Spatial diversity of chlorine residual in a drinking water distribution system: application of an integrated fuzzy logic technique

Donatella Termini and Gaspare Viviani

ABSTRACT

A reduction in the concentration of chlorine, which is used as a chemical disinfectant for water in drinking water distribution systems, can be considered to be an index of the progressive deterioration of water quality. In this work, attention is given to the spatial distribution of the residual chlorine in drinking water distribution systems. The criterion for grouping the water-quality parameters normally used is highly subjective and often based on data that are not correctly identified. In this paper, a cluster analysis based on fuzzy logic is applied. The advantage of the proposed procedure is that it allows a user to identify (in an automatic way and without any specific assumption) the zonation of the network and easily calibrate the unknown parameters. An analysis of the correlation between the sampling sites for the residual chlorine has been used to assess the applicability of the procedure.

Key words | calibration, clustering, fuzzy logic, networks, reactions, water quality

Donatella Termini (corresponding author)
Gaspare Viviani
Dipartimento di Ingegneria Civile, Ambientale,
   Aerospaziale, dei Materiali (DICAM),
University of Palermo,
Palermo 90128,
Italy
E-mail: *donatella.termini@unipa.it*

INTRODUCTION

The quality of the water released in distribution systems can change, in space and time, because of various phenomena: reactions with organic or inorganic compounds in the bulk liquid phase and with the walls of pipes as a result of corrosion, material releases, and/or biofilm formation; and the mixing of water with different origins and/or characterized by different residence times in the network (LeChevallier *et al.* 1990; De Rosa *et al.* 1998; Munavalli & Kumar 2004; Environmental Protection Agency (EPA) 2005).

In particular, the bacterial population is generally increasing in the network, both because of the growth of the microorganisms that are initially present in the water or infiltrated from outside, and because of the re-growth phenomena in parts of the network where favorable conditions occur (long hydraulic detention times, nutrient availability, and high temperature) (Lund & Ormerod 1995; De Rosa *et al.* 1998).

The chlorine, which is often used as water disinfectant, can be regarded as an index of the progressive deterioration of the water quality (Castro & Neves 2003; Ghinelli 2004). In fact, the concentration of chlorine released in the network from the treatment plant is reduced as a result of reactions with inorganic and organic compounds present in the bulk liquid and in the pipe walls (Van der Kooij 1992; Lu *et al.* 1999).

To ensure that a sufficient chlorine concentration (around 0.2 mg/l) is present in every part of the network, booster chlorination facilities can be installed within the network (EPA 2005). For this reason, it is important to estimate the correct residual chlorine content at nodes (or at specific 'control nodes') of the drinking water distribution network.

To identify locations where the chlorine concentration drops below a certain level or exceeds a given maximum level, the resource manager of the distribution network must be able to predict the spatial distribution of the residual chlorine.

For this purpose, the resource manager should acquire the necessary computational instruments to evaluate the

amount of chlorine in every part of the network. To efficiently assist with this management, the simplest model of chlorine decay with sufficient accuracy should be identified (Fisher et al. 2011).

Most of the currently available models of chlorine decay (see as examples Biswas et al. 1993; Rossman et al. 1994; Ozdemir & Ger 1998; Ozdemir & Ucak 2002; Munavalli & Kumar 2004, 2005; Kohpaei & Sathasivan 2011) consider pipes as 'chemical reactors' where chlorine enters at one end, decays because of reactions with organisms in the water (bulk decay), is consumed around the pipe walls (wall decay) because of the biofilm attached to them, and is transported toward the other end of the pipe. Thus, the law of chlorine decay is described by a combination of first-order kinetics explaining the reactions occurring within the water and in the pipe's walls (Fisher et al. 2011). These kinetics include three parameters (k_b, k_w, and k_f– referred to hereafter as water-quality parameters), which have to be determined using experimental data (Rossman et al. 1994; Hallam et al. 2002).

It should be taken into account that the instantaneous decay rate models for the bulk and wall reactions should be integrated simultaneously (Fisher et al. 2011). Previous researches (Powell & Hallam 2000; Ozdemir & Demir 2007) have demonstrated that the parameter determining the bulk-reaction (k_b) does not vary with the pipe's diameter. Ozdemir & Demir (2007) instead observed a variability in parameter k_b with the Reynolds number, Re. However, the verification of the relation k_b–Re still requires more detailed experiments. Several experimental analyses indicate that k_b varies (in the range 0.3–1.398 d^{-1}) with an inverse proportionality with the initial chlorine concentration (see as an example Rossman et al. 1994; Furlani & Marinelli 1998). Other authors (Castro & Neves 2003; Lee et al. 2004) have observed a low variation of the chlorine concentration with the parameter k_b so that they suggest to assume a unique value of k_b for all pipes of the network.

In contrast, the literature (among others Yeh 1986; Yang & Ko 1996) shows that the value of the parameter that determines the wall reactions (k_w) depends both on the local geometrical characteristics of the pipes (such as the pipe diameter, pipe material, and pipe roughness) and on the flow characteristics (such as the flow velocity and the level of bacterial colonization).

Thus, the spatial variability of the water-quality parameters plays an important role in the accurate evaluation of the chlorine distribution in the network (Yang & Ko 1996; Lee et al. 2004).

The number of parameters included in the model could also be considered to be a reasonable measure of the model's complexity. On the other hand, because the number of unknown parameters affects the sensitivity in the prediction error of a model's results (Clark & Coyle 1990; Datta & Sridharan 1994; Munavalli & Komar 2005), it is necessary to reduce the number of parameters as much as possible or to group them into 'clusters'.

On the basis of the aforementioned, it is clear that the estimation of the unknown parameters requires an efficient criterion for grouping them and the application of calibration techniques to easily and automatically identify their values.

Some researchers (as examples, Garcia-Villanova et al. 1997; Clark & Sivagenesan 1998) have estimated the spatial distribution of chlorine in a network by applying regression and/or correlation analyses. They have obtained useful relationships between the water-quality parameters and other specific parameters such as the reaction time, pH, and water temperature. However, this approach may not be suitable to automatically determine the spatial distribution of the chlorine concentration in a network.

