Evolutionary optimization for water losses recognition in water supply networks

Journal:	*European Journal of Environmental and Civil Engineering*
Manuscript ID:	Draft
Manuscript Type:	Original Papers
Date Submitted by the Author:	n/a
Complete List of Authors:	Mambretti, Stefano; Politecnico di Milano, DICA Becciu, Gianfranco; Politecnico di Milano, Martins, Paulo; Universidade Estadual de Campinas,
Keywords:	leakages, water supply network, modelling, management

Evolutionary optimization for water losses recognition in water supply networks

G. Becciu[1], S. Mambretti[1], P.S. Martins[2]

[1] *DICA – Politecnico di Milano, Italy*

[2] *School of Technology, UNICAMP, Limeira, Brazil*

Gianfranco Becciu
DICA – Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano
Italy
Email: gianfranco.becciu@polimi.it

Stefano Mambretti
DICA – Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano
Italy
Email: stefano.mambretti@polimi.it

Paulo Martins
Faculdade de Tecnologia – UNICAMP
Rua Paschoal Marmo, 1888
Jd Nova Italia
CEP: 13484-332
Limeira, SP - Brasil
Email: pmartins@ft.unicamp.br

Evolutionary optimization for water losses recognition in water supply networks

A methodology to localize the losses in the water supply networks has been developed, which requires the installation of a number of flowmeters and pressure transducers on the network and the building of a numerical model. The calibration of the model to match the recorded network parameters (pressures and discharges) is done by searching an optimal set of water demands at network nodes. The comparison between the optimal set and the standard one allows the identification of the areas where the leakages are most likely to be present. The optimal set of water demands is identified by the minimization of an objective function.

In the paper the coupling of this objective function with three evolutionary optimization methods, based on Simulated Annealing (SA), Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have been discussed and tested on a case study.

The simulations show SA risks to be trapped in unfeasible zones in its search, while the methods based on GA and PSO perform very well, because in these latter methods the individuals constituting a population work mainly in groups.

Moreover, the solution obtained by GA and PSO can be further improved by means of a simple Hill Climbing (HC) procedure.

Considerations on the possibility of having more than one maximum of the Objective Function and how they can be detected are presented.

Keywords: leakages; water supply network; modelling; management.

1 Introduction

Water loss in water distribution networks is gaining more attention recently in the research community due to their scale (up to 50% - 70% in some Countries) and economic impact on the society. Non-Revenue Water (NRW) or lost water is the difference between the volume entering a distribution system and the volume billed to customers. This volume is a serious economic damage for the companies, and the challenge is compounded by the fact that sources might become scarcer due to pollution and the increase in demand. To this end, methodologies that aim at detecting, predicting, preventing or avoiding water losses are welcome, in order to help management to make well-informed decisions and ultimately mitigate (or eliminate) this problem.

In particular, management systems need to know where and how to intervene (e.g. repair or substitution of a pipe) (Alvisi *et al.*, 2006), a challenge that is usually formulated as a multi-objective optimization problem. The objective function (O.F) is represented by the performance of the network and the costs of the rehabilitation (Halhal *et al.*, 1999; Giustolisi *et al.*, 2006). Common objectives functions are the not-delivered water volumes or the number of customers affected by interruptions caused by pipe bursts (Engelhardt *et al.*, 2000).

Such condition led to the development of models that are either able to detect the position of pipe breaks (Alvisi *et al.*, 2006) or have available good databases about previous breakages (Male *et al.*, 1990; Sundahl, 1996). Another objective to be pursued is the increase of the network efficiency through the reduction of water losses. However, the limited funds available constrain the invested annual budget and increase the importance of scheduling interventions.

Reduction of water losses have been considered using an appropriate pressure management (Walski *et al.*, 2006; Puust *et al.*, 2010), while Almandoz *et al.* (2005) proposed a method based on a water balance while and Wu and Sage (2006) applied Genetic Algorithms in order to calibrate a mathematical model, which have to be applied to a subarea (district) in which the incoming discharge has to be known.

Approaches similar to that presented in this paper can be found in Islam *et al*. (2011) who used a fuzzy based technique to analyse the losses and the other uncertainties of the network; or in Aksela *et al.* (2009) who used the self-organized map (SOM) method. A comprehensive review of the methods to detect and manage the leakages is presented in Puust *et al.*(2010).

