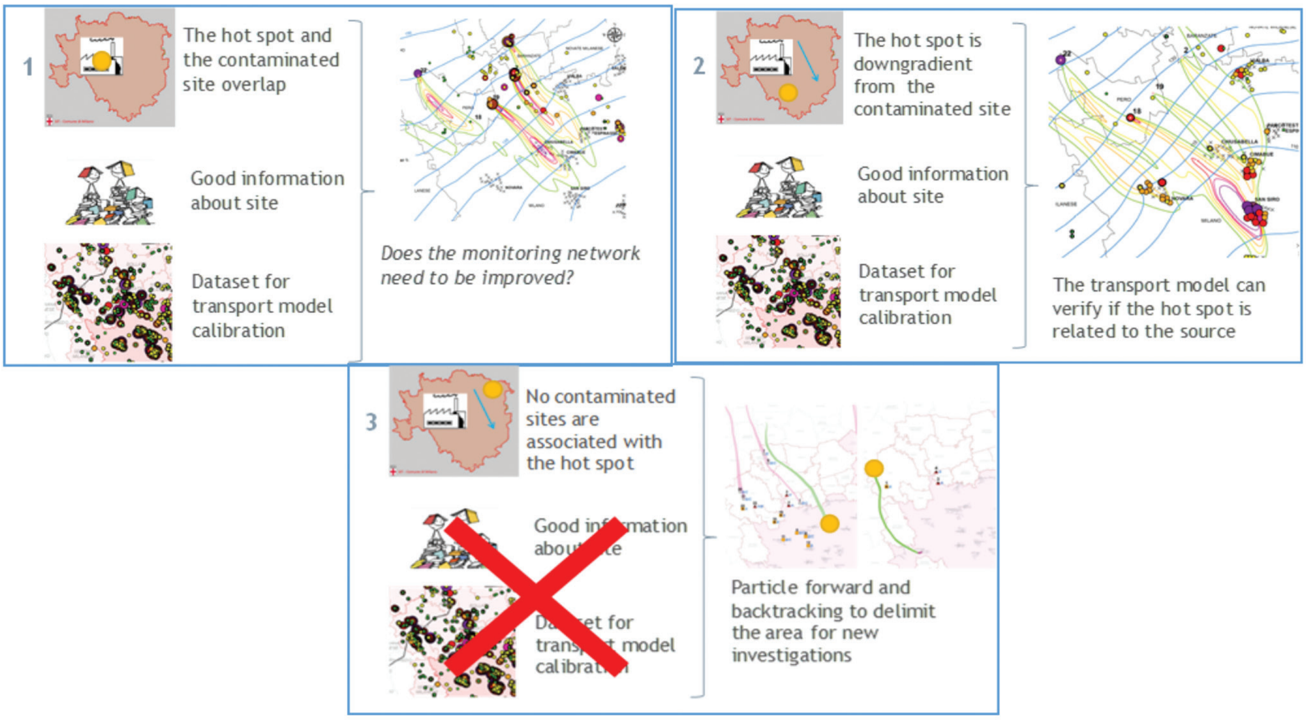

Fig. 2 - 1) hotspot and contaminated site overlapping. Sources were simulated into numerical transport model in order to delimit the plume extension 2) hotspot downgradient from the contaminated site and the plumes moved downgradient the source 3) no contaminated sites related to hotspots allowed only a particle tracking in order to identify the flow line interested by contaminant.

Fig. 2 - 1) hotspot e sito contaminato coincidenti. Le sorgenti sono state simulate nel modello numerico di trasporto per delimitare l'estensione dei pennacchi 2) hotspot a valle di siti contaminati. Il pennacchio si sviluppa a valle della sorgente 3) nessun sito contaminato coincide con gli hotspot. Solamente un tracciamento delle particelle è utile per determinare le linee di flusso interessate dalla contaminazione.

considering the groundwater contamination status in 2014.

