

# Knowledge-based planning of groundwater treatment trains for an efficient drinking water supply system in urban areas

M. Antonelli, A. Azzellino, E. Ficara, A. Piazzoli and G. Saccani

## ABSTRACT

This paper presents knowledge-based planning of the most appropriate treatment schemes for drinking water supply (DWS) systems, jointly applying statistical tools and experimental tests. Milan City was chosen as the case study, its DWS system being composed of more than 20 DWS units, widespread in the urban area. First, multivariate statistical techniques (factor analysis and cluster analysis) were applied to groundwater monitoring data, to identify specific contaminations of the captured groundwater and their spatial distribution. In detail, eight typical compositions of captured groundwater and their distribution in DWS units were identified, leading to the selection of different treatment processes to be adopted in DWS units. Then, process performances were experimentally evaluated on representative groundwater samples. Heterotrophic denitrification was tested at pilot-scale on groundwater contaminated by nitrate, identifying optimal operating and design parameters for complying with regulation limits. Granular activated carbon (GAC) adsorption was tested at laboratory-scale for the removal of volatile organic compounds (VOCs) and pesticides, present in various ratios in groundwater samples: removal efficiencies were obtained and the occurrence of competition phenomena was highlighted. Air stripping was tested at pilot-scale for removing the most volatile VOCs (trichloroethylene, tetrachloroethylene), highlighting its usefulness as pre-treatment before adsorption to increase GAC lifetime.

**Key words** | activated carbon, biological nitrate removal drinking water, multivariate statistics, VOCs

M. Antonelli (corresponding author)  
A. Azzellino  
E. Ficara  
A. Piazzoli  
G. Saccani  
Politecnico di Milano,  
DICA – Environmental Section,  
Piazza Leonardo da Vinci, 32,  
Milano 20133,  
Italy  
E-mail: [manuela.antonelli@polimi.it](mailto:manuela.antonelli@polimi.it)

## INTRODUCTION

Drinking water supply (DWS) in industrialized and densely populated metropolitan areas is often assured through multiple DWS units that intercept various punctual and diffuse pollution plumes.

In such a complex context, planning the appropriate treatment trains can be a hard task. The knowledge of raw groundwater composition and its variability, in order to individuate the water parameters that need to be removed, is critical. The water quality parameters of interest being many and characterized by a complex spatial distribution, data-mining techniques may be extremely supportive (Olsen *et al.* 2012; Selle *et al.* 2013) in this perspective. In the literature, different statistical tools have

been applied for the analysis of water quality data; among these, data driven techniques such as principal components analysis (PCA), factor analysis (FA) and cluster analysis (CA) are reported as effective and unbiased methods to extract meaningful information from a large dataset (Papaioannou *et al.* 2010). Despite the fact that the use of multivariate statistics for water quality assessment is widely reported in the literature (Papaioannou *et al.* 2010; Olsen *et al.* 2012; Page *et al.* 2012), its application as a supporting tool for the planning and management of DWS systems is not conventional.

The planning of a DWS system requires the selection of the appropriate treatment schemes, based on the

comparison of the available technologies, considering technical and non-technical factors (Hamouda *et al.* 2009). Since knowledge acquisition is often case-specific, experimental and pilot studies are needed in order to obtain representative information about the technical and economic efficiency of the alternatives (Joksimovic *et al.* 2006).

The present work considers a case study of the Milan City DWS system. The system provides about  $230 \times 10^6 \text{ m}^3 \text{ y}^{-1}$  of drinking water, collecting groundwater from three aquifers by 522 wells spread across the city area and organized in clusters of about 20 wells, each one conveying water to a DWS unit. This study applies a knowledge-based concept to the planning of the proper treatment train to have in the DWS units. As the first step of the knowledge-based concept, a multivariate statistical analysis was applied to the available groundwater monitoring data in order to identify specific contaminations of the captured groundwater and their spatial distribution. Then the second step of the knowledge framework, the experimental testing of some treatment processes, was performed at laboratory- and pilot-scale to assess the optimal operating conditions. Statistical analyses were used as a supporting tool to guide the selection of (1) the most appropriate treatment technologies and (2) the most representative groundwater samples to be tested, in order to obtain results that could be extended to a high number of cases. Specifically, as treatments are concerned, activated carbon (AC) adsorption, air stripping and heterotrophic biodegradation were evaluated respectively to address the removal of organics (i.e. volatile organic compounds (VOCs) and pesticides) and nitrate.