A methodology that identifies the areas where losses are mostly expected has already been presented (Mambretti and Orsi, 2012); it is based on data collection (discharge and pressure) from instruments positioned on the water supply network, and successive comparison of the data collected with those simulated by software. The results of the model should match the readings of the instruments. With this proposed method, it is possible to identify the position of breakages, for smaller networks without districtualization, and for larger network, with larger areas of each district. Under the hypothesis that the model is a good representation of the real network, the differences between simulated and recorded data are due to the different demands imposed at the nodes. After a description of the method, the paper focuses on different methods of Evolutionary Computation in order to establish the best procedure to minimize the Objective Function.

The described method has also been applied to the case study of a real water supply network, in the town of Castegnato (North of Italy) and assuming five different scenarios of losses, which have been reconstructed with the different algorithms of EC.

Moreover, we assessed whether or not the value of the Objective Function is a reliable indicator of the goodness of the presented solution, and whether there are more than one maximum (and therefore whether the solution can be univocally determined).

The remainder of this paper is organized as follows: Section 2 presents the problem and describes the methods we propose for its solution; Section 3 discusses the optimization methods used in this work. Section 4 introduces the case study (city of Castegnato), while in Section 5 we discuss the results, with some comments related to the unicity of the optimum in Section 6. Finally, in Section 7 we present our remarks and conclusions.

2 Model and problem definition

The goal to be pursued in this work is the increase of the network efficiency through the reduction of water losses. The limited funds available constrain the invested annual budget and increase the importance of prioritizing in scheduling rehabilitation works.

The methodology employed consists of the following major steps (Figure 1).

(1) *Evolutionary computation*: in this step**,** the evolutionary algorithm generates a population of outflows Q_i^{out};

(2) *Hydraulic Analysis*: the hydraulic analysis calculates the values of Q_i^{comp} and H_i^{comp} for a given set of synthetic outflows Q_i^{out};

(3) *Calibration*: in this step the O.F. is computed;

(4) *Loss analysis*: Calculates the losses Q_i^{loss} for all the links L_i where a loss is identified. Once the model is calibrated, a loss in a node[1] N_i is identified and obtained from the difference between the initial values of $Q_i^{out\ expected}$ (assigned function of the users related to a given node) and the values resulting from the calibration $Q_i^{out\ calibrated}$.

These steps are iterated a number of times, as specified in Section 5, and shown in Figure 1. This number of interactions are large enough so that they do not interfere with the results obtained with this analysis, as shown in Section 5.

[1] As known, losses are in the pipes and therefore in the links of the network; however, in the models the discharges required by the users are ascribed to nodes, and so they are the losses.

A number of instruments that collect hydraulic data have to be positioned in the network in order to be used by the calibration module. These instruments are normally pressure transducers, which are positioned in the nodes of the network to record the pressures H_i (heads); and flowmeters, which are positioned at the links of the network and record discharges Q_i.

The requirements and the products of the model are described in Figure 2.

Supply: A set of input discharges $Q^{in} = \{Q_1^{in}, Q_2^{in}, \dots, Q_n^{in}\}$ that are supplied to the distribution network via a number of pipes and represented in the model by their respective links. This flow is directly measured in the real network by field devices. In our case study, there are two pipes supplying the entire flow to the town of Castegnato (as it will be discussed in Section 4). At any time, the network is fed with a given discharge Q^{in}; this discharge is known because it is normally provided by pumps or reservoirs which are continuously monitored. In general, the discharge delivered to the users (inhabitants) $Q^{out} = \{Q_1^{out}, Q_2^{out}, \dots, Q_n^{out}\}$ is inferior to that provided due to leakages. In our case study we have 440 nodes where the demand is set. Unfortunately, so far the knowledge of the delivered discharge is not as precise as that provided to the network, because the meters installed on final users are of inferior quality and their measurement is made monthly or even yearly.

