Near-optimal scheduling of device activation in water distribution systems to reduce the impact of a contamination event

S. Alvisi, M. Franchini, M. Gavanelli and M. Nonato

ABSTRACT

This paper proposes an innovative procedure for identifying, in the event of accidental or intentional contamination of a water distribution system, the optimal scheduling of activation of a pre-selected set of flow control devices which will serve to minimise the volume of contaminated water consumed by users after the detection of the contaminant in the system. The constraints are represented by the number of available response teams and the *maximum* speed at which these teams can travel along the roadway. The optimal scheduling of device activation is sought by means of an optimisation process based on a genetic algorithm (GA) which interacts with a *mixed integer linear programming* (MILP) solver in order to ensure the feasibility of the scheduling identified. The optimisation procedure is coupled to a hydraulic and quality simulator, which enables a calculation of the volumes of contaminated water consumed by users, and a dynamic cache memory, which, by storing information on the system's behaviour as the optimisation process progresses, serves to limit the computational times. The application of the procedure to a highly complex real water distribution system shows that the optimisation process is robust and efficacious and produces a smaller volume of contaminated water consumed by the users than when the activation of all the devices was completed in the shortest amount of time.

Key words | contamination, genetic algorithm, mixed integer linear programming, optimal scheduling, reaction, water distribution systems

S. Alvisi (corresponding author)
M. Franchini
M. Gavanelli
M. Nonato
Dipartimento di Ingegneria,
Università degli Studi di Ferrara,
Via Saragat, 1,
I-44122 Ferrara,
Italy
E-mail: *stefano.alvisi@unife.it*

NOTATION

CIR	Contaminant injection mass rate	$L_{\text{link},i}$	Length of the ith link
$c_{k,m}$	Concentration of contaminant at node m at time step k	l, u	Individuals in the population
DL	Concentration death limit	M	Constant represented by a 'large' term (sufficient to render the inequality of Equation (17) valid even when devices i and j are not visited in sequence by the same team)
dt_{in}	Duration of contaminant injection into the water distribution system		
$dt_{\text{op,hyd}}$	Interval of time necessary to open a hydrant	n_{dev}	Total number of devices to be operated
$dt_{\text{op,val}}$	Interval of time necessary to close a valve	n_{elite}	Number of individuals maintained from one generation to the next using elitism
$dt_{\text{op},i}$	Interval of time necessary to operate device i		
fit(l)	Fitness of the generic individual l	n_{hyd}	Number of hydrants to be opened
L_{ij}	Length of shortest travel route between device i and device j	n_{ind}	Number of individuals in the population
		n_{link}	Number of links to be closed

nn Number of demand nodes in the network

n_{opt} Number of parallel optimisations carried out in the *multi-start* approach

n_{scen1} Number of contamination scenarios considered for sensor placement

n_{scen2} Number of contamination scenarios associated with a sensor

n_{sens} Number of sensors located in the network

n_{teams} Number of available response teams

$n_{\Delta t,s}$ Number of computational time steps between t_{start} and the time of removal of the contaminant from the network in the generic scenario s

O.F.(l) Value of the objective function for the individual l

$p(l)$ Probability associated with the lth individual in a population

$q_{k,m}$ Demand at node m at time step k

rn Random number generated between 0 and 1

t_i Time in which activation of a device is completed

t_{start} Time at which the response teams set out from the point of departure in order to reach the devices to be operated.

V_s Volume of contaminated water consumed by users under scenario s

v_{team} Maximum travel speed of the teams along the roadway

$\mathbf{X}$ Matrix of binary variables in which partial ordering of device activation is encoded

x_{ij} binary variable equal to 1 if devices i and j are operated by the same team, with device j being operated immediately after device i, 0 otherwise

Δt_k kth computation step

$\Delta t_{\text{trav},i}$ Travel time for moving from one end of the link the to the other one

$\delta_{k,m}$ Variable that takes into account the 'contamination status' at node m at time step k

ε_i Distance between the ideal activation time of the ith device and the feasible one

τ_{ij} Minimum interval of time necessary to reach and operate device j on leaving device i

Ω Set of devices to be operated when a given sensor raises an alarm

Subscripts

a,b End nodes of a link to be closed

i,j,h Devices to be operated

k Time step

m Node of the distribution network

o Point of departure of the teams

s Contamination scenario

Superscripts

f Feasible time

INTRODUCTION

Water distribution systems are vulnerable to both intentional and accidental contamination events. Water distribution systems may, in fact, be characterised by a vast planimetric extent and have a large number of network nodes and pipes; solely protecting the most important elements of the system (tanks, water treatment plants, pumping stations, etc.) by measures that limit physical access to them is insufficient to ensure an acceptable level of safety for end users (Ostfeld & Salomons 2005). It is thus also important for utility companies to have tools and procedures which may enable them to deal effectively with a possible contamination event. From a practical standpoint, protecting a water distribution system against a contamination event entails two main steps: the first consists in detecting the presence of the contamination event, while the second is the response implemented with the aim of limiting its effects.

The presence of a contamination event can be determined by means of a system of sensors appropriately placed across the network and capable of detecting the presence of a contaminant (the presence of a contaminant is usually certified when its concentration is beyond a preset threshold). Once it has been ascertained that a risky/dangerous contamination event is underway it will be necessary to respond promptly in order to limit the negative effects of the contamination itself on the population served. The response can be implemented in different ways such as a simple

alerting of the population or the injection of substances into the system that can neutralise the effects of the contaminant; alternatively, the devices that control the flow and functioning of the system, such as isolation valves and hydrants, can be operated with the aim of limiting the spread of the contaminant and its consumption by users, while simultaneously facilitating its removal from the network (Baranowski & LeBoeuf 2006). These operations represent quantitative response actions, in opposition to the conservative approach of the entire shutdown of the system (Preis & Ostfeld 2008). In addition, it is worth noting that such an operation is not feasible when the system is extremely large and not divided into DMAs (District Metered Areas); furthermore many customers with special needs (hospitals, clinics, factories, etc.) may be severely damaged by a complete shutdown while fire fighting problems could arise.

Both the problems of detection and response to a contamination event have been addressed in the technical literature. In particular, as regards the problem of detecting the presence of a contamination event, numerous procedures have been proposed both for identifying the minimum number of stations necessary to guarantee *total coverage* of a water distribution system (see, for example, Lee & Deininger 1992), which are, however, applicable only theoretically and on small-sized networks, and for identifying the *optimal* placement of a pre-established number of stations in relation to different objectives (see, for example, Kumar *et al.* 1997). These objectives can be maximisation of the likelihood of event detection (see, for example, Ostfeld & Salomons 2004), minimisation of the time elapsing between the injection and detection of the contaminant, (see, for example, Watson *et al.* 2004) or minimisation of the contaminated volumes consumed by users (see, for example, Propato 2006). Moreover, these objectives are addressed either one at a time (see, for example, Ostfeld & Salomons 2004, 2005) so as to formulate a single-objective optimisation problem, or else combined (two or more objectives at a time) so as to formulate a multi-objective optimisation problem (see, for example, Ostfeld *et al.* 2008).

With respect to the response which follows detection of the contamination event (achieved by means of one of the previously mentioned procedures), there are fewer available studies compared with the number of studies dealing with

the first problem and their publication is on a very recent timescale. Among them, it is worth mentioning the studies by Baranowski & LeBoeuf (2006), Poulin *et al.* (2006, 2008, 2010), Preis & Ostfeld (2008), Guidorzi *et al.* (2009) and Alfonso *et al.* (2010). Baranowski & LeBoeuf (2006), in particular, propose a procedure based on use of the 'first-order reliability method' and 'parameter estimation method' (Doherty & Johnston 2003) to define the optimal distribution of demand/discharge, both entering and exiting, to be allocated to the nodes of a water distribution network in order to minimise the concentration of contaminants in the system under a *specific* contamination scenario (that is, assuming that it is known where, how and when the contamination occurred). Poulin *et al.* (2006, 2008, 2010) propose procedures based on a heuristic approach and on simple rules for isolating a contaminated area through the *simultaneous* closure of a number of valves in the system, assuming an *unlimited* number of response teams and subsequent removal of the contaminant by 'unidirectional flushing'. Preis & Ostfeld (2008) propose a multi-objective procedure for identifying the devices to be operated (hydrants to be opened and valves to be closed in order to isolate appropriate links), given a *specific* contamination event, with the aim of minimising the number of operations and the contaminated volume consumed following detection. It is worth pointing out that in the latter study no consideration is ever given to the problem of *when* to operate the devices (i.e. the scheduling of device activation) or of *how many* response teams are actually available to operate such devices.

Guidorzi *et al.* (2009) and Alfonso *et al.* (2010) similarly propose multi-objective procedures for a *preliminary* identification of the devices to be operated (among those present in the water distribution system) in order to minimise the number of operations and the contaminated volume consumed following detection. In these approaches, unlike in the one by Preis & Ostfeld (2008), when searching for the optimal set of devices to be operated, it is assumed that the characteristics of the ongoing contamination event, i.e. where, how and when the contamination occurred, are *not* known *a priori*. Whereas in Alfonso *et al.* (2010) the problem of *when* to operate the devices continues not to be taken into consideration, in Guidorzi *et al.* (2009) the problem is considered by looking for the optimal scheduling of

operation given a limited number of available teams. The search is carried out, however, by means of an 'enumerative' procedure among a limited number of solutions hypothesised *a priori* on the basis of pragmatic considerations and subjective assessments, and hence not by means of automatic optimisation processes.