The horizontal domain of the model was defined by a 20 m x 20 m grid and covered an area of about 1120 km^2. The geometries of aquifers, provided by previous works (Carcano and Piccin 2002), were implemented in the model considering three layers: Layer 1 (Group A) with a 10-60 m average thickness, Layer 2 (aquitard) with a 2-10 m thickness and Layer 3 (Group B) with a 60-120 m average thickness. The available log-stratigraphies of the ARPA Lombardia database were used to calculate the equivalent hydraulic conductivities considering the thickness of each layer and considering specific hydraulic conductivity value (K) for each grain size class. The vertical anisotropy ratio was assumed equal to 10% (Anderson and Woessner 1992). Computed values were initially assigned to the model and finally calibrated with an inverse procedure using pilot point technique (Doherty et al. 2005). The regular grid of pilot point was supplemented with individual pilot points in correspondence with pumping tests (30) where the value of hydraulic conductivity could considered with a low uncertainty. For this reason, a narrow range were assigned to these special pilot points (+/-50% of pumping test value) whereas for the regular grid range was between 10^{-5} and 10^{-2} m/s. BCs boundary condition was assigned considering the hydraulic head values collected in May 2014 whereas internal boundary conditions of the model involved main streams (Lambro and Seveso), withdrawals (public and private) and recharge (three different zones were considered respectively urban (10^{-9} m/s), irrigative ($2*10^{-8}$ m/s) and green ($1.2*10^{-8}$ m/s). For more details (Alberti et al. 2016; ARPA Lombardia 2016).

Once the flow model was calibrated, a PCE transport model for only hotspots was developed with MT3DMS. The simulations were divided in six stress periods (each of a decade length, from '50 to nowadays) in order to reconstruct the history of sources with constant concentration time variant conditions. The calibration procedure considered all available dataset from 2003 to 2014 to represent groundwater chemical status in 2014. It consisted on changing the parameters (Tab. 2), based on literature previous works (Gehlar et al. 1992; Hill and Tiedeman 2007), that influence the advection-dispersion equation of the PCE (i.e. dispersivity coefficient, distribution coefficient Kd and half-time life).

Tab. 2 - Hydrodinamic calibrated parameters for the transport model.

Tab. 2 - Parametri idrodinamici utilizzati per il modello di trasporto.

Dispersivity (m)			f_{oc} (-)	k_d (m^3/kg)	$t_{1/2}$ (year)
20	3	0.03	0.001	$4.2*10^{-4}$	10

Diffuse contamination assessment: geostatistical maps and diffuse contamination values

Once the plumes were modeled, from the chemical dataset were excluded those points affected by the plumes and whose quality was obviously not determined by diffuse pollution but from a PS contamination. Using kriging interpolation, different zones in FUA were identified with different levels of diffuse contamination respectively in aquifer A and aquifer B.

Combining maps with PCs - CA statistical results and referring to the two critical diffuse pollution zones with values higher than 1.1 µg/l, the summary statistics can be assumed as reference values for diffuse contamination (ARPA Lombardia 2016).

Here, two approaches were used:

1. B1: considering the dominant cluster obtained with PCA (herein with higher frequency in each different kriged zones of diffuse contamination), its summary statistics (50° percentile for yellow zone and 75°percentile for red zone) were used in order to assess the threshold diffuse contamination values for FUA;
2. B2: considering all clusters in each different kriged zones of diffuse contamination, an average value weighted on different clusters (50° percentile for yellow zone and 75°percentile for red zone) frequency provided the threshold diffuse contamination values for FUA.

Stochastic numerical modeling assessing diffuse contamination

Diffuse contamination is clustered in a large area due to the presence of MPS which are unidentifiable and very small. Because of the uncertainty related to the exact position and strength of MPS, it is hard to implement a numerical model able to simulate the fate and transport of a diffused contamination. To overtake this problem, a numerical stochastic model (code MODFLOW/MT3DMS) was implemented in a pilot area located in the North-eastern part of the Milano FUA. The deterministic model was previously developed and calibrated in steady state condition (ARPA Lombardia 2015). The proposed methodology allowed to consider the uncertainties linked to the diffuse contamination sources (MPS) using a Monte Carlo (MC) procedure (Doherty 2014; Tonkin et al. 2007; Tonkin and Doherty 2009). Several calibrated models were generated considering the effect of some parameters (K hydraulic conductivity heterogeneity and mass released from unknown sources) governing groundwater flow and transport.

Two stochastic approaches were developed and compared:

1. particle backtracking (BT): through 400 MC realizations considering variability and uncertainty of K (within a range of plausible interval 10^{-5} -10^{-2} m/s). Using MODPATH (Pollock 1994), placing particle starting points where a PCE measurement diffuse concentration is available, it was possible to highlight, in terms of frequency, the cells crossed by a high number of particles, i.e. the cells that most probably can host a MPS;
2. clustered MPS: through 100 MC realizations considering variability of contaminant mass released into the shallow aquifer (computed as a product of concentration and Darcy's flow is within a range of 10^{-2} – 1 µg/s). Using MT3DMS, for each domain sector, it was possible to assess in a probabilistic way, the MPS contaminant distribution and the mass releases frequency (Alberti et al. 2017).

From many different realizations all calibrated within NSMC (Null Space Monte Carlo suite in PEST), it was therefore possible to provide results in terms of frequency of occurrence that identified areas of high probability to find MPS.