Distribution Network: The isolated network under consideration, for which we wish to establish the location and amount of water losses. The network is modelled as a graph, and therefore by a number of nodes and links which are characterized by a set of heads $H^{comp} = \{H_1^{comp}, H_2^{comp}, \dots, H_n^{comp}\}$, and discharges $Q^{comp} = \{Q_1^{comp}, Q_2^{comp}, \dots, Q_n^{comp}\}$ respectively. The network is subject to a set of losses $Q^{loss} = \{Q_1^{loss}, Q_2^{loss}, \dots, Q_n^{loss}\}$ to be determined. Mathematical models of the networks are nowadays widespread and we used, in this work, the well-known

EPANET Toolkit (Rossman, 2000) to compute the heads H^{comp} and discharges Q^{comp} inside the network. The network is modelled according to network topology, pipe characteristics (diameter, length, material, …) and to each node N_i a demand $Q_i^{out\ expected}$ (discharge that flows out of the network) is assigned based on a function of the number of inhabitants residing in that area. The hydraulic solver calculates the values H^{comp} and Q^{comp} for each node and for each link (respectively) of the distribution network considering the measured values of Q^{in} and the synthetic demands $Q_i^{out\ expected}$. The network represented by the EPANET is an ideal network, without losses. For the case study under consideration, there are 440 nodes connected by 460 links.

The loss for a given node N_i is determined by $Q_i^{loss} = \{Q_i^{out\ expected} - Q_i^{out\ calibrated}\}$.

Demands: A set of output discharges $Q^{out} = \{Q_1^{out}, Q_2^{out}, \dots, Q_n^{out}\}$. This is the demand imposed on the network, i.e. the flow set Q^{out} corresponds to the discharges delivered to the users. These values are simulated (or synthesized) in the model by the evolutionary algorithm (i.e. population). There are currently 440 demands specified for the case under consideration. As in the case supply, these are modelling outgoing pipes and also represented as unidirectional links in the model.

The goal of the calibration is to find a set of discharges $Q_i^{out\ calibrated}$ that minimizes the O.F. function. The evolutionary algorithm produces a combination (i.e. population) of the discharges Q_i^{out} outflowing from the network, and imposed as "demand" in each node. The hydraulic solver calculates the H^{comp} and Q^{comp} values and the calibration module computes the O.F. function. When the O.F. reaches the minimum, the network model is deemed to be "calibrated". A calibrated model represents an ideal water distribution network without losses. One calibration represents one possible hydraulic solution among possible many solutions, where the hydraulic balance requirement is met.

The Objective Function *O.F.* (1) is formulated in order to minimize the differences between the measured and the computed values, i.e. to calibrate the modelled network in order to mirror, as far as possible, the real one. It is composed of the following terms:

(1) the differences among the measured and modelled pressures are minimized $(H^{meas} - H^{comp})$;

(2) the differences among the measured and modelled discharges are minimized $(Q^{meas} - Q^{comp})$;

(3) the differences among the discharge measured and modelled discharges provided to the users are minimized $(Q^{Globally\ Expected} - Q^{Globally\ Computed})$.

Therefore, the following Objective Function (O.F.) is computed (Mambretti and Orsi, 2012; Mambretti et al., 2013):

$$O.F. = \min\left(\sum_{i=1}^{N}\frac{|H^{meas}-H^{comp}|}{|H^{meas}|}\cdot W_H + \sum_{i=1}^{L}\frac{|Q^{meas}-Q^{comp}|}{|Q^{meas}|}\cdot W_Q + \frac{|Q^{Globally\ Expected}-Q^{Globally\ Computed}|}{|Q^{Globally\ Expected}|}\cdot W_{GE}\right) \quad (1)$$

where: *N* is the number of control nodes, *L* is the number of control links, *H* are pressures, *Q* are discharges; the subscript *meas* refers to the measured values, the subscript *comp* to the computed values; $Q^{Globally\ Expected}$ is the discharge introduced into the network, $Q^{Globally\ Computed}$ is the sum of the demands of the network; *W* are the weights, which depend on the expected precision of the real field devices. They are currently set to 1 (theoretical value) but they should be properly adjusted to allow for a more realistic analysis.

The evolutionary algorithm produces combination of the discharges outflowing from the network, and imposed as “demand” in each node, in order to minimize the O.F. When the O.F. reaches the minimum, which theoretically is zero, the network is deemed to be *“calibrated”*, i.e. the calibrated demands are close to the discharges actually outflowing the nodes.

The comparison between the calibrated demands and the discharges that would be expected considering the number of inhabitants of each area allows the identifications of abnormal areas, i.e. the areas that should be further investigated.

Clearly, the presence on an abnormal area in a single instant of a generic day would have minor significance in real scenarios, as it may be due to an oscillation of the requests; However, if the same procedure is applied for a number of days and in different hours of the day, and the same area(s) is (are) identified, the assigned discharge due to the people resident in the area is incorrect and, therefore, non-revenue water is to be expected in that area.