## METHODS

### Statistical analysis

Monitoring data of groundwater quality collected during 2007–2012 by the Milan City DWS service were analyzed through a data-driven statistical analysis approach constituted by a sequence of Factor and CA (Affi & Clark 1996). The analyzed dataset was constituted by 8,477 cases (corresponding to different capturing wells) and 88 parameters: six physical–chemical parameters, seven ionic species, 18 metals and transition metals, 19 VOCs, 23 pesticides and 15

polycyclic aromatic hydrocarbons (PAHs). PCA allowed the eigenvalues and eigenvectors to be extracted from the covariance matrix of the original variances. FA was chosen to reduce the contribution of the less significant parameters within each component, by extracting a new set of varifactors through rotating the axes defined by the PCA extraction. The varimax rotation criterion was used to rotate the PCA axes, allowing the axes' orthogonality to be maintained. The number of factors to be retained was chosen on the basis of the 'eigenvalue higher than 1' criterion (i.e. all the factors that explained less than the variance of one of the original variables were discarded). That allowed the selection of a few factors able to describe the whole dataset with minimum loss of original information. Moreover, a hierarchical cluster analysis (HCA, according to Affi & Clark 1996) was used to analyze the similarities among the water quality profiles at the different monitoring stations, using the Euclidean distance as the distance metric (see Equation (1)):

$$d_2(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^q (\mathbf{x}_{ik} - \mathbf{x}_{jk})^2} \quad (1)$$

where  $i$  and  $j$  refer to a pair of stations, and  $k$  to the considered parameters.

HCA was run based on the FA extracted varifactors and Ward's method was used as the clustering method. All the data analysis was performed through the IBM SPSS Statistics 21 statistical package.

### Experimental tests

As far as the experimental setup is concerned, a series of laboratory-scale batch experiments was performed to assess organics removal in groundwater samples by adsorption on AC. A commercial AC (Cecacarbon GAC 1240) was tested through bottle-point isotherm measurements (ASTM D5919-96): different doses (from 1 to  $35 \text{ mg L}^{-1}$ ) of AC were dosed in 1 L pyrex bottles and kept mixed on a magnetic stirrer (400 RPM) at ambient temperature (20–25 °C) for 48 h before measuring the residual concentration of VOCs and pesticides in the liquid phase.

Air stripping was tested through a pilot-scale steel column (0.15 m diameter, 1.8 m height, 3.2 min HRT,

15.8 m<sup>3</sup> m<sup>-2</sup> h<sup>-1</sup> air flow rate), analyzing inlet and outlet concentrations of VOCs and pesticides.

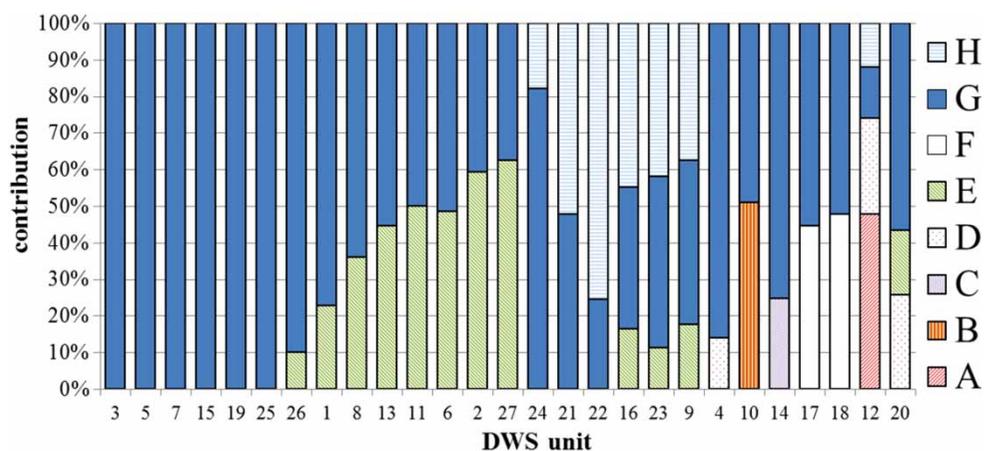
Finally, heterotrophic denitrification was tested in a pilot-scale biofilter consisting of a PVC column (0.33 m inner diameter, 2.40 m height, 21 min EBCT, 0.38 porosity) filled with expanded clay media (Biolite<sup>®</sup>). After a start-up period of about 50 days, aimed at allowing biomass colonization, the biofilter was fed up-flow with groundwater extracted by a supply well at a flow rate of 0.6 m<sup>3</sup> h<sup>-1</sup>. Average influent concentration of NO<sub>3</sub> during the experimentation was 8.9 ± 0.65 mg N L<sup>-1</sup>. The process was evaluated for 10 weeks by monitoring NO<sub>3</sub>, NO<sub>2</sub>, volatile suspended solids (VSS) and total organic carbon (TOC) in water at the biofilter inlet, outlet and at four intermediate biofilter heights. During this period, different operating conditions were tested: (i) nutrient availability was varied between 40% and 135% of the stoichiometric request, evaluated according to Henze *et al.* (2008), by supplying different doses of sodium acetate and potassium phosphate; (ii) cycle durations were varied (24, 48 and 72 h), by periodically backwashing the reactor in order to remove excess biomass.