Various inverse-based procedures have also been developed to estimate water-quality parameters, and clustering techniques that were previously applied to find solutions both in water networks (Bascià et al. 1999) and aquifers (Yeh 1986) have also been applied (Lee et al. 2004; Morga et al. 2007) to solve the problem of the spatial distribution of chlorine in networks. Lee et al. (2004) suggested the classification of sampling sites into groups with similar levels of chlorine. However, this type of approach is limited because the results are directly dependent on measurements and their locations. More recently, Aubry et al. (2011) proposed an FEM (finite element method) approach to solve the inverse problem connected with estimations of the water-quality coefficients. Although the procedure appears promising, more applications and experiments are required to generalize it. In contrast, Morga et al. (2007) used fuzzy logic (Ross 1995; Yang & Ko 1996) as a clustering technique by assuming as an initial condition an empirical

relationship between the residual chlorine and the flow velocity. This procedure is limited because it is only partially automatic, and the resulting zonation depends strongly on the values of the hydraulic variables (flow velocity and water pressure) obtained as outputs of the simulation process. However, the results obtained have substantially confirmed that, as noted by previous researchers (see as an example Silvert 2000; Li *et al*. 2013), fuzzy logic could be applied in controlling available information in water-quality management. Thus, Termini & Viviani (2012) recently improved the approach of Morga *et al*. (2007) by paying particular attention to the choice of the initial partition and the identification of the membership condition for the 'cluster'.

In recent years, there have been numerous attempts at online monitoring in water distribution systems. This is because of the technological innovation associated with data transfer and the fact that cost reductions for sensors allow the collection and automatic transfer of data from increasing numbers of sensors to the point of use. 'Online' modeling can be used to continually assess standards of service complaints for pressure and provide timely warnings of changes in the water-quality conditions. As an example, Machell *et al*. (2010) presented an online modeling package (Aquis) that is driven by near real-time flow and pressure field data gathered using GPRS communication. Landa *et al*. (2010) presented a real-time monitoring system for water distribution networks based on a new formulation of second-order chlorine decay equations, in which the notion of the concentration of the 'equivalent reducing reacting contaminant agent' plays the same role as the concentration of chlorine.

The point is that, because the online simulation models require a continuous feed of flow and pressure data from the real networks being simulated, the proper design and management of sensor systems are critical to achieve the necessary level of efficient and reliable operation. Consequently, the limitation of online monitoring is that it depends on the security of the sensor system performance, and well-designed tools for supporting constant system monitoring are required (see as an example Ediriweera & Marshall 2010).

Thus, the aim of the present work is to present an integrated procedure to determine the spatial distribution of residual chlorine in a real-size network. This procedure is based on interfaced codes (commercial and/or appositely elaborated) and includes a fuzzy-clustering procedure to automatically determine the unknown parameters. The advantage of this procedure is that it allows users to automatically identify (without any specific assumption) the chlorine level at any location of the network. The procedure includes the following steps: (1) the simulation step to estimate the flow rate and flow direction in each pipe of the network and the water pressure (hydrodynamic module) and the chlorine concentration (water-quality module) at each node of the network; (2) the clustering step to estimate the spatial diversity of water-quality parameters; and (3) the calibration step to estimate the value of the unknown parameters. It should be noted that the procedure has also the option of by-passing the simulation process for the water pressure and flow rate estimation at 'control nodes' of the network and applying the calibration of water-quality parameters by using real-time measured data.

This paper has to be viewed as an extension of the work by Termini & Viviani (2012). This paper is organized as follows. First, the simulation model is described. Then, the procedures for clustering and calibrating the water-quality parameters are explained. Finally, the procedure is applied to a sub-network in Palermo City (Italy), and a correlation analysis is used to assess the procedure's performance.

THE SIMULATION MODEL

Simulations have been performed with the help of the EPANET modeling software package developed by the US Environmental Protection Agency's (EPA) Water Supply and Water Resources Division (Rossman *et al*. 1993, 1994) and extensively used in the literature (among others Furlani & Marinelli 1998; Ozdemir & Ucak 2002; Castro & Neves 2003; Ho & Khalsa 2008).

The EPANET code includes two modules: a hydrodynamic module, which estimates the flow rate and flow direction in each pipe and the water pressure at each node of the network, and the water-quality module, which simulates the movement and transport of substances in the water under varying conditions.

The hydrodynamic module (EPA 2005) solves the following system of equations

– for each head tank of the network:

$$\begin{cases} \dfrac{\partial y_s}{\partial t} = \dfrac{q_s}{A_s} \\ q_s = \sum_i q_{is} - \sum_j q_{sj} \\ h_s = E_s + y_s \end{cases} \qquad (1)$$

– for each pipe i–j and for each node k:

$$\begin{cases} h_i - h_j = \varphi(q_{ij}) \\ \sum_i q_{ik} - \sum_j q_{kj} - Q_k = 0 \end{cases} \qquad (2)$$

where y_s is the water level at node-tank s, q_s is the entering flow rate at node-tank s, q_{ij} is the flow rate at a pipe connecting node i to node j, h_i is the hydraulic head at node i, A_s is the transverse area of node-tank s (which is infinite for the sources), E_s is the topographic level of node-tank s, Q_k is the water demand at node k, and φ is a function that expresses the relation between the head loss and the flow rate at each pipe. In the present study, the Hazen–William equation has been used as follows:

$$\varphi(q_{ij}) = 4.72 \varepsilon_{ij}^{-1.85} d_{ij}^{-4.87} L_{ij} q_{ij}^{1.85} \qquad (3)$$

where d_{ij}, L_{ij}, and ε_{ij} are the diameter, length, and roughness coefficient of the pipes i–j, respectively.

The water-quality module (EPA 2005) solves as follows
– the transport equation for each pipe i–j of the network:

$$\frac{\partial c_{ij}}{\partial t} = -\left(\frac{q_{ij}}{A_{ij}}\right)\left(\frac{\partial c_{ij}}{\partial x_{ij}}\right) + \vartheta(c_{ij}) \qquad (4)$$

– the mass conservation equation, for each node k:

$$c_{ij}(0, t) = \frac{\sum\limits_k q_{ki} c_{ki}(L, t) + V_i}{\sum\limits_k q_{ki} + Q_{si}} \qquad (5)$$

where c_{ij} is the substance concentration in pipes i–j, which is a function of distance x_{ij} along pipes i–j and of time t; A_{ij} is the cross-section of pipes i–j; V_i is the mass of the entering substance at node i; Q_{si} is the flow rate in the reach si; and ϑ is called the 'reaction term'

because it combines the decay reactions of the considered substance inside pipes i–j.