Results and discussion

Statistical results

Statistical tool provided two main results:

- CA allowed to separate the PS from the MPS (Fig. 3a, on the next page.);
- PCA/FA transformed the dataset containing the 13 variables (analytical constituents in Tab. 1) interrelated or correlated to various degrees, to a new dataset containing 5 new orthogonal, uncorrelated variables called principal components (PCs).

Factors loadings are the standardized regression coefficients of the multiple linear regression characterizing the components (absolute loading values higher than 0.6 are significant and are identical to the correlation coefficients). PCE and TCE are the variables with the highest correlation within the third component (0.85) whereas Ca, Mg, nitrates and pH are the variables with the strongest correlation within the second component. Finally the first component, contain higher correlations (0.75-0.80) between monovalent ions (K, Na, Cl, SO_4^{2-}).

A new CA was applied to multivariate set of data considering the 5 PCs as variables. Five clusters (3, 5, 6, 7, 13) were identified as representative of diffuse contamination as shown in Fig. 4 as they represent an homogeneous sub-set of database which have the same multi-parameter characteristics (i.e. they are the most populated clusters and they have an average PCE concentration value next to zero).

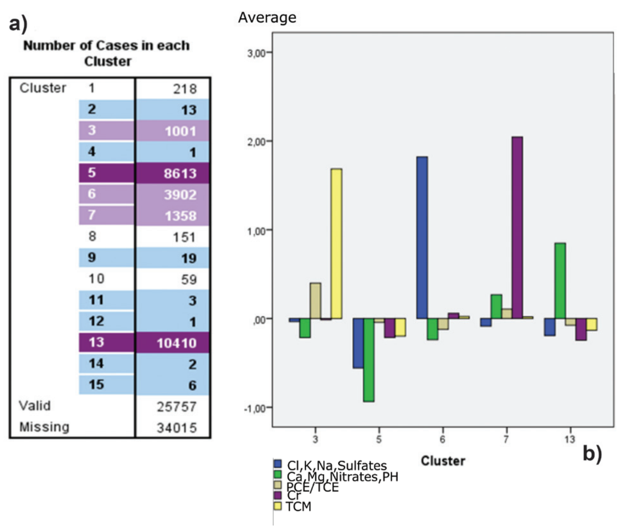

Fig. 4 - a) Clusters identified with cluster analysis after PCA. Ten clusters are representatives of hotspots (samples with high concentration are less than 1%). Only 5 clusters (violet) are representatives of a diffuse contamination b) characteristics of mean PCE-TCE centroids cluster obtained through the multivariate k-means (ARPA LOMBARDIA 2016).

Fig. 4 - a) Cluster individuati dopo la PCA. 10 cluster sono rappresentativi di sorgenti puntuali (i campioni con alte concentrazioni sono minori del 1%). Solo 5 cluster sono rappresentativi invece di una contaminazione diffusa b) valore medio dei centroidi dei cluster diffusi PCE-TCE ottenuti mediante l'analisi multivariata (ARPA LOMBARDIA 2016).