Regarding analytical methods, the concentrations of VOCs, pesticides, nitrite and nitrate ions and VSS were measured according to EPA 524.3-09, APAT-IRSA 5060/03, EPA 300.1 and APHA 2540 methods, respectively; TOC was measured photometrically by a kit method (LCK 380, Hach Lange).

## RESULTS AND DISCUSSION

FA was applied to the groundwater quality dataset in order to reduce the number of variables and to identify sets of parameters with the same pattern of variability. Starting from the 88 analyzed parameters, 16 relevant factors were extracted, explaining 57.6% of data variance and identifying different kinds of contamination in the studied area. Among factors summarizing parameters usually present in natural groundwater, such as physical-chemical parameters or alkaline metals, 13 of the extracted factors could be attributed to pollution of anthropic origin, such as fertilizer constituents (nitrate, sulphate and chloride), PAHs, pesticides, heavy metals (Sb, Cd, V, Se, As and Cr) and VOCs. So FA allowed the identification of the main pollution components affecting the groundwater captured in the studied DWS system.

Hence, CA was performed to recognize typical compositions of captured groundwater and their distribution in DWS units: 98.8% of groundwater samples were grouped in eight clusters, each having similar factor scores and, thus, similar contamination characteristics. CA results (reported in Figure 1) showed that the majority of cases belong to cluster G (59.6% of total samples), grouping groundwater samples displaying average values for all the extracted factors ( $-0.25\sigma \div +0.25\sigma$ ), therefore identifying groundwater of average quality. Two other main clusters can be distinguished: cluster E (15.3% of total samples), grouping groundwater samples affected by nitrate and



**Figure 1** | Distribution of the eight clusters in DWS units (contributions <10% are not reported). According to the factor scores, each cluster represents a typical groundwater composition: G: average groundwater quality; E: contamination by nitrate and pesticides; H: contamination by chromium; A, B, C, D and F: local contaminations by groups of two to three pesticides and/or VOCs at relevant concentrations.

pesticide contamination at concentrations slightly higher than the average values (relative factor scores  $< 1\sigma$ ); and cluster H (11.6% of total samples), grouping samples having high concentrations of chromium (factor score  $> 1.5\sigma$ ), above the Environmental Quality Standard of  $50 \mu\text{g L}^{-1}$ . The remaining clusters (A, B, C, D and F) identified groundwater affected by local contaminations, mainly due to the presence of groups of two to three pesticides and/or VOCs at concentrations higher than the average (factor scores  $> 2.5\sigma$ ). As can be noticed in Figure 1, CA also showed that four main kinds of DWS units could be recognized, according to the captured groundwater composition: units totally fed by water belonging to cluster G, units collecting a mix of waters belonging to clusters G and E, units with the further presence of water from cluster H and, finally, units characterized by the presence of at least one of the other clusters, associated with point contaminations.

Therefore FA and CA results provided useful indications for identifying the most proper technologies to be adopted locally for treating groundwater in the studied DWS system.

Nitrate, pesticides and VOCs were the species higher or very close to the national regulation limits so different types of treatments were experimentally tested to address the removal of these pollutants, particularly: (1) heterotrophic denitrification, for nitrate removal, (2) AC adsorption, for both pesticide and VOC removal, and (3) air stripping for the removal of the most volatile VOCs, i.e. trichloroethylene (TCE), tetrachloroethylene (PCE). Experiments were carried out on a limited number of groundwater samples, extracted from representative DWS units which were selected based on the results of the statistical analysis.