In this work, attention is given only to the residual chlorine, which is an excellent parameter for studying the water quality in a distribution system. The EPANET code simulates the water reaction, which determines the chlorine decay caused by its reaction with organic and inorganic substances, neglecting the presence of the pipe; the wall reaction, which is caused by the pipe's deterioration; and the mass transport toward the walls, which depends on the reactions determined by the exchange of mass between the pipe's walls and the flow. The application of the superimposition effects law determines the following reaction term (Datta & Sridharan 1994; Rossman *et al*. 1994; Yang & Ko 1996), which is included in the code:

$$\vartheta(c_{ij}) = -\left[k_b c_{ij} + \frac{k_w k_f}{r_h(k_w + k_f)}\right] c_{ij} \quad \text{with} \quad k_f = \frac{\text{Sh } D}{d_{ij}} \qquad (6)$$

where Sh is the Sherwood non-dimensional number that depends on the flow regime, D is the molecular diffusivity (equal to $1.21 \times 10^9 \text{ m}^2/\text{s}$), r_h is the hydraulic radius, and k_b and k_w are parameters that depend, respectively, on the kinetics developing within the water (bulk decay) and in the pipe's walls (wall decay). From Equation (6), it is clear that the parameter k_f is a function of the dimensionless Sherwood number, Sh (which depends exclusively on the hydraulic flow regime), the molecular diffusivity of chlorine in water, D, and the diameter of pipes i–j, d_{ij}. Consequently, k_f is uniquely determined once the simulation problem (1)–(3) is solved. In contrast, parameters k_b and k_w have to be estimated by using field data.

Thus, it can be concluded that the solution of the hydrodynamic module (1)–(3) depends on the estimation of parameter ε_{ij} (roughness coefficient of pipes i–j), while the solution of the water-quality module (4)–(6) depends on the estimation of parameters k_b and k_w.

SPATIAL DIVERSITY OF RESIDUAL CHLORINE

As mentioned in the 'Introduction', Morga *et al*. (2007) developed an empirical clustering technique based on the geometrical and hydraulic characteristics of the network.

Lee *et al.* (2004) identified the clusters based on a distance matrix (Euclidean distance) determining the distance to a cluster of selected sampling sites. The point is, to enable the automatic application of the procedure, it would be necessary to identify the 'clusters' according to a criterion that is not directly dependent on the values of variables from the simulation process, but that is defined from an arbitrarily chosen number of parameters.

With the aim of satisfying this requirement, in this work, an integrated fuzzy logic technique has been applied. Fuzzy logic (Ross 1995) is, in fact, particularly suitable for analyzing data sets characterized by the uncertainty caused by random factors. Specifically, in the present paper, the fuzzy C-means algorithm in the MATLAB code has been used (Termini & Viviani 2012). The algorithm consists of choosing an arbitrary initial fuzzy partition (or an arbitrary number of centroid partitions) and then iteratively minimizing the difference between two successive partitions. Such a procedure consists of minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade. Thus, the algorithm requires the definition of a 'membership' function for each cluster and the choice of an initial partition. The choice of the initial condition is particularly important for efficient model calibration/scenario runs.

In agreement with Fisher *et al.* (2011), the initial partition could be assumed on the basis of the chlorine concentration resulting from the initial chlorine dose. To apply a criterion that is not dependent on the initial chlorine dose, in this work, the 'membership' condition is defined by the matrix, U_{rc}, which consists of a number of rows, r, equal to the number of grouping parameters (i.e., the number of measured residual chlorine values at the consumption times t: $r = 1, \dots, t \times n$; with $n =$ the number of pipes) and a number of columns, c, equal to the number of initial parameters, nP, arbitrarily chosen (i.e., $c = 1, \dots, nP$). The elements of matrix U_{rc} have variable values between 0 and 1, depending on the degree of 'membership' of each pipe in each time considered ($0 =$ no membership). Thus, once the matrix is defined, the 'membership' condition for a particular cluster is identified by the maximum value of the degree of 'membership' of each pipe of the network.

This fuzzy-clustering procedure would also allow a user to take into account both the invariance of parameter

k_b over time and with demand and the introduction of a reacting parameter to estimate the variation of k_b caused by re-chlorination at every boosting location, according to specific second-order reaction equations (Kastl *et al.* 1999).

The advantage of the proposed cluster procedure is that it is a user-friendly, modular system for the dynamic simulation of complex multiple reactions, which may or may not be linked by hydraulic flows. Once the clusters are found, the optimal parameter values are derived by minimizing the sum of the squared difference between the measured chlorine concentrations and the model estimates at corresponding times during the simulation.

PARAMETER CALIBRATION

Roughness coefficient of pipes

The water pressure in each node and flow rate in each pipe of the network are estimated as the result of the simulation problem (1)–(3) (hydrodynamic module). To obtain reliable results, the parameter ε_{ij} has to be calibrated.

The calibration of parameter ϵ is performed by solving the inverse problem and imposing the following objective function (F.O.):

$$\text{F.O.} = \min[\sigma_h] \quad \text{with} \quad \sigma_h = \left\{ \sqrt{\frac{\sum_{i=1}^{n_t} (h_{i,m} - h_{i,s})^2}{n_t}} \right\} \tag{7}$$

where $h_{i,m}$ and $h_{i,s}$, respectively, represent the measured and estimated pressure heads at node i, σ_h is the mean square error between the measured and simulated water heads, and n_t is the number of nodes controlled by a remote control system. The decision variable of Equation (7) is the unknown parameter ϵ, and the state variables are the measured pressure heads. The optimization problem in Equation (7) is solved by applying the numerical code previously elaborated by Tucciarelli & Termini (1998). For the simulations performed in the present work, the daily water consumption law for each node is determined by considering the daily variation coefficient reported in Figure 1.
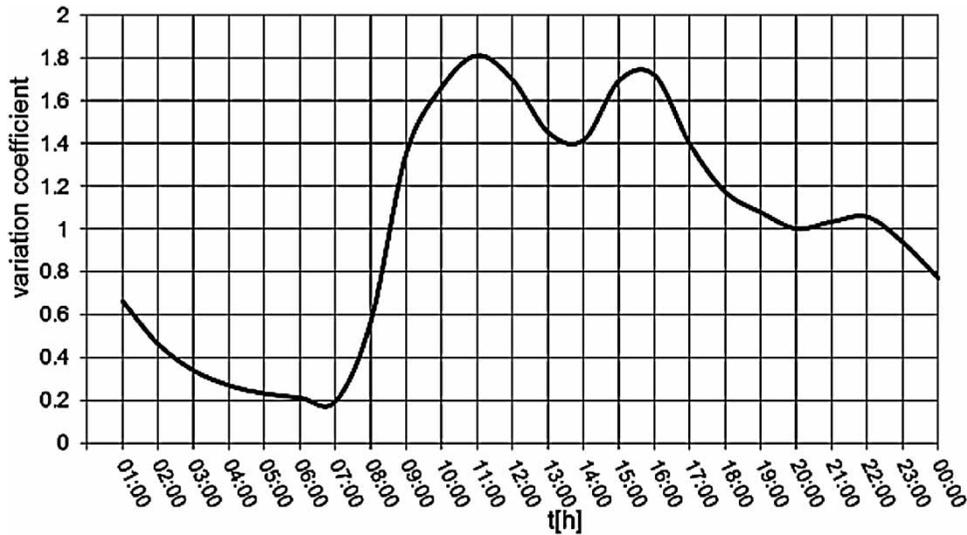
Figure 1 | Daily variation coefficient to estimate the daily water consumption law.