Another challenge to be addressed is the possible existence of more than one minimum, i.e. the instruments are not able to identify a single scenario of non-revenue water. This is a signal of uncertainty and it will be addressed in Section 6.

In this work, however, we subjected the model to five different loss scenarios, where each scenario covers a specific (geographic) area of the network (Section 4). Therefore, the position of the losses is known before the simulations are run. Then, they have been reconstructed applying the proposed methodology in order to check whether it is able to correctly identify the areas where we imposed the losses.

3 Evolutionary Computation

Complex and multi-objective optimization problem are often solved by means of Evolutionary Computation. The term Evolutionary Computation (EC) (Back *et al.*, 1997) represents a large spectrum of heuristic approaches to simulate evolution, including (for example) Genetic Algorithms (GA) (Holland, 1962; Holland, 1975), Simulated Annealing (van Laarhoven and Aarts, 1987), Particle Swarm Optimization (Zhang *et al.*, 2003), and others.

In this work, three approaches have been tested (Simulated Annealing, Genetic Algorithms, Particle Swarm Optimization). Moreover, in this work the results obtained by GA and by PSO have been refined through an Hill Climbing (HC) procedure. All these approaches are described in the following sections.

3.1 Hill Climbing (HC)

The procedure is developed to create small variations at the discharges provided at the nodes in order to define a rather casual new scenario around the existing values; if the new scenario performs better than the existing one, it is kept and used as a new base scenario.

The variations around the average values of the discharges provided at the nodes are computed using a Weibull distribution. In other word, a random number $F \in]0,1[$ is generated and the new discharge Q is computed with the formula:

$$Q = \lambda \cdot [-\ln(1-F)]^{1/k} \quad (2)$$

where λ and k are the parameters of the distribution. The average of the distribution μ is:

$$\mu = \lambda \cdot \Gamma\left(1 + \frac{1}{k}\right) \quad (3)$$

where Γ is the function Gamma. As we wanted to reach the top with slow but safe step, we let:

$$k = 1000 \Rightarrow \Gamma\left(1 + \frac{1}{1000}\right) = 0.999423772 \tag{4}$$

and we computed:

$$\lambda = \frac{\mu(=\mathrm{OldDischarge})}{\Gamma\left(1+\frac{1}{k}\right)} \tag{5}$$

3.2 Simulated Annealing (SA)

Simulated Annealing was originally inspired by formation of crystal in solids during cooling i.e., the physical cooling phenomenon (Kirkpatrick *et al.*, 1983). As discovered a long time ago by iron age blacksmiths, the slower the cooling, the more perfect is the crystal formed. By cooling, complex physical systems naturally converge towards a state of minimal energy. The system moves randomly, but the probability to stay in a particular configuration depends directly on the energy of the system and on its temperature.

The actual application of the procedure is very similar to that developed for HC. However, it admits the possibility of a solution that initially worsens the objective function (O.F.) in order to explore a larger space. This could potentially avoid being trapped in local optima, as it would occur with the simple hill climbing techniques.

Therefore the equation (2) is still applied for the determination of the new scenario, with the parameters computed as described in the paragraph 3.1.

Letting δf the variation of the O.F., the new scenario is kept if:

$\delta f < 0$ i.e. if the O.F. of the new solution is better than the old one;

$\delta f > 0$ (i.e. if the O.F. of the new solution is worse than the old one) the solution is accepted if, obtained a random number $F \in]0,1[$ from a uniform distribution, it results: $F < e^{-\delta f / T}$

where T is the temperature. In other words, there is a possibility that the worse scenario is kept; this possibility is reduced with the number of simulations being tied to the temperature, which is reduced at each iteration with the formula:

$$T_{i+1} = T_i \cdot W = T_0 \cdot W^i \tag{6}$$

where W=0.99995, and i the iteration; having decided to perform one million of runs, a temperature equal to 10 at the first run, at the 1-millionth run is equal to 1.9263E-21. The initial temperature is set in order to allow a given probability p_0 of acceptance of a positive variation of the O.F., with the formula:

$$T_0 = -\delta f^+ / \ln(p_0) \tag{7}$$

Obviously, when $T = 0$ the implemented procedure is a simple hill climbing; increasing the initial temperature and the value of k brings the procedure to have more variability. As mentioned, one million simulations have been performed for each couple of parameters (k, T), as described in the following paragraphs.