## Nitrate removal

Groundwater belonging to cluster E showed a nitrate concentration above the Italian quality standard for drinking waters (i.e.  $50 \text{ mg L}^{-1}$ ). The suitability of a heterotrophic denitrification treatment was then experimentally assessed using a pilot-scale biofilter, evaluating the main aspects of the process: (1) nitrate removal efficiency, (2) residual nitrite concentration in the effluent, and (3) risk of bacterial regrowth in the distribution network triggered by the release of biomass and the availability of readily biodegradable organic carbon. All these aspects were evaluated with respect to the main process parameters, namely organic carbon availability and nitrogen volumetric load.

Figure 2 shows  $\text{NO}_3^-$  and  $\text{NO}_2^-$  concentration profiles along the biofilter, observed during the pilot-scale testing for different carbon dosage conditions, expressed as volumetric load ( $B_V$ , in  $\text{kg N m}^{-3} \text{ d}^{-1}$ ). Results showed that high nitrate removal efficiencies (i.e.  $> 86\%$ ) and complete denitrification could be achieved by supplying a carbon source over 90% of the stoichiometric request, corresponding to  $21 \pm 1.5 \text{ mg C L}^{-1}$ . On the other hand, C-limiting conditions (corresponding to a C dosage  $< 90\%$  of the stoichiometric request) resulted in unsatisfactory nitrogen removal efficiencies of the order of  $46\% \pm 32\%$  (mean  $\pm$  standard deviation) and nitrite concentrations in the effluent of  $1.4 \pm 2.20 \text{ mg N L}^{-1}$ .

In order to assess the risk of bacterial regrowth in distribution networks, the release of organic carbon and biomass during pilot tests was also evaluated: residual TOC and VSS

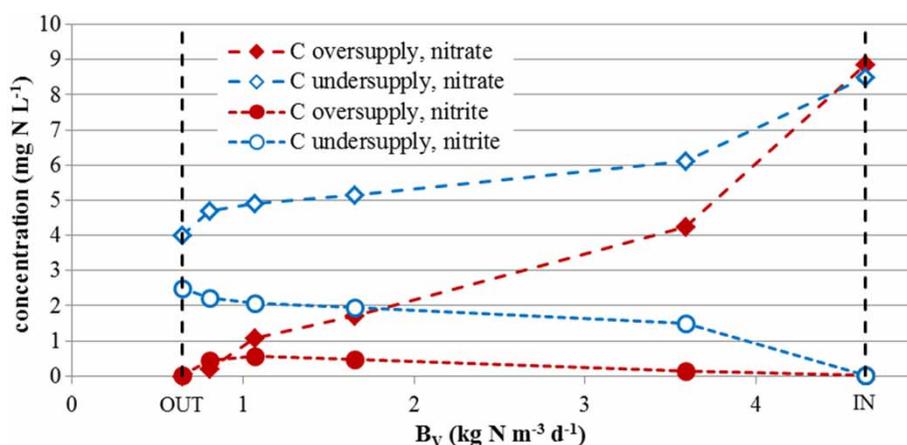


Figure 2 | Average nitrate and nitrite concentrations observed along the biofilter for tested carbon supply conditions.

in the effluent were equal to respectively  $4.6 \pm 1.34 \text{ mg C L}^{-1}$  and  $3.2 \pm 1.57 \text{ mg VSS L}^{-1}$ . No significant differences (Student's *t*-test, *p*-value > 0.05) among the tested carbon dosage conditions were found in TOC and VSS effluent concentrations.

The effect of the volumetric load on the denitrification process was evaluated considering the process efficiency along the biofilter height. Under C-limiting conditions, none of the tested  $B_V$  provided nitrate removal efficiencies higher than 60%, resulting in the highest nitrite concentrations in the effluent (up to  $6.2 \text{ mg N L}^{-1}$ ). Under carbon oversupply conditions, removal efficiencies were strongly affected by the treatment cycle duration ( $t_{TR}$ ), which was modified by varying the biofilter backwashing frequency: for  $t_{TR}$  of 24 and 72 h and volumetric loads between  $0.64$  and  $1.07 \text{ kg N m}^{-3} \text{ d}^{-1}$ , complete nitrate removal and nitrite concentrations below the regulation limit ( $0.5 \text{ mg L}^{-1}$ ) were obtained; on the other hand, at  $t_{TR}$  of 48 h, nitrate removal efficiencies >90% were obtained only for  $B_V$  up to  $0.80 \text{ kg N m}^{-3} \text{ d}^{-1}$ . Moreover, nitrite concentrations in the effluent below the regulation limit were observed at  $t_{TR}$  of 48 h only when  $B_V$  was set to  $0.64 \text{ kg N m}^{-3} \text{ d}^{-1}$ . This behavior may probably be attributed to transitory or less stable conditions achieved at the intermediate  $t_{TR}$  of 48 h; however, further investigations are needed to draw conclusions about this behavior.