The daily water consumption law of Figure 1 was determined by applying a stochastic model calibrated in previous studies conducted by Fontanazza et al. (2006a, 2006b) by using measured water consumption data collected in three residential buildings in the same network being considered in the present work. The water consumption data were obtained using a volumetric pulse flow meter installed at the building's service connection to the distribution network considered for the application.

However, by installing an automated meter reading (AMR) device on the user's property, it would be possible to track the user's water consumption up to four times daily and to use the real-time data instead of the water consumption law (Figure 1). Thus, the use of an AMR system would allow a network manager to perform simulations using real-time data on a daily basis.

Water-quality parameters

The calibration of the water-quality parameters, k_b and k_w, is performed by solving the following optimization problem:

$$\text{F.O.} = \min[\sigma_C] \quad \text{with} \quad \sigma_C = \left\{ \sqrt{\frac{\sum_{i=1}^{n_m} (C_{i,m} - C_{i,s})^2}{n_m}} \right\} \quad (8)$$

where $C_{i,m}$ and $C_{i,s}$ are, respectively, the measured (for example, by a remote control system) and calculated chlorine concentrations, σ_C is the mean square error between the measured and calculated chlorine concentrations, and n_m is the number of measurements.

CASE STUDY

Characteristics of the network

An analysis has been conducted in the sub-network 'Oreto-station' in the city of Palermo (Italy). A plane-view of the network is given in Figure 2. This network includes 394 HDPE (high-density polyethylene) pipes (with diameters ranging from 110 to 225) and 263 nodes. The network is fed through two inflow pipes with diameters of 500, connecting inflow nodes 1 and 107 to the head reservoir (S. Ciro Basso – 54 m a.s.l.). The sub-network is provided with a remote control system that makes it possible to monitor the water pressure (and flow rate) at eight internal nodes (nodes 1, 47, 57, 107, 134, 162, 210, and 248) of the network. The chlorine concentration is also recorded through a chlorine residual meter installed at these nodes. Both the water pressure and the chlorine concentration data were made available by the municipal water service management company (AMAP s.p.a.) for the network.

Figure 2 │ Plane-view of sub-network and sampling sites (grey/red circles). The full color version of this figure is available online at http://www.iwaponline.com/jh/toc.htm.

To increase the number and the spatial distribution of the field data, the residual chlorine was also measured at other sites (opportunely selected) of the network. The sampling sites are also reported in Figure 2 (grey/red circles). The sampling sites were selected to include commercial users with direct links to the network and residential users without domestic storage tanks (the presence of such tanks could affect the results because of a stagnant water condition). The residual chlorine measurements were carried out using two multi-parameter photometers made by Hanna Instruments (mod. HI 9310). Three field measurement campaigns were conducted. The first (13 October 2004) included 13 measurement sites, while the second (8 February 2005) and third (14 March 2005) included 21 measurement sites. During each campaign, four simultaneous series of measurements were performed, spaced 3 hours apart (i.e., at 09:00, 12:00, 15:00, and 18:00). Finally, a total of 220 measurements was obtained.

Because the examined network is composed of new pipes made of the same material (HDPE), a unique value for the roughness coefficient, ϵ, was assumed for the whole network.

As a result of the optimization problem in Equations (1) and (2), a Hazen–Williams coefficient equal to 140 was obtained.

Application and results

To assess the goodness of the procedure, first, an empirical clustering technique based on the physical and hydraulic characteristics of the network and previously developed by Morga *et al.* (2007) was employed. Then, the results obtained were compared with those determined by applying the proposed procedure. In the following analysis, a constant value of k_b has been used. In agreement with other authors (Lee *et al.* 2004), it has been assumed that $k_b = 0.55\ \text{d}^{-1}$.

Empirical clustering and calibration

The analyzed network consists of new pipes made of the same material. Thus, in accord with Morga *et al.* (2007), the empirical zonation of k_w was performed only by considering the pipe's diameter and the flow velocity.

Some studies (Levi & Mallevialle 1995; Lu *et al.* 1995; Kiènè *et al.* 1998) have shown an inverse proportionality between the chlorine concentration and the ratio between the wet surface and the water volume per unit pipe length. This result might indicate a high consumption of chlorine in pipes with a small diameter. On the other hand, it has been observed that a high consumption of chlorine occurs with low flow velocities.

Thus, for the examined network, the first zonation map was produced by considering the distribution of the diameters of the pipes, as reported in Figure 3(a).

Then, for each considered consumption time (at 09:00, 12:00, 15:00, and 18:00), the flow velocity in each pipe was estimated by solving the hydrodynamic module shown in Equations (1)–(3). Thus, the pipes were grouped by fixing the variability ranges for the flow velocity, and a second zonation map was produced. Finally, the solution of the water-quality module, Equations (4)–(6), under the assumption of a unique value of $k_w = 0.625$ m/d (i.e., the mean value of k_w obtained from Equation (8) – see previous section) gave the chlorine concentration in each pipe of the network. Thus, the pipes were again grouped by fixing the variability ranges of the chlorine concentration so as to obtain a third zonation map. A comparison between the flow velocity–zonation map and the chlorine–zonation map allowed us to verify that, in accord with the previous literature results (Levi & Mallevialle 1995; Lu *et al.* 1995), the flow velocity–zonation and chlorine–zonation are very similar. Thus, a final velocity/chlorine zonation map for the network was obtained for each considered consumption time. As an example, Figure 3(b) reports the velocity/chlorine zonation map obtained at 09:00.

The final zonation of the network was defined by comparing the velocity/chlorine zonation map with the distribution of the pipe diameters. Finally, five zones were identified for the examined network (zone 1 = Policlinico, zone 2 = Archirafi, zone 3 = Perez, zone 4 = Oreto, and zone 5 = Stazione), as Figure 4 shows; the corresponding variability ranges of the flow velocity and chlorine concentration are reported in Table 1. From Table 1, it should be noted that, although zones 2 and 3 are characterized by the same flow velocity and chlorine concentration variability ranges, two different zones were considered because they were characterized by different pipe diameter distributions.

Then, the calibration of the 'zonated' k_w parameters was performed by using the measured chlorine data and solving the optimization problem, Equation (8).