DOI: 10.7343/as-2017-301

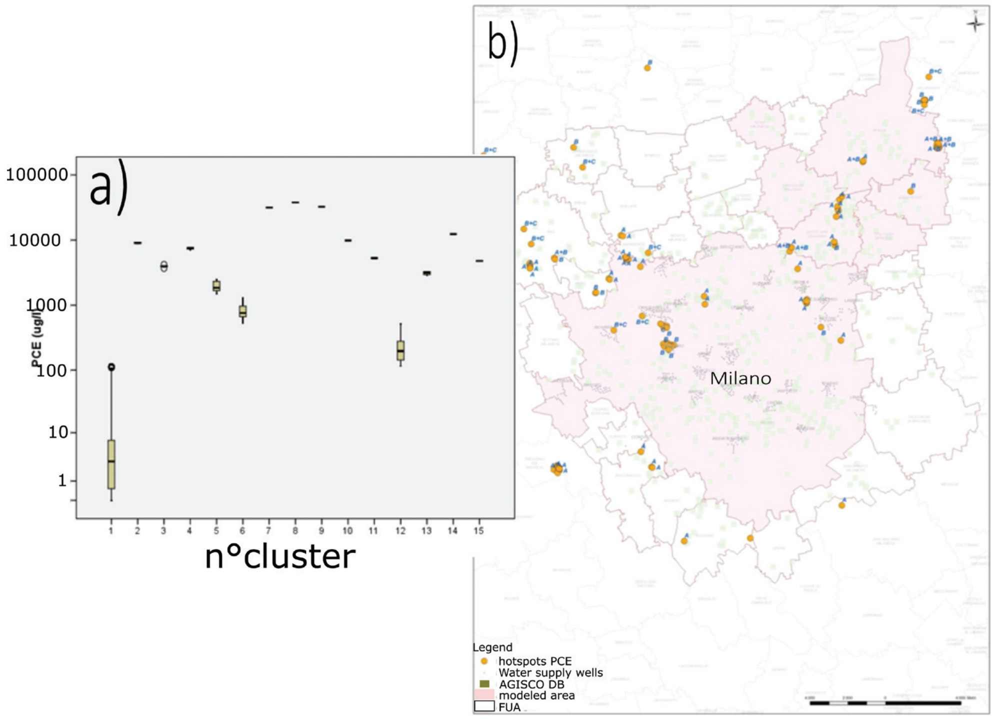

Fig. 3 - a) Box plot in a logarithmic scale of the 45602 PCE measurements (µg/l) subdivided into 15 clusters (k-means). The cluster 1, the highest populated ones, refers to PCE values lower than the PCE values of the other clusters. This could be considered as representative of the diffuse PCE contamination whereas the others 14 may be possibly attributed to hotspots b) the clusters found with cluster analysis are compared with potential contaminant sources identified by documental research (ARPA LOMBARDIA 2016).

Fig. 3 - a) Box-plot in scala logaritmica dei 45602 valori misurati di PCE (µg/l) suddivisi in 15 cluster (k-means), Il cluster 1 che è il più popoloso si riferisce ai valori di PCE inferiori a tutti gli altri e può essere considerato come rappresentativo dell'inquinamento diffuso da PCE mentre i restanti possono essere attribuiti a sorgenti puntuali b) confronto dei cluster con le sorgenti potenziali di contaminazione identificati mediante ricerca documentale (ARPA LOMBARDIA 2016).

Modelling results

The deterministic numerical flow model calibrated showed a good fit to measurements (Fig. 5) with an absolute residual mean (ARM) of the hydraulic heads less than 1 meter for the shallow aquifer (blue dots) and less than 2 meters for the semiconfined one (orange dots).

Considering for simplicity a pseudo-steady state flow model and assuming that the effect of change of flow lines have a little effect on the plumes, the transport model once calibrated with parameters showed in Tab.1, have simulated plumes for those hotspots obtained with CA (Fig. 6a plumes for shallow aquifer and Fig. 6b plumes for semiconfined aquifer).

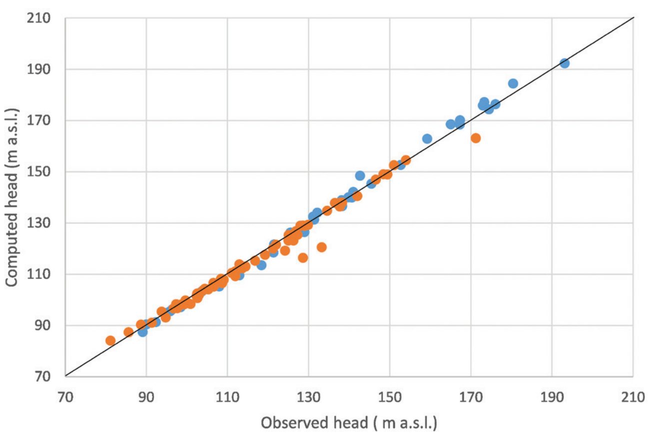

Fig. 5 - Observed vs computed head for targets in shallow aquifer (blue dots) and semiconfined one (orange dots). The residual standard deviations are respectively about 1.2 m and 2.2 m.

Fig. 5 - Retta di calibrazione tra valori osservati e valori simulati per acquifero superficiale (punti blu) e acquifero semiconfinato (punti arancioni). I residui della dev.st sono rispettivamente 1.2 m and 2.2 m..

Furthermore, the transport model allowed to:

- study the plumes (Fig. 6) influence on quality status of downgradient areas (i.e. the influence on the water supply wells in Milan);
- find monitoring points not interested by the presence of plumes and considered related to a diffuse contamination.

Maps of diffuse contamination

Results of the kriging are shown in Fig. 7. Two critical kriged zones for diffuse pollution were identified (Fig. 7): the yellow zone is between the Maximum Contaminant Level of PCE (1.1 µg/l) and the drinkable water threshold (10 µg/l) whereas the red zone is over 10 µg/l. Considering the 2 different approaches listed in the materials and method section, the threshold values for diffuse contamination in FUA is different:

- in yellow zone: the threshold values can vary from 5 (B2) to 8 (B1) µg/l in shallow aquifer, from 6 (B2) to 7.5 (B1) µg/l in semiconfined aquifer;
- in red zone: the threshold values can vary from 8 (B2) to 12 (B1) µg/l in shallow aquifer, from 15 (B1) to 18 (B2) µg/l in semiconfined aquifer.