Summing up, in the scientific literature the problem of sensor placement for detecting the presence of a contaminant in the system is the one that has been most frequently addressed. The subsequent problem of identifying which devices to operate in order to 'contain' the negative effects on the population has drawn some attention from the scientific community only recently. However, the problem of scheduling activation of the control devices once they have been identified has not yet been addressed by researchers, though it is of considerable importance as it fully characterises the response mode under conditions where the devices to be operated and the number of response teams put into action are known.

In this paper reference will be made to the study by Guidorzi *et al.* (2009) insofar as regards identification (a) of the sensors to be placed in the network and (b) of which devices should be activated in the event of an alarm generated by one of the sensors (without there being any knowledge of the contaminating source). With reference, therefore, to a context in which the location of the sensors is known, as are the devices to be activated when one of these triggers an alarm, a procedure is developed to enable automatic identification of the optimal scheduling of activation of the *assigned* set of devices (hydrants to be opened and valves to be closed in order to isolate the appropriate links) in order to minimise the contaminated volumes consumed by users *after* the presence of a contamination event (the source of which is unknown) *has been ascertained* (i.e. the preset concentration threshold has been exceeded) and given an assigned number of available response teams.

The following sections present the formulation of the problem considered and discuss its possible extrapolation to contexts other than the one taken as reference for its presentation. A description of the algorithm follows. The procedure is then applied to a real water distribution network, the results obtained are discussed and, finally, the conclusions are presented.

FORMULATION OF THE PROBLEM

The problem of identifying the optimal scheduling of activation of an assigned set of control devices is addressed in the following context:

- There exists a set of n_{sens} sensors capable of detecting the presence of a contaminant under a number of contamination scenarios (n_{scen1}). This set is identified using the procedure described by Guidorzi *et al.* (2009), who assume that: (a) each node of the water distribution system can be a point of injection, (b) injection can occur at the beginning of any of the time steps (usually one hour) into which the generic day is divided, (c) the duration of injection is fixed and equal to dt_{in} and (d) the contaminant injection mass rate (CIR) is fixed (see also Ostfeld & Salomons 2004);
- $n_{scen2} < n_{scen1}$ is the number of contamination scenarios that may be underway when a generic sensor first raises an alarm within a certain time interval (Guidorzi *et al.* 2009);
- the sensor that first raises an alarm within a certain time interval has associated with it a set Ω of devices to be activated. This set is identified on the basis of the procedure described by Guidorzi *et al.* (2009), which takes account of the n_{scen2} scenarios connected to the sensor considered.

It should be observed that the set Ω of devices to be operated consists of n_{hyd} hydrants to be opened in order to allow the contaminated water to be removed from the system and n_{link} links to be closed in order to limit the spread of the contaminant within the system and its consumption by users connected to the links considered, for a total of n_{dev} elements to be operated:

$$n_{dev} = n_{hyd} + n_{link}. \tag{1}$$

Let t_i (with $i = 1:n_{dev}$) indicate the time at which a hydrant is completely open, or the time at which a link is completely isolated. Hereinafter the times t_i will be called *activation times*, where activation means both the complete opening of the 'device' in the case of a hydrant and the complete closure of the 'device' in the case of a link. The scheduling of the activation times t_i (with $i = 1:n_{dev}$) is

defined with the objective of minimising the volume of con-taminated water consumed by users under the n_{scen2} contamination scenarios associated with the sensor considered.

The activation times t_i (with $i = 1{:}n_{\text{dev}}$) are counted start-ing from a reference time t_{start}, which represents the time at which the n_{teams} response teams set out from the point of departure in order to reach the devices to be operated (this time obviously falls after the time at which the sensor con-sidered detected the contaminant's presence in the system).

With reference to the generic scenario s among the n_{scen2} considered, let V_s be the volume of contaminated water consumed by users, calculated by taking into account the volumes consumed from t_{start} until the contaminant's removal from the network, i.e.

$$V_s = \sum_{k=1}^{n_{\Delta t,s}} \sum_{m=1}^{nn} \delta_{k,m} q_{k,m} \Delta t_k \qquad (2)$$

where $n_{\Delta t,s}$ is the number of computational time steps falling between t_{start} and the time of complete contaminant removal from the network in the generic scenario s, nn represents the number of demand nodes in the network, $q_{k,m}$ the demand at node m at time step k and Δt_k is the kth computational step (the latter being linked to the hydraulic and quality simulation model used). Finally, $\delta_{k,m}$ is a variable that takes account of the 'contamination status' at node m at time step k, defined as

$$\delta_{k,m} = \begin{cases} 0 & \text{if } c_{k,m} < \text{DL} \\ 1 & \text{if } c_{k,m} < \text{DL} \end{cases} \qquad (3)$$

where DL is the concentration death limit (Ostfeld & Salomons 2004) and $c_{k,m}$ the contaminant concentration at node m at time step k. On the basis of Equation (3) the volume of water consumed is contaminated only if the con-taminant concentration exceeds DL. More precisely, the contaminated volume of water that is considered in the pro-cedure is the one that would prove deadly for users (Guidorzi *et al.* 2009).

The problem of identifying device activation times t_i (with $i = 1{:}n_{\text{dev}}$) with the aim of minimising the volume of contaminated water consumed by users under the n_{scen2} contamination scenarios can thus be formulated as a

problem of minimising the following objective function (O.F.):

$$\text{O.F.}: \quad \frac{1}{n_{\text{scen2}}} \sum_{s=1}^{n_{\text{scen2}}} V_s = \frac{1}{n_{\text{scen2}}} \sum_{s=1}^{n_{\text{scen2}}} \sum_{k=1}^{n_{\Delta t,s}} \sum_{m=1}^{nn} \delta_{k,m} q_{k,m} \Delta t_k. \qquad (4)$$

The O.F. described by Equation (4) thus represents an average value of the contaminated volume consumed over the set of possible contamination scenarios connected to the sensor which first raised the alarm in a certain time interval (see Guidorzi *et al.* 2009).

The decision variables of the problem are thus rep-resented by the n_{dev} device activation times t_i. Incidentally, the decision variables, i.e. activation times of the devices in the network (hydrants to be opened and links to be closed), influence the value of the O.F. as they bring about a different hydraulic functioning of the system over time and hence a different contamination status $\delta_{k,m}$ at the generic node m (with $m = 1{:}nn$) at time step k (with $k = 1{:}n_{\Delta t,s}$).

The constraints of the problem are represented by the fact that (a) a pre-established number of teams n_{teams} are available to operate the devices, (b) all of the teams depart from the same point o at time t_{start} and (c) to get from point o to the first device and from one device to the next, the teams travel along a roadway at a speed that is not necessarily constant (depending on the activation times and the distance between two subsequent devices) but has a maximum limit equal to v_{team}, which in turn limits the minimum time interval between two subsequent activation times. In other words, the speed can be interpreted as an 'indirect' decisional variable which can vary between 0 and v_{team}.

Furthermore, still with reference to point (c), the road-way system is here reasonably assumed to be coincident with the layout of the water distribution network; thus the roads correspond to the links. This assumption is only prag-matic and not restrictive and thus the procedure described here can easily be extended to a case in which the roadway system does not (perfectly) coincide with the layout of the water distribution system. Finally, it is assumed that the 'operation' of each device requires a certain amount of time which will vary according to whether the device

concerned is a hydrant to be opened or a link to be isolated. In the case of a hydrant, the opening time is equal to $dt_{\mathrm{op,hyd}}$. In the case of a link to be closed, it is assumed that an isolation valve is present at both ends of the link and that they will be closed in sequence. More precisely, it is assumed that the valve located on the end nearest to the point the team is coming from will be closed first and that this operation will require a time $dt_{\mathrm{op,val}}$. The team will then travel to the other end at a *constant* speed equal to v_{team} and close the second valve in a time equal to $dt_{\mathrm{op,val}}$.

Incidentally, the closure in sequence of the valves located at either end of the link to be isolated will have the effect of altering the demand at the end nodes according to the scheme illustrated by Guidorzi *et al.* (2009), to which reference should be made.

Summing up, the term 'operation' means, in the case of a hydrant, the opening process that takes place in a time interval $dt_{\mathrm{op},i} = dt_{\mathrm{op,hyd}}$, and, in the case of a link, the closing process which is completed in a time interval equal to

$$dt_{\mathrm{op},i} = 2dt_{\mathrm{op,val}} + \frac{L_{\mathrm{link},i}}{v_{\mathrm{team}}} \tag{5}$$

where $L_{\mathrm{link},i}$ is the length of the link i to be closed. It is worth highlighting once again that during its closure the teams travel along the link at a *constant* speed equal to v_{team}. On the other hand, the teams travel at a speed that is lower than or equal to v_{team} when they go from one device to another. As a result, the time interval necessary to close a link is fixed and does not constitute a decision variable.

The imposition of these conditions, which reflect the actual conditions the response teams operate under, means that the device activation times are mutually constrained, i.e. not all of the times represent acceptable (or 'feasible') solutions of the problem. Formally speaking, remembering that the activation times t_i are calculated based on the mobilisation time t_{start} and assuming that the device i is the first device operated by a generic team, it follows that

$$t_i \geq \frac{L_{oi}}{v_{\mathrm{team}}} + dt_{\mathrm{op},i} \tag{6}$$

where $dt_{\mathrm{op},i}$ is the time it takes to complete the operation on

device i and L_{oi} is the length of the shortest travel route (along the water distribution system in our study) between the point of mobilisation o and device i.