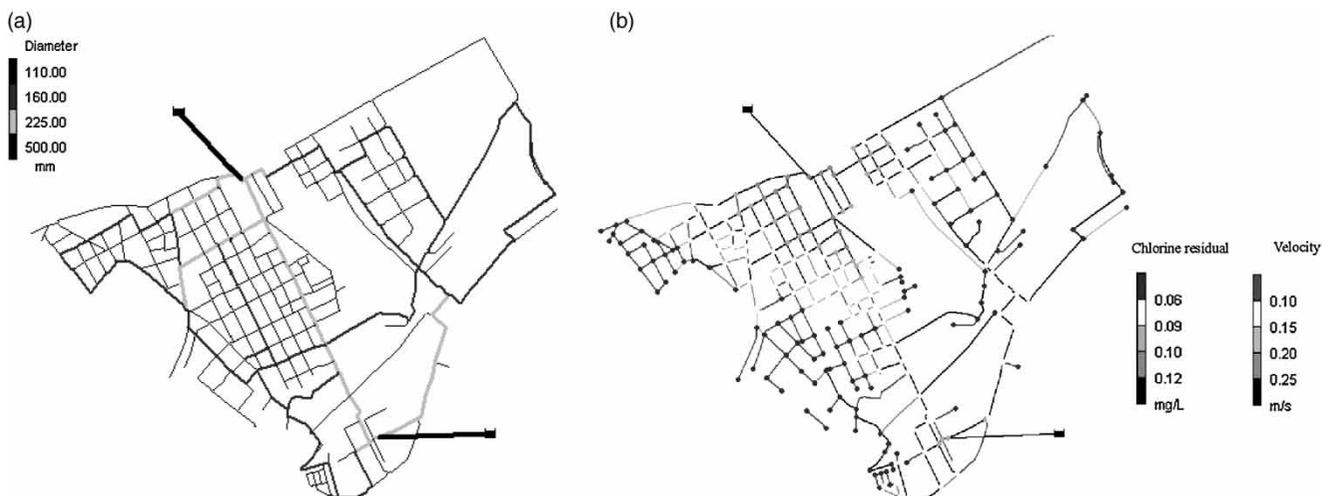Finally, to investigate the relation between the considered variables (the pipe's diameter and pipe's flow

Figure 3 | (a) Distribution of pipes diameters. (b) Velocity/chlorine zonation map (at 09:00).
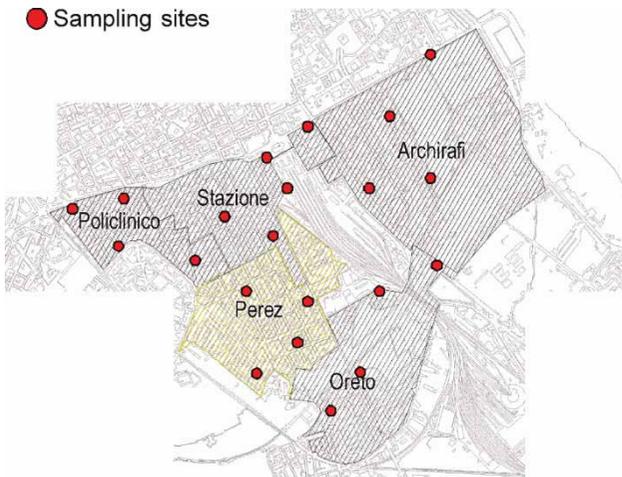
Figure 4 │ Empirical zonation.

Table 1 │ Empirical zonation: flow velocity and chlorine concentration variability ranges at 09:00

Zone	Flow velocity [m/s]	Chlorine [mg/l]
Policlinico	<0.1	<0.07
Archirafi	0.1–0.15	0.06–0.08
Perez	0.1–0.15	0.06–0.08
Oreto	0.15–0.2	0.09–0.09
Stazione	>0.2	>0.09

velocity) and parameter k_w, the weighted averages of the diameters and flow velocities of the pipes included in the pth zone (the weight was defined on the basis of the water volume in each included pipe), $\overline{d_p}$ and $\overline{U_p}$, were estimated. The regression analysis, performed using $\overline{d_p}$ and $\overline{U_p}$ as independent variables, yielded the following relationship (with $R^2 = 0.998$):

$$k_{w,p} = \alpha\overline{d_p}^{\,x} + \beta\overline{U_p}^{\,y} + \xi \qquad (9)$$

where the coefficients α, β, ξ, x, and y have to be determined by using experimental data. For the considered study case: $\alpha = 35.419$, $\beta = -259.820$, $\xi = -18.515$, $x = 0.308$, and $y = 6.313$.

The high value of the determination coefficient R^2 confirms the strong relation between the parameter k_w and the aforementioned variables (pipe's diameter and flow velocity), in agreement with the literature findings (Levi &

Mallevialle 1995; Lu et al. 1995; Kiènè et al. 1998). Equation (9) suggests that the spatial distribution of the unknown parameter, k_w, could be defined on the basis of this empirical relationship and by using simulated data. However, it should be clear that the use of Equation (9) is restricted to the data considered in the present work; further data would be necessary to generalize the expression.

Clustering by fuzzy logic and calibration

The empirical clustering procedure previously applied presents two main limitations: (1) the resulting zonation depends on the initial condition (determined by assuming a unique value of k_w for the entire network), and (2) it does not allow the user to work with a sufficient level of automation. Thus, clustering by fuzzy logic has been applied.

The algorithm requires the definition of a 'membership' function for each cluster and the choice of an initial partition. To assess the advantages of the proposed procedure, two initial partitions were considered for the application. First, the initial partition of Figure 4 (empirical initial partition) was applied, and the membership condition was defined on the basis of the estimated chlorine and flow velocity values. Then, the procedure was applied by considering an initial partition based on the definition of a 'membership' function.

As a result of the first application (i.e., by considering the initial empirical partition), four zones for parameter k_w were determined. The values of k_w, calibrated as the solution of Equation (8), are reported in Table 2.

Then, the procedure was applied again by considering, as the initial partition, the one determined through a specific 'membership' condition for each pipe of the network. The

Table 2 │ Calibrated values of k_w – by empirical initial partition (k_w) and by the 'membership' function ($k_{w,c}$)

Zones (empirical partition)	k_w [m/d]	Clusters (membership function partition)	$k_{w,c}$ [m/d]
1	0.2	3, 8, 9	0.2
2	0.5	2, 7	0.5
3	0.6	4, 5, 6	0.6
4	0.9	1, 10	0.9

'membership' condition was defined by the matrix, U_{rc} (because it is $t = 4$, $r = 1, \dots, 4 \times n$; and $c = 1, \dots, nP$). For the examined network, it was assumed that $nP = 10$. Thus, once the matrix is defined, the 'membership' condition for a particular cluster is identified by the maximum value of the degree of 'membership' of each pipe of the network.

As a result of the algorithm application, four zones were identified by grouping clusters 3, 8, and 9; clusters 4, 5, and 6; and clusters 1 and 10 of the initial partition. Then, the calibration of parameter k_w (indicated as $k_{w,c}$) was performed using Equation (8). Table 2 also reports the final zonation of the network and the value of parameter $k_{w,c}$ determined for each zone.