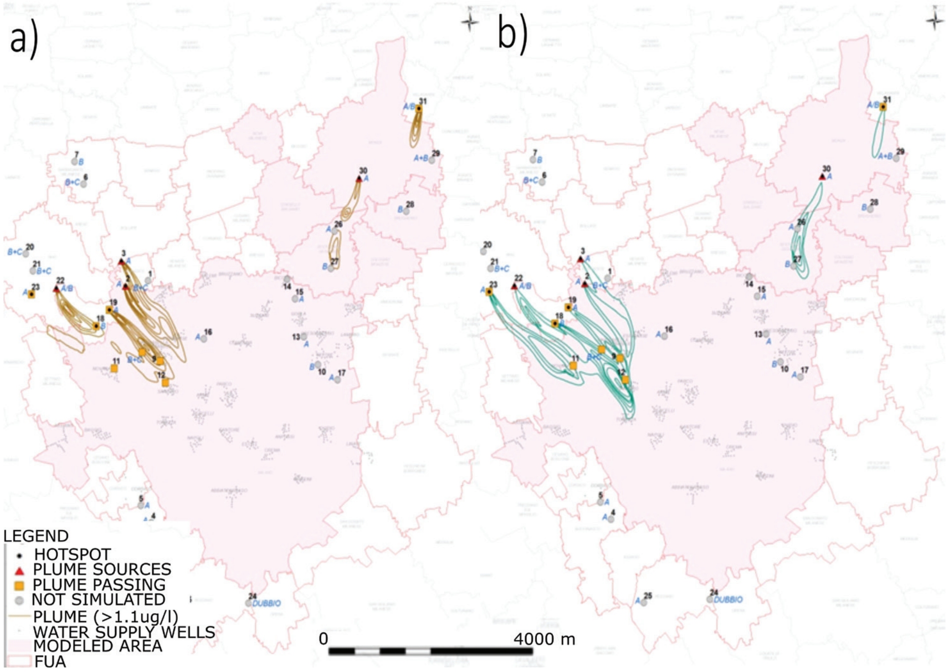

Fig. 6 - PCE transport model results in a) Group A (Layer 1) and b) Group B (Layer 3).

Fig. 6 - Risultati del modello numerico di trasporto di PCE nel a) Gruppo Acquifero A (Layer 1) e b) Gruppo Acquifero B (Layer 3).