Where i and j indicate two generic devices operated one after the other it follows that

$$t_j \geq t_i + \frac{L_{ij}}{v_{\mathrm{team}}} + dt_{\mathrm{op},j} \tag{7}$$

where L_{ij} is the length of the shortest travel route between device i and device j and $dt_{\mathrm{op},j}$ is the time it takes to complete the 'operation' on device j.

Summing up, a set of times $t_1, t_2, \ldots, t_i, \ldots, t_{n_{\mathrm{dev}}}$, for each of which the constraints represented by Equations (6) and/or (7) are met and where t_i represents the activation time of generic device i, constitutes a feasible solution of the problem and will be indicated hereinafter with the following set of symbols: $t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\mathrm{dev}}}^f$. It is worth highlighting that the above considerations imply that a solution is feasible when a *possible* division of tasks may be identified among the n_{teams} teams such that each team can activate the devices assigned to them in an order (first, second, etc.) that is consistent with the respective activation times, which are in turn consistent with the constraints expressed by Equations (6) and (7). This means that a solution $t_1, t_2, \ldots, t_i, \ldots, t_{n_{\mathrm{dev}}}$, once confirmed as feasible, will also give rise to a division of tasks among the teams. This aspect will be thoroughly discussed and clarified in the next section, in which the optimisation algorithm will be described.

To conclude the above considerations, it is important to observe that the formulation of the problem of identifying the optimal scheduling of activation of flow control devices, presented here as an outgrowth of the problem as formulated by Guidorzi *et al.* (2009), can also be extended to cases in which the set of contamination scenarios, locations of the sensors and the set of devices to be operated have been defined on the basis of assumptions and using procedures other than the those proposed in Guidorzi *et al.* (2009). Given the O.F. of Equation (4) and the decision variables and constraints described earlier, for the purpose of formulating the problem it is necessary and sufficient to fix a set of devices to be operated and a set of possible contamination scenarios under which the contaminated volume consumed by users needs to be minimised.

Finally, the formulation presented here can also be extended to the case in which the aim is to identify the sequence of activation of devices under a *single* contamination scenario: this situation refers to the case where the contamination source is known/identified (see, for instance, Guan *et al.* 2006; Preis & Ostfeld 2006; Liu *et al.* 2011). In this case as well it is necessary to fix the set of devices to be operated. The objective function of the problem expressed in Equation (4) is reduced in this case to volume V_s alone (see Equation (2)), where s indicates the single specific contamination scenario for which the sequence of device activation needs to be identified. Further considerations on this latter case will be developed in the section 'Analysis and discussion of the results'.

STEPS OF THE PROPOSED PROCEDURE

The procedure developed to solve the problem presented above is based on the use of a genetic algorithm (GA) optimiser, which 'manipulates', according its own 'internal logic', the scheduling of device activation, 'hybridised' with a mixed integer linear programming (MILP) solver whose function is to bring the scheduled sequence into line with 'feasibility' constraints as soon as they are generated by the GA. The optimisation package thus structured is coupled to (a) a hydraulic and quality simulator (EPANET, Rossman 2000) and (b) a cache memory: the hydraulic and quality simulator is used to compute the O.F., i.e. the contaminated volumes consumed under the contamination scenario(s) considered and given the current solution (control device activation times) by assuming a conservative contaminant (Guidorzi *et al.* 2009) and a purely convective transport scheme (Rossman 2000); the cache memory is used to store the O.F. values associated with solutions already analysed in previous computational steps in order to reduce computational times by avoiding the repetition of hydraulic simulations followed by the corresponding contaminant diffusion simulations. The choice of using GA is tied to their efficiency in solving optimisation problems in which the objective function is discontinuous, non-derivable and highly nonlinear, with a large number of local minima/

maxima and a domain that is not necessarily continuous, as in the case concerned.

It is important to observe that, at least to the authors' knowledge, there exist no other examples in the literature of optimisation procedures for identifying the optimal scheduling of device activation in a water distribution system on occurrence of a contamination event. However, there do exist several examples of GAs (and, more generally, of evolutionary algorithms) being used in problems analogous to the one considered here, such as, for example, the multiple travelling salesman problem (MTSP) (e.g. Lawler *et al.* 1985), the time irrigation scheduling problem or other similar problems in the hydro-science field (e.g. Nixon *et al.* 2001; Montesinos *et al.* 2002; Ulanicki *et al.* 2007).

The logical scheme characterising the procedure is presented here below, while the sections that follow provide a detailed description of the individual components used to search for the optimal solution, namely the GA and its operators, the cache memory and the MILP model. Figure 1 shows a block diagram of the proposed procedure, corresponding to the following logical steps:

(0) Initial values and parameter assignments.

(1) Creation of an initial population made up of n_{ind} individuals. This population is generated by means of a procedure described further in the next subsection which guarantees its feasibility. The activation times $t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\text{dev}}}^f$ of the n_{dev} devices to be operated are encoded in the generic solution, that is, in the generic individual l. The subsequent steps 2 and 3 are applied for each individual l of the population .

(2) Hydraulic and quality simulation relative to the contamination scenario(s) and set of activation times $t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\text{dev}}}^f$ of the n_{dev} devices encoded in the individual l and calculation of the value of the objective function (see Equation (2)/Equation (4)).

(3) Updating of the cache memory of solutions already analysed, where alongside the scheduling of activation $t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\text{dev}}}^f$ the corresponding value of the objective function is placed.

(4) Check on stopping criterion. If the stopping criterion (represented, for example, by the maximum

Figure 1 │ Block diagram of the proposed procedure.

time elapsed, the maximum number of generations or 'calls' of the hydraulic and quality simulation model) is satisfied, the procedure goes to step 5, otherwise to step 6.

(5)   The procedure terminates by outputting the near-optimal feasible solution: in addition to defining the scheduling of activation of the various devices, this solution also contains the division of tasks among the different teams, i.e. which devices are assigned to each team and the order (first, second, etc.) in which they are activated.

Steps 6–14 serve to create a new population made up of n_{ind} individuals.

(6)   Transfer of the n_{elite} best individuals of the current population into the new population (elitism). Steps 7–14 are applied iteratively until n_{ind}-n_{elite} individuals are created; together with the n_{elite} individuals inserted using elitism, they form the new population.

(7)   Selection of a pair of individuals in the current population (parents) which are called on to generate one/two individuals in the new population (offspring).

(8)   Generation of a random number between 0 and 1. If the number generated is less than 0.5 the procedure goes to step 9, otherwise to step 10.

(9)   Generation of a new individual (offspring) by means of a crossover operator called HX, described in the next subsection and based on the application of the mixed integer linear programming (MILP) solver. This operator leads to the generation of an individual for which the feasibility of the solution encoded in it is guaranteed. The procedure goes on to step 12.

(10)   Generation of two new individuals (offspring) by means of a crossover operator called UX, described in the next subsection. This operator leads to the generation of two individuals for which the feasibility of the solution encoded in them is *not* guaranteed.

(11)   Application of the MILP solver to the two individuals considered to check their feasibility and, when necessary, to render the solutions encoded in them feasible. This logic block can also be accessed after mutation (see step 13 below). In this latter case only one individual is 'elaborated'.

(12)   Check for the presence of the individual/individuals just generated in the current population. If the individual is already present in the current population, the procedure goes to step 13, otherwise the individual is inserted in the current population and the procedure goes to step 14.

(13)   Mutation of the individual. The mutation leads to the formation of a new individual for which the feasibility of the solution encoded in it is *not* guaranteed, so that in order to restore the feasibility of the solution encoded in the mutated individual the procedure goes back to step 11.

(14)   Check on the number of individuals present in the current population. If the number of n_{ind} individuals in the population has been reached, the procedure goes to step 15, otherwise to step 7.

   Steps 15 and 16 are applied for each individual l of the population just generated.

(15)   Check for the presence of the solution in the cache memory of solutions already analysed by means of the hydraulic and quality simulator in the previous optimisation steps. If the solution is already present, the procedure goes to step 16, otherwise to step 2.

(16)   The value of the objective function corresponding to the solution encoded in the individual is drawn from the cache memory and the procedure goes to step 4.

   The characteristics of the GA operators, cache memory and MILP model used to seek the optimal solution are presented in detail below.

Genetic algorithm

The general structure of a GA (Holland 1975; Goldberg 1989) can be summed up in the following steps (Davis 1991). The algorithm starts off with an initial population of n_{ind} individuals. The decision variables of the problem are encoded in the individual. Each individual is evaluated and the individuals in the population of the current generation are used to create new individuals via selection and reproduction operators inspired by genetics (crossover, mutation); these go to make up a new population, in which the best individuals of the current population may also be maintained (elitism). A detailed description of the encoding used and the individual GA operators accordingly developed follows, with specific reference to the problem considered here, in order to render the proposed algorithm efficient and robust.

Encoding

An integer-valued encoding is used. In particular, the individual contains a number of chromosomes equal to the number of the devices n_{dev} to be operated. Each chromosome is associated with a specific device and the corresponding activation time t is represented in it by an integer value. The integer-valued encoding is used as: (a) the hydraulic and quality simulations are conducted with a prefix computation step Δt; and (b) for simulation purposes the device activation time cannot fall within the generic computation step but only at the beginning or end of it. Thus, assuming for example a computation step $\Delta t = 1$ min, in the generic chromosome associated with the device i, the activation time t_i is expressed by means of an integer variable which represents the minute, counted from the reference time t_{start}, at which the procedure of device operation ends, i.e. the minute when the device is considered 'activated'.