Table 2 highlights that the values of parameter $k_{w,c}$ obtained by applying the membership condition are equal to those obtained by applying the empirical zonation.

Sensitivity analysis of chlorine concentration with water-quality parameters

Parameter k_b

To verify the sensitivity of the chlorine concentration with parameter k_b, Equation (8) was solved by assuming a constant value for parameter k_w and a value of k_b that varied in the range of $0.1 \div 1.398 \, \text{d}^{-1}$ (Furlani & Marinelli 1998). As a result, very similar values for σ_C were obtained for different values of k_b. This can be observed from Table 3, which reports the values of σ_C obtained for different values of k_b. Thus, a low sensibility of the chlorine concentrations with k_b was observed.

Parameter k_w

To verify the sensitivity of the residual chlorine with the spatial variability of parameter k_w, Equation (8) was solved

by considering separately the data measured during each field campaign, by assuming a constant value for coefficient k_b and varying the value of k_w (which was unique for all pipes of the network). In agreement with the results obtained in previous works (Rossman et al. 1994; Furlani & Marinelli 1998), the variation in k_w was in the range 0.1–1.15 m/d. Table 4 reports the values of k_w and the corresponding values of σ_C obtained, for each case, as a result of the optimization problem of Equation (8). It can be noted from Table 4 that k_w assumes very similar values (it varies between 0.6 and 0.65 m/d), but the mean squared error assumes increasing values. This confirms that the spatial variability of k_w could determine different values of residual chlorine in a network.

CORRELATION ANALYSIS FOR SAMPLING SITES AND VALIDATION

To further verify the zonation determined by applying the aforementioned procedure, the corresponding associations among sampling sites were checked using a correlation analysis.

In fact, the sampling sites with a high correlation coefficient should be included in the same zone of the network (Lee et al. 2004). For this purpose, a correlation matrix was generated for all the sampling sites. The terms of the correlation matrix are the correlation coefficients of the chlorine concentration between a sampling site and the other ones. The calculated correlation matrix is shown in Table 5. The correlation matrix is a squared matrix (21×21) with diagonal terms equal to 1.0; the maximum (less than 1.0) value of the correlation matrix has been boxed. Table 5 shows that the sampling sites with a high correlation coefficient (squared) belong to the same zone of the network. Figure 5 reports the

Table 3 │ Parameter k_b and mean squared error σ_C

K_b [d^{-1}]	σ_C [mg/l]
0.1	0.027
0.3	0.026
0.55	0.028
1.398	0.029

Table 4 │ Parameter k_w and mean squared error σ_C

Field campaign	k_w [m/d]	σ_C [mg/l]
13 October 2004	0.6	0.023
8 February 2005	0.65	0.018
14 March 2005	0.6	0.016

Table 5 │ Correlation matrix between the sampling sites

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
P1	1.00	0.89	0.74	0.86	0.57	0.95	0.25	0.76	0.87	0.93	0.96	0.90	0.65	0.94	0.98	0.68	-0.43	0.84	0.81	0.55	0.84
P2	0.89	1.00	0.42	0.82	0.19	0.71	0.49	0.90	0.64	0.66	0.74	0.76	0.32	0.73	0.93	0.54	-0.18	0.54	0.58	0.61	0.60
P3	0.74	0.42	1.00	0.77	0.97	0.80	0.15	0.13	0.96	0.89	0.83	0.55	0.99	0.92	0.73	0.31	-0.88	0.98	0.97	0.57	0.97
P4	0.86	0.82	0.77	1.00	0.61	0.71	0.65	0.50	0.91	0.76	0.77	0.55	0.74	0.89	0.94	0.22	-0.70	0.79	0.90	0.89	0.89
P5	0.57	0.19	0.97	0.61	1.00	0.70	0.00	-0.08	0.87	0.80	0.71	0.42	0.99	0.80	0.54	0.23	-0.88	0.92	0.89	0.43	0.89
P6	0.95	0.71	0.80	0.71	0.70	1.00	-0.05	0.64	0.86	0.99	1.00	0.94	0.71	0.93	0.87	0.79	-0.43	0.90	0.80	0.32	0.83
P7	0.25	0.49	0.15	0.65	0.00	-0.05	1.00	0.16	0.32	0.00	0.03	-0.16	0.19	0.24	0.43	-0.47	-0.41	0.09	0.36	0.90	0.32
P8	0.76	0.90	0.13	0.50	-0.08	0.64	0.16	1.00	0.35	0.52	0.63	0.82	0.00	0.51	0.73	0.74	0.23	0.30	0.25	0.21	0.29
P9	0.87	0.64	0.96	0.91	0.87	0.86	0.32	0.35	1.00	0.92	0.89	0.65	0.93	0.98	0.88	0.36	-0.81	0.97	0.99	0.70	1.00
P10	0.93	0.66	0.89	0.76	0.80	0.99	0.00	0.52	0.92	1.00	0.99	0.87	0.82	0.97	0.87	0.69	-0.57	0.96	0.88	0.40	0.91
P11	0.96	0.74	0.83	0.77	0.71	1.00	0.03	0.63	0.89	0.99	1.00	0.92	0.74	0.96	0.90	0.75	-0.48	0.92	0.84	0.40	0.87
P12	0.90	0.76	0.55	0.55	0.42	0.94	-0.16	0.82	0.65	0.87	0.92	1.00	0.43	0.78	0.80	0.93	-0.10	0.70	0.56	0.13	0.61
P13	0.65	0.32	0.98	0.74	0.98	0.71	0.19	0.00	0.93	0.82	0.74	0.43	1.00	0.86	0.65	0.18	-0.93	0.94	0.96	0.59	0.95
P14	0.94	0.73	0.92	0.89	0.80	0.93	0.24	0.51	0.98	0.97	0.96	0.78	0.86	1.00	0.93	0.53	-0.69	0.97	0.95	0.62	0.97
P15	0.98	0.93	0.73	0.94	0.54	0.87	0.43	0.73	0.88	0.87	0.90	0.80	0.65	0.93	1.00	0.53	-0.49	0.81	0.84	0.70	0.85
P16	0.68	0.54	0.31	0.22	0.23	0.79	-0.47	0.74	0.36	0.69	0.75	0.94	0.18	0.53	0.53	1.00	0.19	0.48	0.27	-0.24	0.33
P17	-0.43	-0.18	-0.88	-0.70	-0.88	-0.43	-0.41	0.23	-0.81	-0.57	-0.48	-0.10	-0.93	-0.69	-0.49	0.19	1.00	-0.77	-0.87	-0.72	-0.85
P18	0.84	0.54	0.98	0.79	0.92	0.90	0.09	0.30	0.97	0.96	0.92	0.70	0.94	0.97	0.81	0.48	-0.77	1.00	0.96	0.51	0.97
P19	0.81	0.58	0.97	0.90	0.89	0.80	0.36	0.25	0.99	0.88	0.84	0.56	0.96	0.95	0.84	0.27	-0.87	0.96	1.00	0.73	1.00
P20	0.55	0.61	0.57	0.89	0.43	0.32	0.90	0.21	0.70	0.40	0.40	0.13	0.59	0.62	0.70	-0.24	-0.72	0.51	0.73	1.00	0.70
P21	0.84	0.60	0.97	0.89	0.89	0.83	0.32	0.29	1.00	0.91	0.87	0.61	0.95	0.97	0.85	0.33	-0.85	0.97	1.00	0.70	1.00

final zonation and the positions of the sampling sites. From Figure 5, it can be seen that the sampling sites that belong to the same zone are characterized by high values for the correlation coefficient. As an example, sampling sites P5 and P13 have a correlation coefficient equal to 0.98 and belong to the same zone.