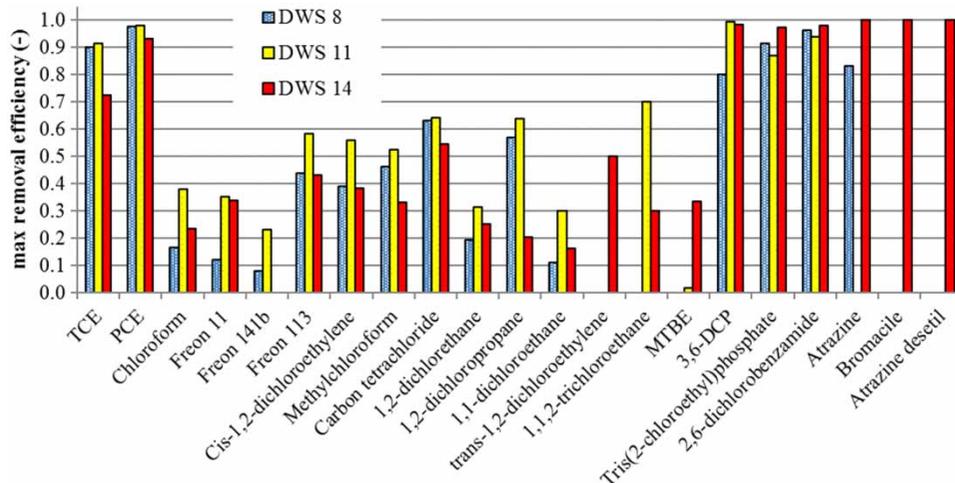
Finally, no significant removal was observed for the three VOCs (PCE, TCE, vinyl chloride) and nine pesticides (2,6-dichlorobenzamide, 3,6-dichloropyridazine, atrazine, atrazine desethyl, atrazine desisopropyl, bromacil, hexazinone, terbuthylazine desethyl, tris(2-chloroethyl)phosphate) which were initially present in the inlet water. Therefore, additional treatments, such as AC adsorption or air stripping, should be designed for groundwater requiring the removal of these organics.

### AC adsorption

Since the presence of VOCs and pesticides affected most of the captured groundwater, adsorption equilibrium batch tests were carried out selecting groundwater samples from two DWS units (DWS8 and DWS11) representative of diffuse and average contamination (clusters G and E) and one DWS unit (DWS14) representative of an organic micropollutant hotspot (tris(2-chloroethyl)phosphate, PCE, and 2,6-

dichlorobenzamide of the order of  $0.4 \text{ } \mu\text{g L}^{-1}$ ,  $242.4 \text{ } \mu\text{g L}^{-1}$  and  $0.7 \text{ } \mu\text{g L}^{-1}$ , respectively – cluster C). Figure 3 shows the maximum removal efficiency observed during batch experiments (obtained at the highest AC dose) for various pollutants: it can be noticed that pesticides, present at initial concentrations  $<0.7 \text{ } \mu\text{g L}^{-1}$ , were satisfactorily removed in all the groundwater samples, with removal efficiencies higher than 80%. On the other hand, VOCs showed different affinity towards AC: the highest removal efficiencies were observed for PCE and TCE (>93% and >72%, respectively), while the other VOCs showed a lower affinity towards AC, depending on their physical-chemical properties (such as solubility, molecular weight and structure, vapor pressure). It must be pointed out that the majority of the VOCs, except PCE, TCE and chloroform, were present in raw water at low concentrations, never being  $>1.0 \text{ } \mu\text{g L}^{-1}$ ; for most of these contaminants, removal efficiencies could be limited by the low initial concentration. The same was not verified concerning chloroform, being present at relatively high concentrations ( $12.9 \text{ } \mu\text{g L}^{-1}$ ,  $15.6 \text{ } \mu\text{g L}^{-1}$  and  $29.2 \text{ } \mu\text{g L}^{-1}$  in samples from DWS8, DWS11 and DWS14, respectively) but removed with low efficiencies ( $<0.4$ ). The low affinity of AC for chloroform should be carefully considered in designing a full-scale granular activated carbon (GAC) system, in particular when final disinfection is obtained by means of chlorine-based products, since trihalomethanes would be formed. In fact, the concentration of trihalomethanes can be easily increased, considering the initial concentrations of chloroform in captured groundwater and its possible transformation during the disinfection phase.