Generation of the initial population

The generic individual of the initial population (see Figure 1, block 1) is generated by randomly attributing each device to one of the n_{teams} teams available and randomly generating the order in which the operations of each team are to be carried out. Based on the sequence/order of operation of the various devices assigned to each team, the corresponding activation times of each device are calculated taking into account the time necessary to operate the device dt_{op} and the speed at which the teams travel along the roadway, which *in this initial phase* where the first population is created (i.e. only for block 1) is assumed to be constant and

equal to v_{team}. The times thus obtained are rounded up to the nearest integer.

This method of generating the initial population guarantees a high variability in the activation times of each single device in the various individuals and moreover guarantees that the initial solutions will all be feasible, having been generated taking into account *a priori* the actual number of available teams and the constraints tied to the travel speed of the response teams and activation times of the devices introduced earlier.

Fitness

The fitness of an individual is closely connected to the corresponding value of the objective function (Davis 1991). In particular, with specific reference to the O.F. to be minimised considered in this problem, the fitness fit(l) of the generic individual l is defined as

$$\text{fit}(l) = \frac{1}{\text{O.F.}(l)} \tag{8}$$

with O.F.(l) representing the value of the objective function of the individual l. The probability $p(l)$ of choosing the individual l from among the n_{ind} making up the population used in the selection process is (Chelouah & Siarry 2003)

$$p(l) = \frac{\text{fit}(l)}{\sum_{u=1}^{n_{\text{ind}}} \text{fit}(u)}. \tag{9}$$

Selection

The individuals in the current population (parents) which will be called on to generate the individuals in the new population (offspring) (see Figure 1, block 7) are selected by means of the biased roulette wheel technique (Goldberg 1989), which uses the degree of adaptation (fitness) of each individual to ensure that the individuals with better fitness have a greater probability of being selected.

Crossover

In the GA proposed, two different types of crossover were implemented and used with equal probability. The first type, called UX (see Figure 1, block 10), is uniform crossover, which enables the creation of *two* offspring individuals. Once two individuals of the current population (parents) have been selected according to the above-described procedures, a zero–one vector of n_{dev} elements, i.e. made up of a number of elements equal to the number of chromosomes of an individual, is randomly generated. If the generic element is equal to 1, the corresponding component in the 'activation times' vector of the first offspring individual will coincide with that of the first parent, otherwise with that of the second parent. The opposite applies for the second offspring individual. Incidentally, with this type of crossover the activation time of the generic device encoded in each offspring individual coincides with that encoded in one of the two parents. However, given the random manner in which these times are recombined, even starting off from two parent individuals in which feasible solutions are encoded, the UX crossover could result in unfeasible solutions encoded in the offspring individuals. In this case the solutions must subsequently be rendered feasible by calling the MILP solver (see Figure 1, block 11).

The second type of crossover implemented, called HX (see Figure 1, block 9), enables only *one* offspring individual to be produced at a time: this offspring individual is directly generated as feasible. This latter crossover works in the following way: once two individuals of the current population (parents) have been selected, a hypercube is generated in the n_{dev}-dimensional space where two of the vertices of the hypercube are represented by the parent individuals, whereas the other vertices are represented by solutions obtained by combining the activation times of the devices encoded in the two parents, and the sides of the hypercube are segments parallel to the reference axes (see Figure 2 relative to the case where $n_{\text{dev}} = 3$). It is worth noting that the two vertices corresponding to the two parent individuals represent feasible solutions while the other vertices represent solutions which *may be* unfeasible.

The feasible offspring individual is generated by searching for the feasible point in the hyperspace which lies nearest to (or coincides with) one of the vertices of the hypercube and does not coincide with either of the two vertices represented by the parent individuals.

Formally speaking, assuming two generic parent individuals indicated, for example, by $g \equiv (t_{1,g}^{f}, t_{2,g}^{f}, \ldots, t_{i,g}^{f}, \ldots, t_{n_{\text{dev}},g}^{f})$
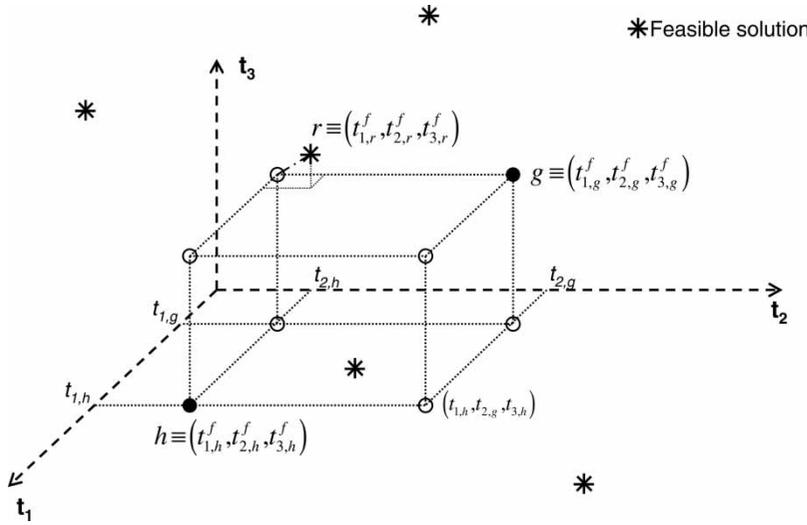
Figure 2 | Example of a three-dimensional hypercube created from two parent individuals $g \equiv (t^f_{1,g}, t^f_{2,g}, t^f_{3,g})$ and $h \equiv (t^f_{1,h}, t^f_{2,h}, t^f_{3,h})$ in order to generate, by applying crossover HX, the offspring individual $r \equiv (t^f_{1,r}, t^f_{2,r}, t^f_{3,r})$.

and $h \equiv (t^f_{1,h}, t^f_{2,h}, \ldots, t^f_{i,h}, \ldots, t^f_{n_{\mathrm{dev}},h})$, the individual offspring r will be given by the set of feasible activation times of the devices $(t^f_{1,r}, t^f_{2,r}, \ldots, t^f_{i,r}, \ldots, t^f_{n_{\mathrm{dev}},r})$ such that

$$\sum_{i=1}^{n_{\mathrm{dev}}} \min(|t^f_{i,r} - t^f_{i,g}|, |t^f_{i,r} - t^f_{i,h}|) \text{ is minimum}$$
subject to
$$\exists i : t^f_{i,r} \neq t^f_{i,g} \quad \text{and} \quad \exists j : t^f_{j,r} \neq t^f_{j,h} \tag{10}$$

In practical terms, the feasible point in hyperspace coinciding with or nearest to one of the vertices of the hypercube is identified by means of the MILP solver. If one or more vertices, beyond the two vertices corresponding to the parent individuals, represent feasible points, the *first* one identified by the MILP solver is used to generate the offspring individual independently of its position. For details concerning the solution of this problem via the MILP solver, reference should be made to section 'The mixed integer linear programming (MILP) model'.

It is important to stress here that two different types of crossover operations were used for balancing the necessity of guaranteeing variability in the new individuals without being too destructive. In fact, on the one hand, the UX crossover randomly combines the activation times of two parents and thus the feasibility must be restored through the MILP leading to *two* offspring individuals that are

rather different from both the parents (variability). On the other hand, the HX crossover, given its very nature, leads to *one* offspring individual that preserves, as far as possible, the information (activation times) encoded in both the parents and thus is more 'conservative' than the previous one. Given the structure of the two crossovers, 2/3 of the generic population are expected to be generated by the UX crossover, while only 1/3 by the other when the probability of calling either UX or HX crossover is equal to 0.5. This latter value was identified after several tests: it guarantees a good balance between variability (UX crossover) – which reduces the risk for the optimisation process of being trapped into local minima and at the same time makes the aging of the population slower – and the necessity of avoiding a too destructive generation process (HX crossover) as this would make the optimisation too slow.

Mutation

The mutation (see Figure 1, block 13) is achieved by swapping over the activation times of two devices randomly chosen from among the n_{dev} times that make up a solution. Incidentally, as in the case of crossover UX, this operation could result in a non-feasible solution, which will need to be subsequently modified and rendered feasible by calling the MILP solver (see Figure 1, block 11).

As previously mentioned, a mutation is activated every time an individual generated for inclusion in the new population is identical to one of the individuals already present in the new population (see Figure 1, block 12, and step 13 in this section).

Elitism

In the creation of a new population from the current population, the elitism technique was used (see Figure 1, block 6). Under this approach the n_{elite} best individuals of the current population, i.e. the individuals with the highest fitness scores, are preserved in the new population. The selection, crossover and mutation techniques described earlier are used to create n_{ind}–n_{elite} new individuals that will replace the worst in the current population.

Cache memory

The implementation of a cache memory was prompted by the observation that the computational times connected to the hydraulic and quality simulations account for the largest part of the computational times of the entire procedure. Thus, a check is performed on each individual of the population in which a feasible solution is encoded in order to determine whether it is already present in the cache memory (see Figure 1, block 15). If it is, the hydraulic and quality simulations are not run and the corresponding value of the objective function is drawn from the cache memory itself (see Figure 1, block 16); if it is not already present, the solution is analysed by running the simulations (see Figure 1, block 2) and subsequently inserted in the cache memory together with the corresponding value of the objective function (see Figure 1, block 3). With this mechanism it is possible to significantly reduce the computational times, in particular towards the final part of the optimisation process, when the population tends to converge, presenting solutions that are often identical to those already analysed in the previous populations (it should be borne in mind that the feasibility constraints increases the likelihood of the presence of identical individuals).