Then, the good fit between measured and estimated values of chlorine concentration has been verified and the root mean squared error ($\sigma_{c,t}$) has been used as indicator. It is as follows:

$$\sigma_{c,t} = \sqrt{\frac{\sum_{i=1,N}\left(C_{i,m} - C_{i,s}\right)^2}{N}} \qquad (10)$$

where N is the total number of measurements considered. Figure 6 reports the pair of values ($C_{i,m}$ and $C_{i,s}$) and the line of perfect agreement (bisector line). The error bar is defined by the value of $\sigma_{c,t}$. As shown in Figure 6, with a few exceptions, the points arrange around the bisector line

and the $\sigma_{c,t}$ is quite low in comparison to the magnitude of the measured values.

CONCLUSION

This work considered the simulation of the decay law of chlorine in drinking water systems. In particular, attention was given to the identification of the spatial diversity of the water-quality parameters. The criterion for grouping the water-quality parameters was based on fuzzy logic. An integrated numerical technique, including interfaced codes (commercial and/or appositely elaborated), was applied to determine the spatial diversity of the residual chlorine in a sub-network in Palermo City (Italy). The procedure included the EPANET code for the estimation of flow rates, water pressure, and the transport of substances in the water of the network (hydrodynamic and water-quality modules); a C-means fuzzy algorithm for the pipe's clustering, and a

Figure 5 │ Final zonation by using the 'membership' function and sampling sites.
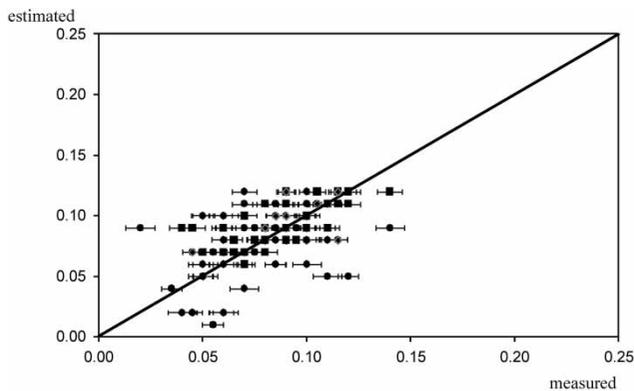

Figure 6 │ Comparison between measured and estimated values of chlorine concentration.

previously developed optimization algorithm (Tucciarelli & Termini 1998) for calibrating the unknown parameters. The advantage of the proposed procedure in comparison with others proposed in the literature is that it allows the user to automatically identify the zonation parameters of the network without any arbitrary assumption. The initial partition of the fuzzy-clustering algorithm is based on the identification of the membership condition for a 'cluster'.

The procedure has the option of by-passing the simulation process to determine the water pressure and flow rate at 'control nodes' of the network and applying the water-quality parameter calibration by using real-time measured data. Thus, the advantage of the proposed clustering procedure is that it is a user-friendly, modular system for the dynamic simulation of complex multiple reactions, which may or may not be linked by hydraulic flows.

The application of this procedure to a real-size network determined a zonation of the quality parameter k_w that was basically equal to that obtained by considering, as the initial partition, the empirical zonation determined by the empirical relationship between the residual chlorine and the flow velocity.

To assess the procedure, the associations among sampling sites were checked using a correlation analysis.

As a result, it was shown that the sampling sites with high correlation coefficients belonged to the same defined zone of the network.

REFERENCES

Aubry, D., Boudon, J. & Landa Paradinas, T. 2011 A FEM approach to water balance and quality parameters identification, towards real time networks monitoring. In: *CCWI 2011, September 5–7*, Exeter, UK.

Bascià, A., Termini, D. & Tucciarelli, T. 1999 A clustering technique for the parameter estimation in flow pipe network models. In: *CCWI '99, September 13–15*, Exeter, UK.

Biswas, P., Lu, C. & Clark, R. M. 1993 A model for chlorine concentration decay in pipes. *Water Research* **27** (12), 1715–1724.

Castro, P. & Neves, M. 2003 Chlorine decay in water distribution system, case study – Lousada network. *Electronic Journal of Environment, Agricultural and Food Chemistry* **2** (2), 261–266.

Clark, R. M. & Coyle, J. A. 1990 Measuring and modeling variations in distribution system water quality. *Journal of the American Water Works Association* **2** (46), 46–53.

Clark, R. M. & Sivagenesan, M. 1998 Predicting chlorine residuals and formation of TTHMs in drinking water. *Journal of Environmental Engineering* **124** (12), 1203–1210.

Datta, R. S. N. & Sridharan, K. 1994 Parameter estimation in water-distribution systems by least squares. *Journal of Water Resources Planning and Management, ASCE* **120** (4), 405–422.

De Rosa, S., Vasta, A. & Greco, S. 1998 Dinamica di Formazione del Biofilm Batterico in Condotte d'Acquedotto. *I.A. – Ingegneria Ambientale* **27** (11), 12. (Novembre-Dicembre (in Italian)).

Ediriweera, D. D. & Marshall, I. W. 2010 Monitoring water distribution systems: understanding and managing sensor networks. *Drinking Water Engineering and Science* **3**, 107–113.

Environmental Protection Agency (EPA). 2005 *Water Distribution System Analysis: Field Studies, Modeling and Management.* EPA/600/R-06/028, Cincinnati, OH, USA.

Fisher, I., Kastl, G. & Sathasivan, A. 2011 Evaluation of suitable chlorine bulk-decay models for water distribution systems. *Water Research* **45** (16), 4896–4908.

Fontanazza, C., Freni, G., LaLoggia, G. & Termini, D. 2006a Un modello previsionale della domanda idrica istantanea in condizioni di scarsità. In: *XXX Convegno di Idraulica e Costruzioni Idrauliche, September 11–15*, Roma, Italy (in Italian).