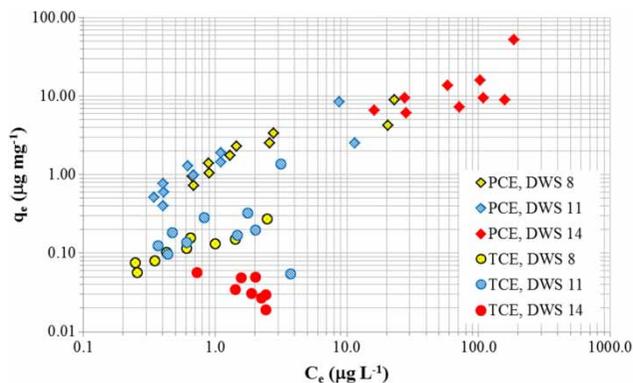
Isotherms were then computed from experimental data on PCE and TCE. Isotherms have an important role in the study of adsorption processes, allowing the evaluation of the affinity between a solute and an adsorbent and the preliminary design of full-scale operations (Foo & Hameed 2010). Residual liquid concentration ( $C_e$ , expressed in  $\mu\text{g L}^{-1}$ ) vs solid phase concentration ( $q_e$ , expressed in  $\mu\text{g mg}^{-1}$ ) data obtained for PCE and TCE in the different water samples are reported in Figure 4: it should be pointed out that no significant differences were observed in PCE removal in water samples coming from the three DWS units, despite the different initial concentrations ( $29.1 \text{ } \mu\text{g L}^{-1}$ ,  $16.2 \text{ } \mu\text{g L}^{-1}$  and  $242.4 \text{ } \mu\text{g L}^{-1}$  respectively in DWS8, DWS11 and DWS14). In contrast, TCE concentrations at the DWS14 unit are different from those



**Figure 3** | Maximum removal efficiency for various VOCs and pesticides observed during adsorption batch experiments in three DWS units (DWS14, DWS11 and DWS8 belonging to clusters G, E and C respectively).

obtained for the other two DWS units: despite the comparable initial concentrations in all groundwater samples ( $2.5\text{--}4.3\ \mu\text{g L}^{-1}$ ), TCE removal efficiency was lower in samples from DWS14, in which the initial PCE concentration was higher than in samples from DWS8 and DWS11. This behavior suggests that adsorption of compounds as TCE, having lower affinity towards AC, may be affected by the presence of competitors for adsorption sites in the water matrix, being in this case PCE.

These results suggested that adsorption onto AC is a suitable process for the removal of organic micropollutants. Nonetheless, when poorly adsorbable organics are present, an upgrade of the process is required, in order to increase the bed lifetime, for example, adopting higher EBCTs or a two-stage in-series configuration.



**Figure 4** | Equilibrium solid phase ( $q_e$ ) vs residual liquid concentration ( $C_e$ ) for PCE and TCE observed in three tested water samples.

### Air stripping

In some cases, especially for DWS units treating groundwater with high levels of VOCs (such as clusters C, D and F), air stripping could be a suitable solution both as a stand-alone treatment or as a pre-treatment before adsorption onto AC. Stripping efficiency was evaluated using a pilot-scale aeration column, fed on groundwater containing various VOCs, mainly PCE and TCE at average initial concentrations of  $16.2\ \mu\text{g L}^{-1}$  and  $2.4\ \mu\text{g L}^{-1}$ , respectively. Significant removal was observed for PCE and TCE, on average  $36\% \pm 11\%$  and  $18\% \pm 13\%$ , respectively. PCE removal was higher than TCE removal, since a lower air-water ratio is required as expected by their values of Henry's constant. Moreover, results for TCE showed a significant correlation between removal efficiency and initial concentration ( $\rho = 0.713$ ,  $p\text{-value} = 0.001$ ,  $N = 10$ ), while the same behavior was not observed for PCE. Results suggested that the adoption of an air-stripping step before GAC could be a valid solution, especially in the presence of PCE at high concentrations: as a matter of fact, the preliminary removal of PCE could reduce competition for adsorption sites, increasing the adsorption of the other organic micropollutants and reducing the frequency of GAC regeneration. However, the choice and the proper design of such a layout must also be based on economic evaluations, considering the costs related to both the treatment phases, depending on the specific water composition.