3.3 Genetic Algorithms (GA)

The method is a simple genetic algorithm with mutation and crossover operators; this method is based on roulette wheel (Goldberg, 1989). In this paper, this algorithm has been tested using one and two points for crossing over the chromosomes. As known, these algorithms are able to find points close to the best solution, but not the best solution itself; therefore, at the end of the application of the GA a procedure that applies the hill climbing is also used in order to find the best possible solution.

As for the GA, the parameters for running the computer program are:

- Number of individuals per population: 2500
- Number of generations: 100

- Elitism: 20%

The size of the population raises problems. The larger the population is, the easier it is to explore the search space. But it has established that the time required by a GA to converge is O (n log$n$) function evaluations where n is the population size. Goldberg (1989) has shown that GA efficiency to reach global optimum instead of local ones is largely determined by the size of the population. To sum up, a large population is quite useful, but it requires much more computational cost, memory and time. Practically, a population size which individuals are equal to a number of around 5 times the number of parameters is quite frequent, but anyway this size can be changed according to the time and the memory disposed on the machine compared to the quality of the result to be reached (Sivanandam, Deepa, 2008). Note that in this job as there are 440 nodes, there are also 440 parameters to be calibrated; the number of individuals per population is set to be more than 5 times the number of parameters.

3.4 Particle Swarm Optimization (PSO)

With this method, the individuals are assigned a position and a velocity, and they change their position according to the new velocity, computed trying to approach the "best" point (Kennedy and Eberhart, 1995).

The procedure in this case is as follows**:**

First, a new random population is created. These first individuals (scenarios) have velocity v_i equal to zero and the "coordinates" stored in an array x_i. New velocity is computed with a formula we modified with respect to the original (Kennedy and Eberhart, 1995):

$$v_i(t+1) = \alpha \cdot v_i(t) + c_1 \cdot rand \cdot \left[x_{pbest}(t) - x_i(t)\right] + c_2 \cdot rand \cdot \left[x_{gbest}(t) - x_i(t)\right] + c_3 \cdot rand \cdot \left[x_{pbest}(\forall t) - x_i(t)\right] + c_4 \cdot rand \cdot \left[x_{gbest}(\forall t) - x_i(t)\right]$$

(8)

and then the new individual as:

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (9)$$

In the above equations, we have:

α is an "inertia" parameter, which is normally < 1 (it can be >1 but it may produce instabilities) and in the range 0.4-0.9; sometime it changes, starting with 0.9 and reducing to 0.4 during the simulation or something leaving a random component (Eberhart and Shi, 2001). For the first version of the program it is let constant and different values have been tested

c_1, c_2, c_3, c_4 are acceleration parameters normally let equal to 2 (Poli *et al.*, 2002), but they have been changed in this job

rand is a random number $\in [0,1]$

x_{pbest} is the individual with the local best (among neighbours)

x_{pbest} is the individual with the global best

In this phase, individuals are randomly grouped as "neighbours" and their link is permanent. This choice can be obviously considered too simplified; mirroring the real world, early topologies were based on proximity in the search space, and therefore the "neighbour" was not permanent and defined in the Euclidean sense. However, besides being computationally intensive, this kind of communication structure had undesirable convergence properties and therefore was abandoned (Poli *et al.*, 2007). Normally, topologies are static, even if more complex algorithms have been tested; however, as the research has not determined yet the best topology to be adopted (Kennedy and Mendes, 2002), it has to be further investigated.

The same is for the acceleration coefficients *c* which calibration is difficult and deserve a deeper investigation, as will be discussed in Section 5.

The simulations are run with the following parameter values:

- Number of individuals: 2000
- Number of “neighbours”: 100
- Number of iteration: 100

4 Case study

The case study is the water supply network of Castegnato, a small town in the North of Italy with around 7900 inhabitants and with a network divided in two disconnected parts. The characteristics of the town and its water supply networks have been presented by Mambretti and Orsi (2012). For the sake of simplicity, in this paper we only address 440 nodes and 460 links.

As over the years the Board of Water Supply managers recorded more than 50% of water losses, a number of transducers have been installed in the network; their position is shown in Figure 3 and they are detailed in Table 1.