It is important to observe that checking for the solution's presence in the cache memory also entails a certain amount of computing time; therefore, in order to provide a significant benefit in terms of reducing total computational times the cache memory has to be structured in such a way as to enable an efficient check on whether or not a solution is present in it. Such considerations led to the adoption of a data structure based on a *binary search tree* (Knuth 1997).

The mixed integer linear programming (MILP) model

Within the framework of the proposed procedure, the MILP model is used to restore the feasibility (see Figure 1, block 11) of the solutions produced by uniform crossover UX (see Figure 1, block 10) or by mutation (see Figure 1, block 13) when they are not feasible, and to identify the solution produced by the crossover HX (see Figure 1, block 9).

Let $\mathbf{t} = (t_1, t_2, \ldots, t_i, \ldots, t_{n_{\text{dev}}})$ be the vector of device activation times representing the solution encoded in the generic individual produced by crossover UX or in the generic individual that has undergone a mutation, or the solution corresponding to a generic vertex of the hypercube. The MILP model is used to search for the vector $\mathbf{t^f} = (t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\text{dev}}}^f)$ which (a) respects the set of constraints that ensure its feasibility and (b) is *nearest*, in terms of norm 1, to vector $\mathbf{t}$.

The model is formulated in the following manner.

The unknowns are:

- The vector $\mathbf{t^f}$ (n_{dev}), made up of n_{dev} device activation times, all feasible;
- a matrix $\mathbf{X}$ ($n_{\text{dev}} + 1$, $n_{\text{dev}} + 1$) of binary variables 0–1, with $n_{\text{dev}} + 1$ representing the total number of points 'visited' by the teams, that is, the n_{dev} devices to be operated plus the point of mobilisation o or departure of the teams; the generic element x_{ij} of the matrix is equal to 1 if the devices i and j are operated by the same team, with device j being operated immediately after device i; device i is operated first by a team if $x_{oi} = 1$ (o point of mobilisation of the teams); finally, it is necessary that $x_{ii} = 0$ $\forall i$ (no self-looping arcs);

- the vector of the differences $\boldsymbol{\varepsilon}$ (n_{dev}), whose generic element ε_i is given by

$$\varepsilon_i = t_i - t_i^f, \quad i = 1 : n_{\text{dev}}. \tag{11}$$

The parameters are:

- the vector $\mathbf{t}$ (n_{dev}), made up of the n_{dev} device activation times to be 'approximated/modified' by means of a feasible solution;
- a matrix $\boldsymbol{\tau}$ ($n_{\text{dev}} + 1$, $n_{\text{dev}} + 1$), whose generic element τ_{ij} represents the minimum time necessary to get from device i to device j and to operate the latter device, whereas τ_{oi} represents the minimum time necessary to reach device i from the point of mobilisation o and to operate the device. In other words:

$$\tau_{ij} = \frac{L_{ij}}{v_{\text{team}}} + \mathrm{d}t_{op,j}. \tag{12}$$

The constraints are:

$$\sum_{i=1}^{n_{\text{dev}}} x_{oi} = n_{\text{team}} \tag{13}$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{n_{\text{dev}}} x_{ij} = \sum_{\substack{h=1 \\ h \neq i}}^{n_{\text{dev}}} x_{hi}, \qquad i = 1 : n_{\text{dev}} \tag{14}$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{n_{\text{dev}}} x_{ij} = 1, \qquad i = 1 : n_{\text{dev}} \tag{15}$$

$$t_i^f \geq \tau_{oi}, \qquad i = 1 : n_{\text{dev}} \tag{16}$$

$$t_j^f + (1 - x_{ij})M \geq t_i^f + \tau_{ij}, \qquad i,j = 1 : n_{\text{dev}}. \tag{17}$$

In detail, Equation (13) guarantees that n_{teams} teams depart from the point of mobilisation o; Equation (14) guarantees that the number of arrivals at a device is equal to the number of departures from the same device; Equation (15) guarantees that all of the devices are 'visited', and hence

activated, once only; finally, Equations (16) and (17) relate the device activation times based on the partial ordering encoded in the matrix $\mathbf{X}$. In particular, Equation (16) makes the activation time of the first device operated by each team dependent on the time τ_{oi} necessary to reach said device from the point of mobilisation o and operate it, whereas Equation (17), in the case where $x_{ij} = 1$ (i.e. devices i and j are activated by the same team, with device j being activated immediately after device i), makes the activation time of the generic device j dependent on the activation time of device i as a function of the time τ_{ij}; M is a constant represented by a term that is sufficiently large to render the inequality of Equation (17) valid even when i and j are not visited in sequence by the same team, i.e. when $x_{ij} = 0$ (so that the activation times of two generic devices i and j not operated in sequence by the same team are mutually unconstrained).

This set of constraints leads to a solution that can be interpreted as a tree with n_{teams} branches and a root which coincides with the mobilisation point. In particular, the branch of the tree corresponding to a generic team is made up of the subset of devices to be operated by that team; these devices are ordered along the branch starting from the root/mobilisation point following the order in which they must be activated by the team.

The objective is to minimise norm 1 of the difference between the vector $\mathbf{t^f} = (t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\text{dev}}}^f)$ of feasible device activation times and the vector $\mathbf{t} = (t_1, t_2, \ldots, t_i, \ldots, t_{n_{\text{dev}}})$, i.e.

$$\min\left(\sum_{i=1}^{n_{\text{dev}}} |t_i - t_i^f|\right) = \min\left(\sum_{i=1}^{n_{\text{dev}}} |\varepsilon_i|\right). \tag{18}$$

In order to linearise this objective function, finally, we introduce the unknowns ε_i^+ which meet the constraints $\varepsilon_i^+ \geq \varepsilon_i$ and $\varepsilon_i^- \geq (-\varepsilon_i)$ (i.e. $\varepsilon_i^+ = |\varepsilon_i|$) and we minimise their sum.

It is worth highlighting that, downstream of the application of the MILP model, given its structure and constraints, a sequence of feasible activation times $t_1^f, t_2^f, \ldots, t_i^f, \ldots, t_{n_{\text{dev}}}^f$ will be associated with a precise division of tasks among the n_{teams} teams; that is, each team will be assigned (a) a precise subset of devices to be operated and

(b) the order (first, second, etc.) in which they must be activated.

APPLICATION

The proposed procedure was applied to the water distribution system of the city of Ferrara, shown in Figure 3. The network extends for a total length of 180 km, is made up of 1,141 mains pipes and 780 nodes and serves a total population of about 100,000. The network is supplied from a single pumping station (see point *o* in Figure 3), which delivers water to a tank located inside the old city walls surrounding the centre of Ferrara and ensures a hydraulic head of approximately 28 m. It is worth noting that this water distribution system is part of a larger water distribution system that covers a very broad area, approximately 1,320 km², and serves 250,000 inhabitants of 11 municipalities in the province of Ferrara. It is important to highlight that the entire shutdown of the system considered in this study serving

the *city* of Ferrara is not a feasible solution for the water agency as in order to do this, the system serving the whole *province* of Ferrara should be shut down, involving technical problems (to hospitals, clinics, factories, etc.) and fire fighting risks.

Based on the results of the application of the procedure devised by Guidorzi *et al.* (2009) (described and applied to the same case study – see section 4 of the cited article), the location of 30 sensors is known after considering $n_{\text{scen1}} = $ 18,240 contamination scenarios. In particular, the contamination scenarios were defined assuming that all nodes could be points of injection and that injection could occur at the beginning of any hour of the day. The sensor locations are indicated in Figure 3. For one of these sensors, sensor 18 (see Figure 3), again following the procedure proposed by Guidorzi *et al.* (2009), we identified the set Ω of devices to be operated when that sensor first detects the contaminant's presence in the network in the time interval falling between 8 and 9 a.m. The set Ω is made up of $n_{\text{hyd}} = 7$ hydrants to be opened and $n_{\text{link}} = 6$ links to be closed, for a
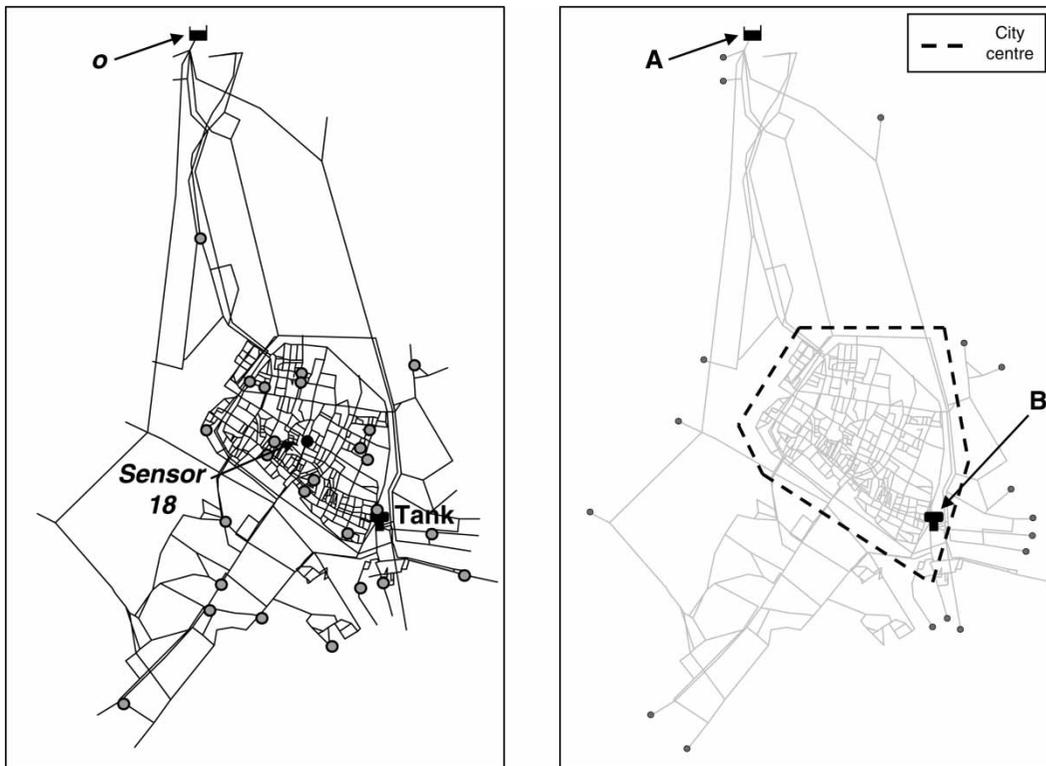