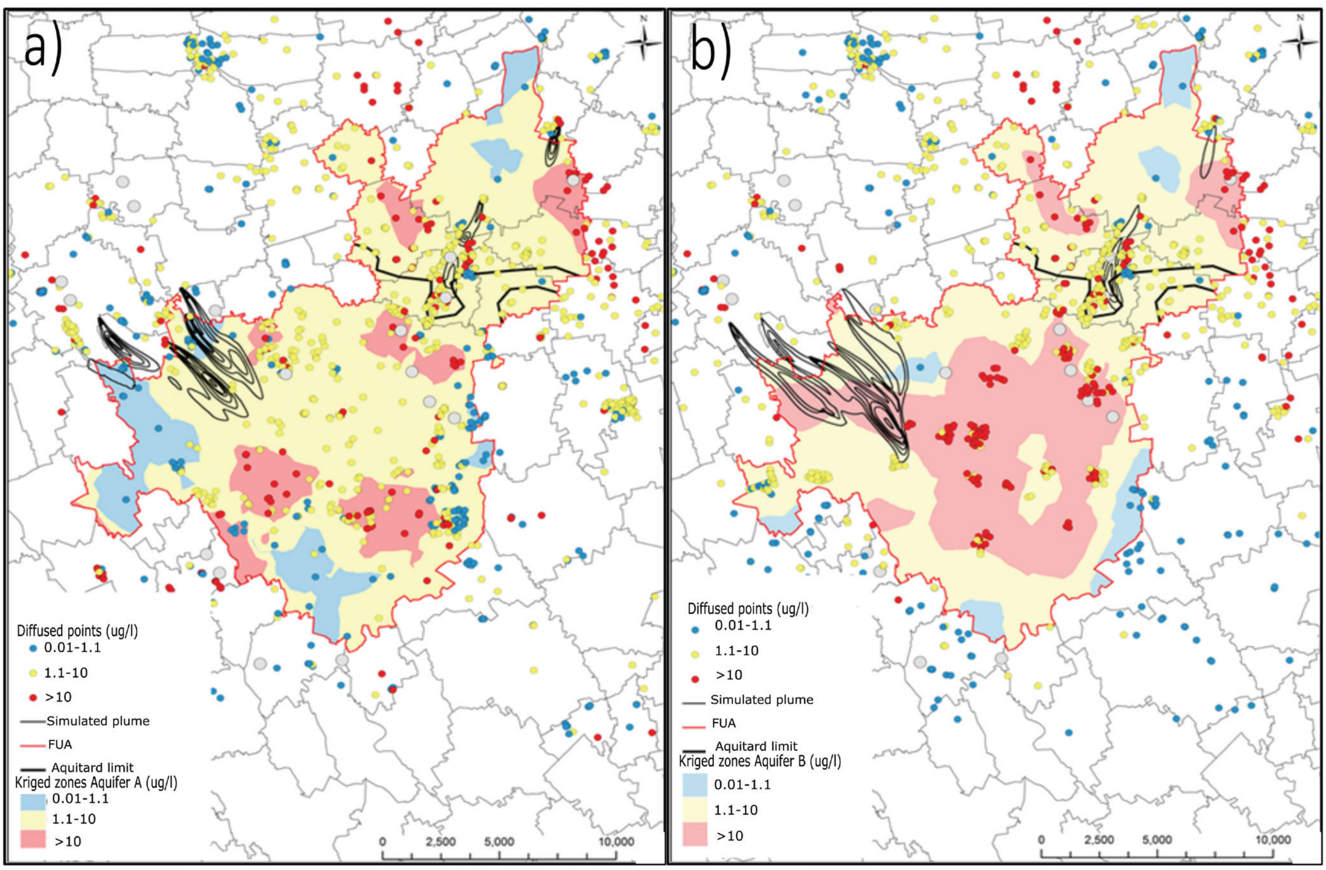

Fig. 7 - Kriging interpolation applied to PCE median concentration values (2010-2014) for a) Group A and b) Group B. Light blue zone is under Law Limit (<1.1 ug/l) whereas red zone exceeds the 31/2001 drinking water standard of 10 ug/l (ARPA LOMBARDIA 2016).

Fig. 7 - Interpolazione kriging applicata ai valori di concentrazione mediana di PCE (2010-2014) per a) Gruppo A e b) Gruppo B. La zona azzurra è al di sotto della CSC (<1.1 ug/l) mentre la zona rossa supera il limite delle acque potabili (31/2001, 10 ug/l.) (ARPA LOMBARDIA 2016).

DOI: 10.7343/as-2017-301

Stochastic numerical modelling results

The BT and clustered MPS methodologies results were compared in order to collect information about those areas that most probably could be responsible of the MPS (circled into the Fig. 8). In most of the cases, the identified areas match. Some improvements are necessary, mostly in areas located close to the boundary conditions.

About computational aspects, it is possible to state that, particle BT is faster than the clustered MPS approach. Moreover, the latter ones require several information for the inverse problem implementation.

Conclusions

In Milan FUA, which is one of the most populated and industrialized area in Europe, chlorinated solvents contamination is affecting groundwater quality since 50s because of the presence of highly industrialized zones. The work presented in this paper gives a contribution to the identification and distinction of PS and MPS contamination and presents the main results of the study related to the PCE diffuse pollution:

- The statistical analysis allowed to distinguish PS from MPS and led to define the threshold values (Approach A and B) of diffuse contamination for different zones mapped through geostatistical analysis in Milano FUA;
- The modelling phase defined the extension of PCE plumes both for shallow and semiconfined Aquifer with reference to 2014 contamination data;
- The stochastic analysis led to define the areas that most probably host MPS responsible for the PCE diffuse contamination (in the upper and central part of the pilot model).

These results become essential for regional administration for identifying the zones interested by PCE contamination.

More in detail, contamination distribution maps represent the base both for the Public Authorities decision-making process and for the development of strategical plans for assessing the contamination problem. The results of this study turn out to be useful for actions as:

- to operate against the diffuse contamination with a management plan (i.e. monitoring network improvement and cooperation between Public Authorities);
- to implement a risk-analysis study in order to evaluate, and eventually exclude, possible effects on human health of vapours from the un-saturated zone ;
- to evaluate the evolution of the diffuse contamination in short-term scenarios (i.e. a new dedicated network to monitor PCE diffuse contamination);
- to drive Public Authorities to focus on those areas where the intervention seems to be more urgent.