Fontanazza, C., Freni, G., LaLoggia, G. & Termini, D. 2006b Comparison of different stochastic models for urban water demand forecasting in drought conditions. In: *7th International Conference of Hydroinformatics – HIC 2006, September 6–8*, Nice, France.

Furlani, G. & Marinelli, A. 1998 Simulazione della Qualità dell'Acqua nella Rete d'Acquedotto della Città di Pesaro.

XXVI Convegno di Idraulica e Costruzioni Idrauliche, Catania (in Italian).

Garcia-Villanova, R. J., Garcia, C., Gomez, J. A., Garcia, M. P. & Ardanuy, R. 1997 Formation, evolution and modeling of trihalomethanes in the drinking water of a town. II: In the distribution system. *Water Resources* **31** (6), 1405–1413.

Ghinelli, P. 2004 Qualità microbiologica delle acque destinate al consumo umano, Infoambiente, sezione Provinciale di Parma (in Italian).

Hallam, N. B., West, J. R., Forster, C. F., Powell, J. C. & Spencer, I. 2002 The decay of chlorine associated with the pipe wall in water distribution systems. *Water Research* **36** (14), 3479–3488.

Ho, C. H. & Khalsa, S. S. 2008 EPANET-BAM: water quality modeling with incomplete mixing in pipe junctions. In Water Distribution Systems Analysis 2008 Conference, August 17–20, South Africa.

Kastl, G., Fisher, I. & Jegatheesan, V. 1999 Evaluation of chlorine kinetics expressions for drinking water distribution modeling. *Journal of Water Supply: Research and Technology – Aqua* **48** (6), 219–226.

Kiènè, L., Lèvi, W. & Lu, Y. 1998 Relative importance of the phenomena responsible for chlorine decay in drinking water distribution system. *Water Science and Technology* **38** (6), 219–277.

Kohpaei, A. J. & Sathasivan, A. 2011 Chlorine decay prediction in bulk water using the parallel second order model: an analytical solution development. *Chemical Engineering Journal* **171**, 232–241.

Landa, T., Sacareau, P., Boudon, J. & Gourbesville, P. 2010 Contribution to new approach of chlorine decay equations for real time monitoring of water quality in distribution networks. In: *9th International Conference on Hydroinformatics – HIC 2010, September 2010* Tianjin, China.

LeChevallier, M. W., Lowry, C. D. & Lee, R. G. 1990 Disinfecting biofilms in a model distribution system. *Journal of the American Water Works Association* **82** (7), 87–99.

Lee, L., Lu, C. & Kung, S.-L. 2004 Spatial diversity of chlorine residual in a drinking water distribution system. *Journal of Environmental Engineering, ASCE* **130** (11), 1263–1268.

Levi, Y. & Mallevialle, J. 1995 Prevention of bacterial regrowth in distribution systems. In: *Water Osaka '95, IWSA Specialized Conference on Advance Water Treatment and Integrated Water System Management, May 15–17*, Osaka, pp. 120–125.

Li, Z., Huang, G., Zhang, Y. & Li, Y. 2013 Inexact two-stage stochastic credibility constrained programming for water quality management. *Resources, Conservation and Recycling* **73**, 122–132.

Lu, C., Biswas, P. & Clark, R. M. 1995 Simultaneous transport of substrates, disinfectants and microorganisms in water pipes. *Water Research* **29** (3), 881–894.

Lu, C., Kiené, L. & Lévi, Y. 1999 Chlorine demand of biofilms in water distribution systems. *Water Research* **33** (39), 827–835.

Lund, V. & Ormerod, K. 1995 The influence of disinfection processes on biofilm formation in water distribution systems. *Water Research* **29** (4), 1013–1021.

Machell, J., Mounce, S. R. & Boxall, J. B. 2010 Online modelling of water distribution systems: a UK case study. *Drinking Water Engineering and Science* **3**, 21–27.

Morga, F., Termini, D. & Viviani, G. 2007 Calibrazione dei parametri di qualità nelle reti di approvvigionamento idrico e caratterizzazione della diversità spaziale–applicazione ad un caso di studio; Approvv. Distr. Idrica: Esperienza, Ricerca ed Innovazione; Ferrara (in Italian).

Munavalli, G. R. & Kumar, M. S. M. 2004 Dynamic simulation of multicomponent reaction transport in water distribution systems. *Water Research* **38** (6), 1971–1988.

Munavalli, G. R. & Kumar, M. S. M. 2005 Water quality parameter estimation in a distribution system under dynamic state. *Water Research* **39** (18), 4287–4298.

Ozdemir, O. N. & Demir, E. 2007 Experimental study of chlorine bulk decay in water supply pipes. *Journal of Hydraulic Research* **45** (6), 811–817.

Ozdemir, O. N. & Ger, A. M. 1998 Realistic numerical simulation of chlorine decay in pipes. *Water Research* **32** (11), 3307–3312.

Ozdemir, O. N. & Ucak, A. 2002 Simulation of chlorine decay in drinking-water distribution systems. *Journal of Environmental Engineering* **128** (1), 3307–3312.

Powell, J. C. & Hallam, N. B. 2000 Factors which control bulk chlorine decay. *Water Research* **34** (1), 117–126.

Ross, T. 1995 *Fuzzy Logic with Engineering Applications*. McGraw-Hill, New York.

Rossman, L. A., Boulus, P. F. & Altman, T. 1993 Discrete volume-element method for network water-quality models. *Journal of Water Resources Planning and Management* **119** (5), 505–517.

Rossman, L. A., Clark, R. M. & Grayman, W. M. 1994 Modeling chlorine residuals in drinking water distribution systems. *Journal of Environmental Engineering* **120** (4), 803–820.

Silvert, W. 2000 Fuzzy indices of environmental conditions. *Ecological Modelling* **130**, 111–119.

Termini, D. & Viviani, G. 2012 Spatial diversity of chlorine residual in a drinking water distribution system: a study case. In: *10th International Conference on Hydroinformatics – HIC 2012, July 14–18*, 2012 Hamburg.

Tucciarelli, T. & Termini, D. 1998 Optimal valves regulation for calibration of pipe networks models. In Hydroinformatics '98, August 24–26, Copenhagen.

Van der Kooij, D. 1992 Assimilable organic carbon as an indicator of bacterial regrowth. *Journal of the American Water Works Association* **84** (2), 57–65.

Yang, M. S. & Ko, C.-H. 1996 On a class of fuzzy c-numbers clustering procedures for fuzzy data. *Fuzzy Sets and Systems* **84**, 49–60.

Yeh, W. W.-G. 1986 Review of parameter identification procedures in groundwater hydrology: the inverse problem. *Water Resources Research* **22** (2), 95–108.