Figure 3 | Layout of the water distribution system serving the city of Ferrara (I). Point *o* hydraulically represents the supply point of the system and coincides with the point of mobilisation of the response teams. The grey circles indicate the sensor locations and the black one the location of sensor 18 considered in the subsequent part of the application.

total of $n_{dev} = 13$ devices to be operated (see Figure 4). Finally, Guidorzi *et al.* (2009) identify the set of $n_{scen2} = 42$ contamination scenarios which sensor 18 is first to detect in the time interval between 8 and 9 a.m. considered. In short, the set of $n_{dev} = 13$ devices to be operated given $n_{scen2} = 42$ different possible contamination scenarios is known.

As regards the constraints of the problem, based on indications provided by the technicians of the utility company that manages Ferrara's water distribution system, it was assumed that (a) the number of response teams n_{teams} available to operate the devices is equal to 3, (b) the roadways along which the teams travel coincide with the layout of the water distribution network, (c) the maximum speed of travel from one device to another (v_{team}) is 30 km/h and (d) the time it takes to open a hydrant ($dt_{op,hyd}$) and to close a valve ($dt_{op,val}$) is 3 min and the speed of travel from one end to the other of a link to be closed is constant and equal to $v_{team} = 30$ km/h.

It was further assumed that the point of departure of the teams coincides with the location of the pumping station that supplies the network (point *o* in Figure 3).

Finally, for the purposes of calculating the value of the objective function, i.e. the volume of contaminated water consumed, the death limit DL was assumed to be equal to 0.3 mg/l (Guidorzi *et al.* 2009).

The procedure was applied:

(a) assuming that the specific scenario underway was not known, which meant searching for the optimal scheduling of activation of the $n_{dev} = 13$ devices that *on average* minimised the volume of contaminated water consumed across the whole set of $n_{scen2} = 42$ possible contamination scenarios (see Equation (4));

(b) assuming that the specific scenario underway was known: such a situation could correspond to a case where the source is identified (e.g. Guan *et al.* 2006; Preis & Ostfeld 2006; Liu *et al.* 2011). In this latter case the procedure was applied in relation to each of the $n_{scen2} = 42$ contamination scenarios; this involved searching, for each scenario *s*, the optimal scheduling of activation of the $n_{dev} = 13$ devices that minimised the volume of contaminated water consumed (see Equation (2)).
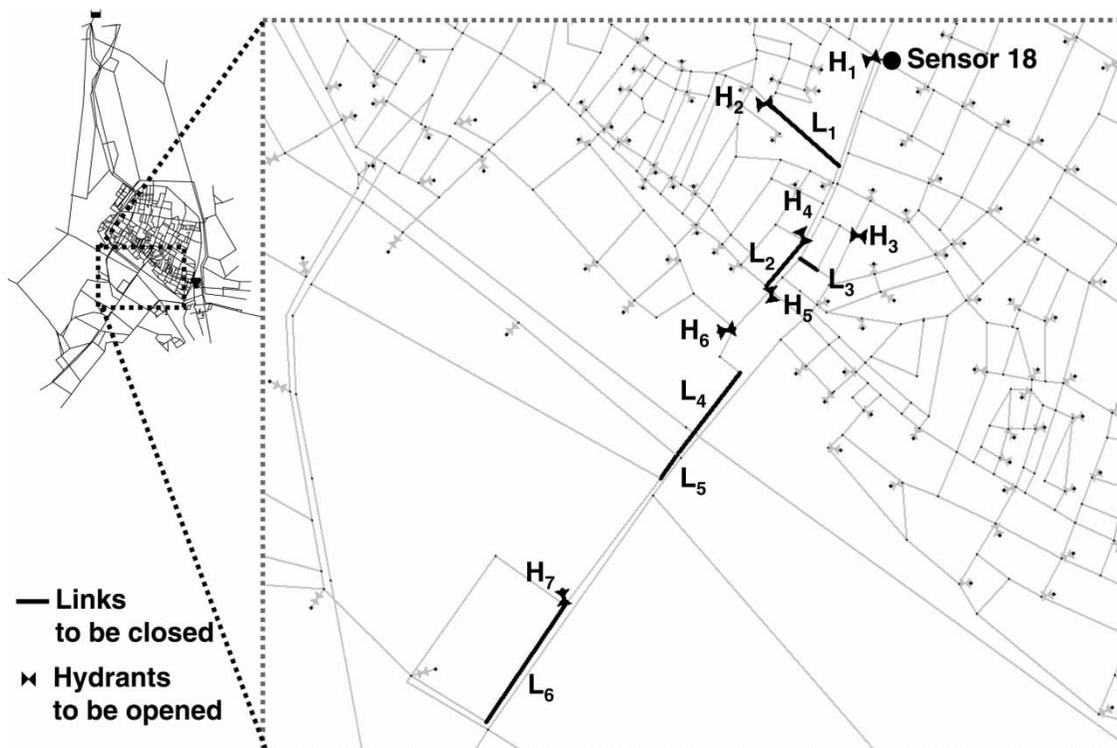

Figure 4 | Location within the water distribution system of the devices to be operated in the event of an alarm raised by sensor 18.

Furthermore, in both cases the procedure was applied by adopting both a *single-start* approach, in which a single optimisation is carried out, and a *multi-start* approach, in which n_{opt} optimisations are carried out in parallel (in this specific case $n_{opt} = 3$), each with a different initialisation seed, and the best solution was chosen from the final ones resulting from the n_{opt} optimisations. In order to make a meaningful comparison between these two different approaches, a constraint was imposed whereby each of the optimisations of the *multi-start* approach would have a '*duration*' (as better explained further below) equal to $1/n_{opt}$ (1/3 in this specific case) of the duration of the optimisation carried out in the *single-start* mode, so that the overall computational burden of the two approaches was equivalent. In particular, as the computational times associated with the hydraulic and quality simulation account for the largest portion of the total computational time of the entire procedure (in fact the generation of each individual of the population – which implies also the application of the MILP solver (see Figure 1, blocks 6–14) – takes less than 0.5 s, whereas the corresponding hydraulic and quality simulation for one contamination scenario (see Figure 1, block 2) takes nearly 7 s), the '*duration*' was assumed to be representable as the number of calls made to the simulator. Therefore, the procedure was stopped after 1,200 calls in the case of the *single-start* approach and at the 400th call in the case of the *multi-start* approach (three optimisations) so that the total number of calls made with both approaches was equal. The maximum number of calls was set equal to 1,200 for the *single-start* approach according to the results of several tests developed considering larger numbers of the maximum number of calls (up to 5,000 calls) and different initial seeds: in fact it was observed that 1,200 calls were sufficient to ensure a stabilisation of the objective function value. In other words, it was observed that the objective function value obtained by stopping the optimisation process at the 1,200th call was not far from the best known solution obtained considering many more calls (up to 5,000 calls) and different initial seeds.

A population of 20 individuals was assumed in all applications; FICO Xpress software (2009) was used for the resolution of the MILP model, whereas EPANET software (Rossman 2000) was used for the hydraulic and quality simulation.

Finally, in order to verify the efficacy of the proposed procedure, the two solutions corresponding to the operation of the devices (1) at times such as to minimise the total time necessary to complete all the operations (hereinafter indicated as 't-min') and (2) all simultaneously at time zero, that is when the teams start off from the point of mobilisation o (hereinafter indicated as 't-0'), were also considered as terms of comparison.

ANALYSIS AND DISCUSSION OF THE RESULTS

Table 1 shows the average contaminated volumes consumed considering the $n_{scen2} = 42$ possible contamination scenarios underway and assuming that the devices are activated (a) at the times corresponding to the optimal solution provided by the procedure using the *single-start* approach, (b) at the times corresponding to the optimal solution provided by the procedure using the *multi-start* approach, (c) at times such as to minimise the total time necessary to complete the operations on all the devices given the imposed constraints (solution 't-min') and (d) all (and simultaneously) at time zero, that is when the teams start off from the point of mobilisation o (solution 't-0'). Table 2 shows, for some contamination scenarios, the contaminated volumes consumed where scheduling has been optimised with reference to each specific contamination scenario using the proposed procedure with the (a) *single-start* and (b) *multi-start* approach; for these specific contamination scenarios, columns (c) and (d) show the contaminated volumes consumed when the scheduling corresponding to the solutions 't-min' and 't-0' is applied.