As it can be seen, 16 pressure transducers are positioned at the corresponding nodes and 3 flowmeters are positioned at the corresponding links. The pressures H_i^{in} and the discharges Q_i^{in} at the two pumping stations are measured by four of these transducers.

In order to validate the model and methodology and also understand whether the number and position of transducers are appropriate to locate the leakages, five different loss scenarios have been simulated to check whether they can be reconstructed (i.e. detected or identified) by the algorithms mentioned in Section 3. The different scenarios are reported in Figure 4 and summarized in Table 2.

More specifically, losses are known because we know the pumped discharges $Q^{in} = \{Q_1^{in}, Q_2^{in}, \dots, Q_n^{in}\}$ and the requests $Q^{out} = \{Q_1^{out}, Q_2^{out}, \dots, Q_n^{out}\}$ from the users.

They have been estimated to be equal to 13.41 l/s. Their position is unknown and therefore, to study whether the method works or not, it was decided to set up five test cases (scenarios), distributing those 13.41 l/s all around the network.

For example, in scenario 1 losses Q_i^{loss} were assigned to the north of the town, i.e. 0.18 l/s were added to the demand of 75 nodes for a total of 13.5 l/s. In the scenario 2, losses were assigned to the Eastern area of the town, i.e. 0.27 l/s were added to the demands of 50 nodes in that area, for a total of 13.5 l/s. In scenario 3 losses were supposed to be equally distributed on the whole catchment, and assigned to the nodes where no demand has been applied – therefore 378 nodes out of 440. In scenarios 4 and 5 three areas (clusters) have been selected to disseminate the losses following the discharges shown in Table 2.

The idea in drawing these scenarios is to have a sample of different cases in order to obtain a reasonable certainty that the model is capable of detecting the real scenarios, when the real data is applied.

5 Results

5.1 Results for SA

Results for the SA algorithm are reported in table 3. They are quite inadequate as the final value of O.F. is often higher than the initial one after one million simulations. The reason for the problematic performance of the SA is probably due to the presence of non-physical potential solutions which have been tested: these are given by a distribution of discharges that would lead the network to situations where pumps cannot deliver enough flow or head, or the system has negative pressures. As shown in Figure 5, once the potential solution travels in a field where the solution is not acceptable (grey points, while black points are related to acceptable solutions), it remains trapped for long time before being able to escape, and only then the OF start again to diminish.

5.2 Results for GA

As GAs have random components, simulations have been run 10 times for each scenario and for each method. Results are reported in table 4 (one-point crossover) and 5 (two-point crossover).

Notice that the results can be further improved by repeatedly applying the HC procedure. The HC has been carried out performing one million simulations; however, the results can be further improved: for instance, the value obtained with 2-point crossover GA improved with HC (Scenario 5, simulation 1). The value OF = 0.1548 (table 5) has been tuned up by repeatedly applying the HC procedure to obtain subsequently OF = 0.1309; OF = 0.0988; and OF = 0.0986.

For the purpose of this research, the GA seems to have a better performance as it works in a group of individuals, and therefore if some of them fall in a field where the solution is not allowed, they are simply discarded in the following population without significantly affecting the final results.

So far it has been found that the procedure based on GA, preferably with 2 points crossover followed by a fine tuning with a hill climbing procedure is able to minimize the O.F. It is now to be shown whether the minimization of the O.F. allows the correct reconstruction of the initial scenario. In Figure 6 scenario 1 and 5 are reproduced, together with their best reconstruction.

As it can be seen, the reduction of the O.F. value allows the reconstruction of the correct (i.e. original) loss scenario.

5.3 Results for PSO

In this case, the quality of results is strongly tied to the coefficients values used in the simulation.

In general, it is positive to allow the algorithm a certain amount of freedom to explore. However, too much freedom might lead the system to an unstable state and therefore without a solution.

Tables 6 to 10 illustrate the simulation results for the five scenarios. The solutions provided by PSO are in the best case comparable to the ones achieved by the GA.

In Figure 7, optimizations that generated stable-and-converging and unstable solutions are reported.

6 Multiple optima

So far we have assumed that only one optimum was present. Therefore, the reduction of the O.F. is a practical parameter in deciding whether or not one solution is better than another.

This assumption holds when the number of instruments deployed throughout the network is sufficient to univocally identify the right scenario.