## CONCLUSIONS

Data mining applied to groundwater quality data permitted the identification of the main kinds of contamination affecting the DWS system of Milan City, defining typical water composition contributing to each DWS unit. According to these results, different treatment processes were selected to be adopted in DWS units and their performances were experimentally evaluated. Experimental tests were performed on the most representative groundwater samples to obtain information generalizable to a high number of DWS units.

Heterotrophic denitrification resulted in being effective for nitrate abatement despite the significant organic contamination due to pesticides and VOCs. Carbon supply over 90% of the stoichiometric requirements and volumetric loads  $\leq 1.07 \text{ kg N m}^{-3} \text{ d}^{-1}$  were found to assure effective and complete denitrification.

Adsorption onto AC was effective for the removal of several VOCs and pesticides. However, variable efficiencies were highlighted towards contaminants due to (1) their physical-chemical properties and (2) the influence of water composition, which magnifies competition phenomena. An air-stripping step before GAC resulted in a successful solution to removing the most volatile compounds (e.g. PCE), favoring the following adsorption of low-adsorbable compounds (e.g. TCE).

Reported results showed that data-mining techniques can effectively support the optimal design of a drinking water treatment train, allowing (1) identification of the main contaminants to be controlled and (2) reduction of efforts in testing the performances of the most suitable treatment processes, supporting the selection of the most representative water samples to be used in the experiments, in order to obtain results of general validity. The applied knowledge-based approach resulted in a suitable supporting tool for the planning of a DWS system in an urban area, such as Milan City, coping with a composite source water quality, a complex groundwater capturing system and many fingerprint constraints.

## ACKNOWLEDGEMENTS

This research has been funded by Metropolitana Milanese S.p.A. (MM), which manages the integrated water service for the city of Milan (I).

The authors thank Carlo Carrettini and Fabio Marelli for having made this study possible, and Angela Manenti, Chiara Pagano and the MM laboratory staff for their helpful cooperation during the experimental work.

## REFERENCES

- Affi, A. & Clark, V. 1996 *Computer-Aided Multivariate Analysis*. Texts in Statistical Science, Chapman and Hall, London, UK.
- Foo, K. Y. & Hameed, B. H. 2010 *Insights into the modeling of adsorption isotherm systems*. *Chemical Engineering Journal* **156**, 2–10.
- Hamouda, M. A., Anderson, W. B. & Huck, P. M. 2009 *Decision support systems in water and wastewater treatment process selection and design: a review*. *Water Science & Technology* **60** (7), 1757–1770.
- Henze, M., van Loosdrecht, M. C. M., Ekama, G. A. & Brdjanovic, D. 2008 *Biological Wastewater Treatment: Principles, Modelling and Design*. IWA Publishing, London, UK.
- Joksimovic, D., Kubik, J., Hlavinec, P., Savic, D. & Walters, G. 2006 *Development of an integrated simulation model for treatment and distribution of reclaimed water*. *Desalination* **188**, 9–20.
- Olsen, R. L., Chappell, R. W. & Loftis, J. C. 2012 *Water quality sample collection, data treatment and results presentation for principal components analysis – literature review and Illinois River watershed case study*. *Water Research* **46** (9), 3110–3122.
- Page, R. M., Lischeid, G., Epting, J. & Huggenberger, P. 2012 *Principal component analysis of time series for identifying indicator variables for riverine groundwater extraction management*. *Journal of Hydrology* **432–433**, 137–144.
- Papaioannou, A., Mavridou, A., Hadjichristodoulou, C., Papastergiou, P., Pappa, O., Dovriki, E. & Rigas, I. 2010 *Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health*. *Environmental Monitoring and Assessment* **170** (1), 87–97.
- Selle, B., Schwientek, M. & Lischeid, G. 2013 *Understanding processes governing water quality in catchments using principal component scores*. *Journal of Hydrology* **486**, 31–38.