Table 1 | Average contaminated volumes consumed under the $n_{scen2} = 42$ possible contamination scenarios underway and considering that the devices are operated (a) at the times corresponding to the optimal solution provided by the procedure using the *single-start* approach, (b) at the times corresponding to the optimal solution provided by the procedure using the *multi-start* approach, (c) at the times corresponding to minimisation of the overall activation time of all devices and (d) at the time when the teams set out from the point of mobilisation o

	(a) Single-start (l)	(b) Multi-start (l)	(c) t-min (l)	(d) t-0 (l)
$\dfrac{1}{n_{scen2}} \displaystyle\sum_{s=1}^{n_{scen2}} V_s$	30,232	30,574	36,276	25,383

Table 2 | Contaminated volumes consumed under some of the $n_{scen2} = 42$ possible contamination scenarios underway and considering that the devices are operated (a) at the times corresponding to the optimal solution provided by the procedure using the *single-start* approach, (b) at the times corresponding to the optimal solution provided by the procedure using the *multi-start* approach, (c) at the times corresponding to minimisation of the overall activation time of all devices and (d) at the time when the teams set out from the point of mobilisation *o*. In the last row the average value of contaminated volumes consumed under each of the $n_{scen2} = 42$ possible contamination scenarios is reported

Scenario	(a) Single-start (l)	(b) Multi-start (l)	(c) t-min (l)	(d) t-0 (l)
1	16,922	15,439	28,065	15,164
2	8,006	7,853	15,503	2,756
3	26,435	25,932	36,750	24,182
4	12,644	12,544	21,442	9,790
5	47,122	47,618	53,344	42,916
6	22,511	22,530	33,477	21,159
7	23,199	22,446	39,558	18,667
8	34,675	36,004	44,165	30,032
9	26,029	25,989	37,435	23,373
10	35,265	35,677	46,009	35,101
…	…	….	…	…
41	7,453	7,160	14,761	2,756
42	42,076	42,513	43,958	37,462
Average	27,257	27,157	36,276	25,383

It is worth noting that solution 't-min' (see column (c) in Tables 1 and 2) represents a scheduling that meets the imposed constraints and depends on the location of the devices to be operated but *not* on the contamination scenario(s) considered. This scheduling is obtained by solving, by means of a linear solver (e.g. FICO Xpress (2009)), a problem of minimising the objective function represented by the activation time of the last device operated, in observance of the constraints represented by the time required by the teams to reach the devices and operate them. This scheduling represents a feasible reference solution of the problem that a manager or technician could intuitively apply with the aim of completing the response operations in as little time as possible.

Solution 't-0' (see column (d) in Tables 1 and 2) represents a second reference solution (equivalent to the one adopted in a number of studies to identify the devices to be operated in a network in case of contamination – see, for example, Preis & Ostfeld (2008) and Alfonso *et al.*

(2010)), but is unfeasible, given that no account is taken of the time it takes to reach the devices, the time necessary to operate them or the number of response teams that are actually available.

When comparing the results furnished by the procedure using the two different approaches, *single-start* and *multi-start* (see solutions (a) and (b), respectively), we can see that the procedure leads to solutions which have very similar contaminated volumes consumed associated with them, both in the case where optimisation of scheduling is carried out over the whole set of possible scenarios underway (Table 1), and in the case of optimisation for each specific scenario (Table 2). In the latter case, in particular, it may be observed that for some scenarios the *single-start* approach produces solutions associated with contaminated volumes consumed that are slightly lower than those resulting from the *multi-start* approach, while the opposite is true for other scenarios. Generally speaking, however, the differences are very modest and not such as to warrant an opinion that one approach is better than the other, at least from this point of view. However, the ease of parallel implementation of a *multi-start* approach and the growing prevalence of cloud computing techniques (see, for example, Moya *et al.* 2010; Xu *et al.* 2010) could make this approach preferable in that it reduces computational times (which are 1/3 as long as in the *single-start* approach in the case considered).

When comparing the contaminated volumes consumed resulting from the application of the optimal scheduling furnished by the procedure in the case of *optimisation over the whole set of possible scenarios* to the volumes derived from the reference solutions (c) (solution 't-min') and (d) (solution 't-0') we can see that the scheduling corresponding to solutions (a) and (b) leads to contaminated volumes consumed which are *smaller* (by about 6,000 l) than the volume corresponding to solution (c) (see Table 1) and *larger* (by about 5,000 l) than the volume corresponding to solution (d) (see Table 1). Similar results are also obtained in the case of *optimisation in relation to each specific scenario*, though in this case we observe a larger reduction in volumes (average reduction of about 9,000 l) when comparing solutions (a) and (b) vs. solution (c) (see last row in Table 2) and a smaller increase (average increase of about 2,000 l) when comparing solutions (a) and (b) vs. solution (d) (see last row in Table 2).

These results are understandable considering that, when the optimisation is carried out for one specific contamination scenario at a time, the procedure is capable of providing for each scenario a specific scheduling that effectively fits the contamination event considered, thus making it possible to limit the contaminated volumes consumed to values just above those that would result in the case of simultaneous activation (solution (d)) and distinctly below those associated with scheduling that minimises the overall device activation times (solution (c)). On the other hand, when optimisation is conducted over the whole set of scenarios, the procedure furnishes a single scheduling solution which, *on average*, allows the contaminated volumes consumed to be minimised over the whole set of scenarios, though clearly the contaminated volumes consumed associated with each specific scenario are slightly greater than the ones obtained where scheduling is optimised for that specific scenario alone. This is confirmed by the fact that the average value of contaminated volumes consumed under each contamination scenario where scheduling is optimised with reference specifically to the individual contamination scenarios (about 27,000 l for both the *single-start* and *multi-start* approaches – see last row in Table 2, columns (a) and (b)) is slightly lower than the average contaminated volume consumed where scheduling is optimised over the whole set of scenarios (about 30,000 l for both the *single-start* and *multi-start* approaches – see Table 1, columns (a) and (b)).

Incidentally, if no response action were to be taken, the average contaminated volume consumed over the whole set of contamination scenarios would be equal to approximately 355,000 l, that is one order of magnitude greater than the volume that would be obtained as a result of the response procedure.

In general, the results obtained show that the assumption adopted up to now in various studies, i.e. that all devices are simultaneously operated at time zero of the response, though unfeasible, results in contaminated volumes consumed which closely approximate, in order of magnitude, but underestimate in detail the volumes that would actually be obtained if an account were taken of the actual number of teams available to operate the devices. The results likewise show that the proposed procedure can provide a scheduling of device activation which would

result in significantly smaller contaminated volumes consumed compared to the feasible solution of minimisation of overall device activation time. In other words, the results obtained show that for this problem, at least with respect to the case study considered here, the principle of 'the sooner the better' does not apply. This is understandable considering that in a tightly meshed water distribution network such as the one analysed in this study, closing a valve too soon, for example, could divert the contaminant toward areas at higher risk, that is with higher consumption/demand, thus causing larger volumes of contaminated water to be consumed by users.

To conclude our presentation of the results, we feel it is important to point out once again that each feasible scheduling solution provided by the procedure corresponds to a division and ordering of tasks among the different teams; by way of example, Table 3 shows the one corresponding to the optimal solution provided by the optimisation procedure in the *single-start* mode considering all the $n_{\text{scen2}} = 42$ possible contamination scenarios underway (see Table 1, column (a)).

Finally, a note on the execution times. The optimal scheduling for a single contamination scenario with the *single-start* approach requires an execution time of around 2.5 h on a Pentium 4 with a 3.4 GHz CPU. The execution time is reduced to nearly 50 min, parallelising the process by running the optimisation with the *multi-start* approach. These execution times are not compatible with a reaction in real-time. Thus the procedure proposed can be useful to create, offline, a sort of scheduling database or a reaction manual indicating for each contamination scenario and/or for the whole set of possible contamination scenarios, the

Table 3 | Division and ordering of the devices to be operated (L: links to be closed, H: hydrants to be opened) among the $n_{\text{teams}} = 3$ teams corresponding to the optimal solution provided by the optimisation procedure using the *single-start* approach under the $n_{\text{scen2}} = 42$ possible contamination scenarios underway (see Table 1, column (a)). Shown in brackets is the activation time of each device expressed in minutes, counted starting from t_{start} (see Figure 4 for identification of the individual devices)

Team	Devices and corresponding activation times (min.)
1	L$_3$ (*16*), H$_1$ (*21*), L$_2$ (*29*), H$_5$ (*34*), H$_4$ (*38*)
2	L$_1$ (*16*), H$_6$ (*20*), L$_4$ (*28*), H$_3$ (*32*)
3	H$_7$ (*18*), L$_5$ (*26*), L$_6$ (*34*), H$_2$ (*40*)

corresponding scheduling of device activations together with the precise division of tasks among the teams, like the one shown in Table 3. The water agency manager or technician can then use this manual in a real-time situation both in the case where the source is unknown and where it is known: in this latter case, it is possible to recognise the most similar scenario previously analysed to that which is currently underway and refer to it to define the scheduling of operations.