Acknowledgment: hydro-chemical values database, database for log-stratigraphies K, porosity values and piezometric survey in May 2014 were provided by ARPA- Lombardia. This pHD thesis was developed thanks to Dott. Luca Alberti (Supervisor), Dott.ssa Azzellino Arianna (Co-supervisor), Prof.ssa Laura Scesi (Tutor) and Prof. Alberto Guadagnini (The Chair of the Doctoral Program).

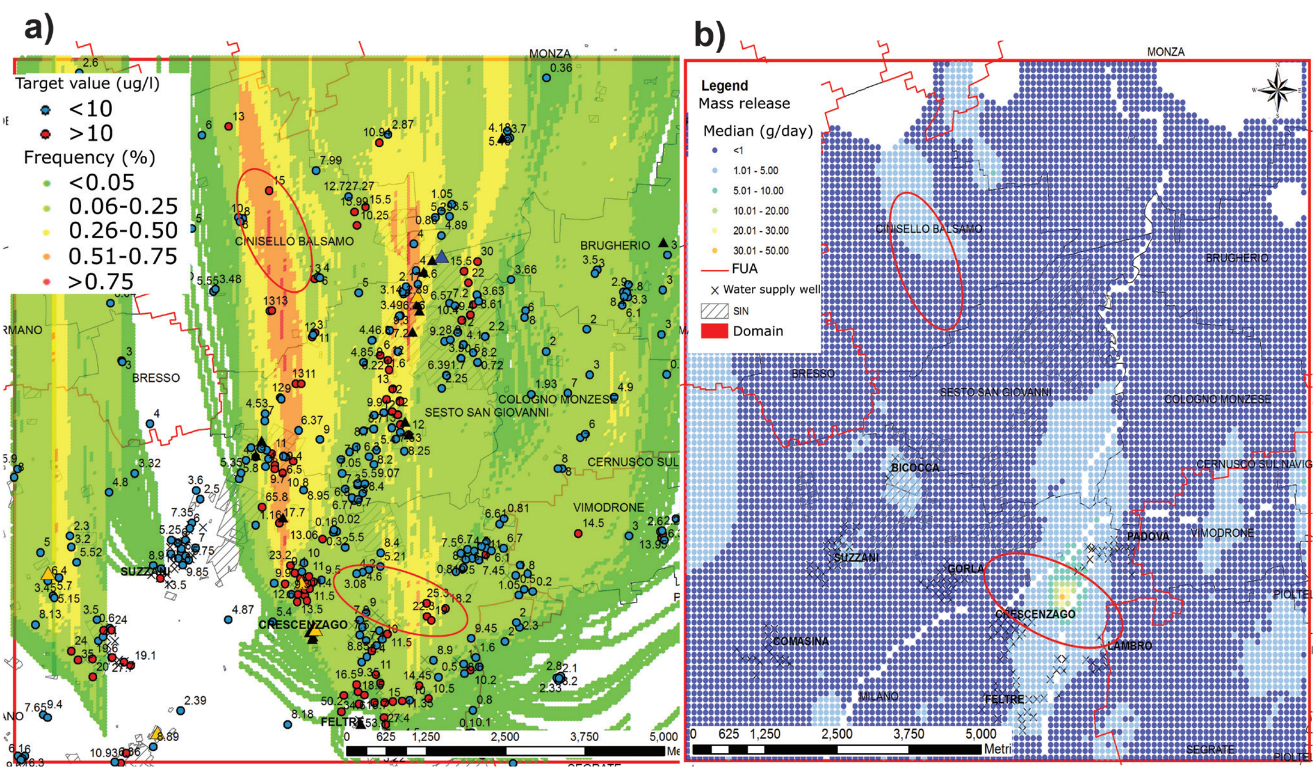

Fig. 8 - Comparison between two stochastic procedure with NSMC a) back-tracking particle with k-uncertainties and b) MPS with release mass uncertainties

Fig. 8 -Confronto tra due metodi stocastici con NSMC a) tracciamento delle particelle all'indietro considerando le incertezze della permeabilità k e b) MPS con incertezza della massa rilasciata dalle sorgenti.

REFERENCES

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

Alberti L, Azzellino A, Colombo L, Lombi S (2016). Use of cluster analysis to identify tetrachloroethylene pollution hotspots for the transport numerical model implementation in urban functional area of Milan, Italy, in: International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM. doi:10.5593/SGEM2016/B11/S02.091

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

Alberti L, Colombo L, Formentin G, Oberto G (2017). Assessment of sources of PCE contamination in groundwater using a Monte Carlo method in the Functional Urban Area of Milano. Rend. Online Soc. Geol. Ital. 42. doi:10.3301/ROL.2017.15

Anderson MP, Woessner WW (1992). Applied groundwater modeling: Simulation to flow and advective transport. J. Contam. Hydrol. 10, 339–340. doi:10.1016/0169-7722(92)90015-7

Arain FM, Pheng LS (2006). A framework for developing a knowledge-based decision support system for management of variation orders for institutional buildings. Electron. J. Inf. Technol. Constr. 11, 285–307. doi:10.1016/j.autcon.2005.06.005

ARPA Lombardia (2016). PROGETTO PLUMES - Final Report.