CONCLUSIONS

This study addresses the problem of determining the optimal scheduling of activation of flow control devices in a water distribution system in the case of a contamination event (accidental or deliberate), given an assigned number of available response teams.

The proposed procedure for solving the problem is based on a GA coupled to a MILP solver, whose function is to bring the schedules generated by the GA into line with 'feasibility' constraints. The proposed procedure represents an innovative element in the technical literature within this sector, as it is typically assumed that the activations of devices in response to a contamination event take place simultaneously. However, such simultaneity is not possible given that the devices must generally be operated on site, utility companies have a limited number of response teams to carry out such operations and the teams must travel along a roadway at a speed which has an upper limit.

The results obtained when the procedure was applied to a complex real case showed it to be robust and efficacious. The procedure is capable of providing feasible solutions for scheduling the activation of devices in the system in such a way as to significantly contain the impact of the contamination event on the population served, which means limiting the contaminated volumes consumed.

It was observed, in particular, that the largest benefits are gained where the specific contamination scenario underway is known, but even where the specific contamination scenario underway is not known the procedure represents a useful tool for defining a feasible reference solution for scheduling activations so as to minimise the contaminated volume consumed on average over a broad range of possible contamination scenarios.

In both cases (known or unknown contamination scenario), the scheduling solutions provided by the procedure result in a smaller volume of contaminated water consumed than would be the case if activation of all devices were to be completed in the least amount of time possible. This result highlights that, given the complexity of the phenomenon of contaminant propagation in a tightly meshed distribution network and its nonlinearity, 'the sooner the better' is not necessarily true.

It is important to observe, moreover, that hydraulic simulations and quality simulations in particular entail lengthy computation. However, the procedure makes use of a cache memory which avoids the need to repeat numerous simulations corresponding to the same set of device activation times, as well as a *multi-start* approach which enables a simple parallel implementation of the procedure without impairing its efficacy; these represent valid means of reducing the computational burden. However, the execution times remain quite high and thus the proposed procedure cannot be applied in real time but, instead, is useful for creating offline a 'reference manual' of the optimal scheduling for each possible contamination scenario, together with the devices' assignment to each team and the order in which the devices must be activated, to be used by the water agency manager in situations of crisis.

To conclude two further considerations can be made. Firstly, it is worth noting that the solution identified is optimal with respect to the minimisation of the volume of contaminated water consumed; the solution, being feasible, is also accompanied by a division and ordering of tasks among the various teams, so that the devices can really be activated in the times indicated by the solution itself. Indeed, one could observe that the division of the tasks among the various teams is not optimised in order to minimise the travel of each team. This aspect is, however, secondary with respect to the minimisation of the volume of contaminated water consumed, which reasonably represents the major concern for a water agency manager/technician. However, if of interest, the division of the tasks among the various teams furnished by the procedure could be further improved *a posteriori* in order to minimise the transfers of the teams along the roadway.

Secondly, deterministic model inputs and parameters are assumed in the procedure here presented. Indeed, model inputs and parameters, such as water demands, pipe roughnesses, etc. can be affected by uncertainty and this uncertainty can be represented through many different techniques such as probabilistic, fuzzy or grey approaches (Alvisi & Franchini 2010). This aspect has not been considered in the presented study but represents a challenge for future research where uncertainty will be considered.

ACKNOWLEDGEMENT

REFERENCES

Alfonso, L., Jonoski, A. & Solomatine, D. 2010 Multiobjective optimization of operational responses for contaminant flushing in water distribution networks. *J. Wat. Res. Plann. Mngmnt.* **136** (1), 48–58.

Alvisi, S. & Franchini, M. 2010 Pipe roughness calibration in water distribution systems using grey numbers. *J. Hydroinf.* **12** (4), 424–445.

Baranowski, T. M. & LeBoeuf, E. J. 2006 Consequence management optimization for contaminant detection and isolation. *J. Wat. Res. Plann. Mngmnt.* **132** (4), 274–282.

Chelouah, R. & Siarry, P. 2003 Genetic and Nelder–Mead algorithms hybridized for a more accurate global optimization of continuous multiminima functions. *Eur. J. Oper. Res.* **148**, 335–348.

Davis, L. D. 1991 *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York.

Doherty, J. & Johnston, J. M. 2003 Methodologies for calibration and predictive analysis of a watershed model. *J. AWRA* **39** (2), 251–265.

FICO™ Xpress Optimization Suite 2009 *Xpress-Optimizer Reference Manual*. Available from: http://www.fico.com/en/Products/DMTools/Pages/FICO-Xpress-Optimization-Suite.aspx (accessed 9 September 2011).

Goldberg, D. E. 1989 *Genetic Algorithms in Search Optimization and Machines Learning*. Addison Wesley, Reading, MA.

Guan, J., Aral, M. M., Maslia, M. L. & Grayman, W. M. 2006 Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. *J. Wat. Res. Plann. Mngmnt.* **132** (4), 252–262.

Guidorzi, M., Franchini, M. & Alvisi, S. 2009 A multi-objective approach for detecting and responding to accidental and intentional contamination events in water distribution systems. *Urban Wat.* **6** (2), 115–135.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.

Knuth, D. 1997 *The Art of Computer Programming vol 3 Sorting and Searching*, 3rd edition. Addison-Wesley, Reading, MA, section 6.2.2, pp. 426–458.

Kumar, A., Kansal, M. L. & Arora, G. 1997 Identification of monitoring stations in water distribution system. *J. Environ. Engng.* **123** (8), 746–752.

Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. & Shimoys, D. 1985 *The Traveling Salesman Problem*. John Wiley & Sons, New York.

Lee, B. & Deininger, R. A. 1992 Optimal locations of monitoring stations in water distribution system. *J. Environ. Engng.* **118** (1), 4–16.

Liu, L., Sankarasubramanian, A. & Ranjithan, R. S. 2011 Logistic regression analysis to estimate contaminant sources in water distribution systems. *J. Hydroinf.* **13** (3), 545–557.

Montesinos, P., Camacho, E. & Alvarez, S. 2002 Application of genetic algorithms for optimal seasonal furrow irrigation. *J. Hydroinf.* **4** (3), 145–156.

Moya, V., Popescu, I., Solomatine, D. P. & Lobbrecht, A. 2010 Monte Carlo uncertainty analysis of hydraulic models using cloud computing. In *Proc 9th International Conference on Hydroinformatics HIC2010*. Chemical Industry Press, Tianjin, China, pp. 913–921.

Nixon, J. B., Dandy, G. C. & Simpson, A. R. 2001 A genetic algorithm for optimizing off-farm irrigation scheduling. *J. Hydroinf.* **3** (1), 11–22.

Ostfeld, A. & Salomons, E. 2004 Optimal layout of early warning detection stations for water distribution systems security. *J. Wat. Res. Plann. Mngmnt.* **130** (5), 377–385.

Ostfeld, A. & Salomons, E. 2005 Securing water distribution systems using online contamination monitoring. *J. Wat. Res. Plann. Mngmnt.* **131** (5), 402–405.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J. P., Dorini, G., Jonkergouw, P., Kapelan1, Z., di Pierro, F., Khu, S. T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., Van Briesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y. & Walski, T. 2008 The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms. *J. Wat. Res. Plann. Mngmnt.* **134** (6), 556–568.

Poulin, A., Mailhot, A., Grondin, P., Delorme, L. & Villeneuve, J. P. 2006 Optimization of operational response to contamination in water networks. In *Proc. 8th Annual Water Distribution Systems Analysis Symposium. CD-ROM*. University of Cincinnati, Cincinnati, OH.

Poulin, A., Mailhot, A., Grondin, P., Delorme, L., Periche, N. & Villeneuve, J. P. 2008 A heuristic approach for operational response to drinking water contamination. *J. Wat. Res. Plann. Mngmnt.* **134** (5), 457–465.

Poulin, A., Mailhot, A., Periche, N., Delorme, L. & Villeneuve, J. P. 2010 Planning unidirectional flushing operations as a response to drinking water distribution system contamination. *J. Wat. Res. Plann. Mngmnt.* **136** (6), 647–657.

Preis, A. & Ostfeld, A. 2006 Contamination source identification in water systems: a hybrid model trees-linear programming scheme. *J. Wat. Res. Plann. Mngmnt.* **132** (4), 263–273.

Preis, A. & Ostfeld, A. 2008 Multiobjective contaminant response modeling for water distribution systems security. *J. Hydroinf.* **10** (4), 267–274.

Propato, M. 2006 Contamination warning in water networks: general mixed-integer linear models for sensor location design. *J. Wat. Res. Plann. Mngmnt.* **132** (4), 225–233.

Rossman, L. A. 2000 *EPANET 2 User's Manual*. National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH.

Xu, Z., Velez, C., Solomatine, D. P. & Lobbrecht, A. 2010 Use of cloud computing for optimal design of urban wastewater systems. In *Proc. 9th International Conference on Hydroinformatics HIC2010*. Chemical Industry Press, Tianjin, China, pp. 930–938.

Ulanicki, B., Kahler, J. & See, H. 2007 Dynamic optimization approach for solving an optimal scheduling problem in water distribution systems. *J. Wat. Res. Plann. Mngmnt.* **133** (1), 23–32.

Watson, J. P., Greenberg, H. J. & Hart, W. E. 2004 A multiple-objective analysis of sensor placement optimization in water networks. In *Proc. World Water and Environment Resources Conference. CD-ROM*. ASCE, Reston, VA.