ARPA Lombardia (2015). PROGETTO PLUMES : Final Report.

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

Devic G, Djordjevic D, Sakan S (2014). Natural and anthropogenic factors affecting the groundwater quality in Serbia. Sci. Total Environ. doi:10.1016/j.scitotenv.2013.09.011

Doherty J (2014). Addendum to the PEST Manual. PEST Man. Add. 313.

Doherty J, Brebber L, Whyte P (2005). PEST: Model-independent parameter estimation. Watermark Comput. Corinda, Aust. 2005, 122. doi:10.1016/B978-141600119-5.50003-2

European Union (2006). Directive 2006/118/EC of the European Parliament and of the council of 12 December 2006 on the protection of groundwater against pollution and deterioration. Off. J. Eur. Union 19, 19–31. doi:http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006L0118

Fabbris L, Gallo F (1993). Bivariate Coefficients of Agreement among any Number of Observers. Educ. Psychol. Meas. 53, 337–349. doi:10.1177/0013164493053002004

Francani V, Beretta GP (1995). Preliminary survey about sustainable use of deeper groundwater in Milano province. Milan.

Frumkin H (2002). Urban sprawl and public health. Public Health Rep. 117, 201–217. doi:10.1016/S0033-3549(04)50155-3

Gattinoni P, Scesi L (2017). The groundwater rise in the urban area of Milan (Italy) and its interactions with underground structures and infrastructures. Tunn. Undergr. Sp. Technol. 62. doi:10.1016/j.tust.2016.12.001

Gelhar LW, Welty C, Rehfeldt KR (1992). A Critical Review of Data on Field-Scale Dispersion in Aquifers Lynn W. Gelhar, Claire Welty, Kenneth R. Rehfeldt. Water Resour. Res. 28, 1955–1974.

Harbaugh A, McDonald M (1996). User's Documentation for MODFLOW-96, an update to the U.S. Geological Survey Modular Finite-Difference Ground-Water Flow Model. Open-File Report, US Geol. Surv. 96–485.

Hill MC, Tiedeman CR (2007). Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty, Books.Google.Com. doi:10.1111/j.1745-6584.2007.00398.x

McDonald M, Harbaugh AW (1988). A modular three-dimensional finite difference ground-water flow model, Techniques of Water-Resources Investigations, 6-A1. doi:10.1016/0022-1694(70)90079-X

Mouron P, Nemecek T, Scholz RW, Weber O (2006). Management influence on environmental impacts in an apple production system on Swiss fruit farms: Combining life cycle assessment with statistical risk assessment. Agric. Ecosyst. Environ. 114, 311–322. doi:10.1016/j.agee.2005.11.020

Nolan BT, Hitt KJ, Ruddy BC (2002). Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. Environ. Sci. Technol. 36, 2138–2145. doi:10.1021/es0113854

Pollock DW (1994). User's Guide for MODPATH/MODPATH-PLOT, Version 3: A particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite-difference ground-water flow model. Open-File Rep. 94-464 0–249. doi:94-464

Provincia di Milano (1992). Indagini sulla presenza di composti organo-alogenati nelle acque di falda della Provincia di Milano. Milano *"Surveys about the solvents compounds presence in the Milan Province groundwater"*.

Reghunath R, Murthy TRS, Raghavan BR (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India. Water Res. 36, 2437–2442. doi:10.1016/S0043-1354(01)00490-0

Rosner B (1975). On the Detection of Many Outliers. Technometrics 17, 221. doi:10.2307/1268354

Stevenazzi S, Masetti M, Beretta GP (2017). Groundwater vulnerability assessment: from overlay methods to statistical methods in the Lombardy Plain area. Acque Sotter. - Ital. J. Groundw. 6. doi:10.7343/as-2017-276

Tonkin M, Doherty J (2009). Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resour. Res. 45. doi:10.1029/2007WR006678

Tonkin M, Doherty J, Moore C (2007). Efficient nonlinear predictive error variance for highly parameterized models. Water Resour. Res. 43. doi:10.1029/2006WR005348

Zheng C, Wang PP (1999). MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. pp239