However, it is surely possible that the number of instruments positioned is not sufficient, because of their cost, and anyway in the phase of planning of the number and position of the devices the knowledge of the presence of other minima is desired.

In the former case, the existence of other minima is a measure of the uncertainty of the scenario under analysis; in the latter case instead, it allows a better determination of the number and position of instruments to be installed.

Therefore, GAs seem to be less flexible, although it is still possible to identify the presence of multiple minima dividing the population in “species”. A “niche” can be regarded as one of the holes and a “species” is a collection of population members well suited for a particular niche. It is possible to create stable subpopulations (species) that are well suited to the niches (De Jong, 1975).

However, from this viewpoint the PSO is surely more efficient, as it is possible to calibrate parameter values in order to try to keep the populations at least partially separated.

To achieve this result, we defined a “distance” D (first introduced by Pétrowski, 1996) between the best individual and a different individual that identifies a suboptimal scenario**,** as:

$$D = \sqrt{\sum\left(x_i - x_i^{best}\right)^2} \tag{10}$$

With the instruments actually installed in Castegnato only one minimum exists whereas a number of different maxima can be found as the number of instruments is reduced. For example, with the instruments reported in table 11, we obtained the results showed in table 12.

As it is shown in this table, an optimum O.F. = 0.092 was identified. It is a good solution compared to the values presented in Section 5. However, different solutions with similar O.F. values were found, meaning that there are other holes very close to the best, and therefore there is a large uncertainty in the results.

7 Conclusions

The ability to identify water losses in water distribution networks is crucial in the modern society, usually allowing better resource planning and overall strategic management of these resources. This requirement is exacerbated by the fact that non-revenue water can reach unacceptable levels (e.g. 70%) in some areas.

In this paper, we proposed and employed a methodology to identify the areas where losses are mostly expected. The procedure required data collection (discharge and pressure) from instruments positioned on the water supply network, and successive comparison of the data collected with those simulated by means of a hydraulic network solver program (EPANET).

Three well-known evolutionary algorithms were used to minimize the objective function and then compared, namely Simulated Annealing, Genetic Algorithms and Particle Swarm Optimization.

In order to validate both the approach and the evolutionary algorithms, five test scenarios were developed where a pattern of water loss was established for each of them within the target area (town of Castegnato).

The results show that the scenarios were properly reconstructed, even if errors are obviously present; however, the goal to identify the areas where losses are concentrated seems to be reached.

GA and PSO have shown a better performance among the selected methods. This is probably due to the fact that they work with groups of individuals, while SA operates with only one individual at a time and it might become trapped in a field of unfeasible solutions.

Moreover, occasionally there is the need to have a method able to identify whether more than one optimum is present, which would support the decision about the minimum number of devices required to be installed and it would also provide a measure of the uncertainties in the identification of the area where losses are expected. To this end, PSO seems to be the preferred algorithm, even considering the fact that the calibration of its parameters is more complex (thus requiring a deeper investigation). So far, the application of this algorithm has not been trivial and therefore it should be performed by a skilled technician.

However, we believe that the precision of the results is important and we are not worried if the procedure is difficult or unstable and requires expertise and time to be applied, as the costs of a few hours of computational work are not comparable with those resulting from an incorrect analysis. To this aim, this methodology could also indicate the need of collecting more data, for instance by means of a portable flowmeter.

Although the case study presented in the paper is a quite small town in the north of Italy, we believe that the method can be applied to any other town, as it is able to identify whether the number of instruments is appropriate or not, or to determine the suitability of a proposed districtualization.

Future developments will also comprise the analysis of real data collected on the network, which will induce even more uncertainties in the evaluations. In fact, in this paper we assumed (without sacrificing the quality of the results) that the “measured” values were not affected by errors, which is obviously not true in the real world; therefore, the need of different weights could emerge and the possibility to insert them in the calibration phase should be investigated. Moreover, the network is not always perfectly known: in the worst case, even the topology or the pipe diameters are incorrectly modelled; often, the roughness of the pipes is considered, again, as a calibration parameter. However, the procedure developed and presented in the paper, implying continuous monitoring of the network, produce an increasing knowledge of the network. In fact, at the time of implementation of the procedure, it is to be expected that the network model would have many errors; however, network losses would be also quite high, thus facilitating their determination. Due to the rehabilitation work, not only the losses would be reduced but also the errors in the model would be corrected, starting in this way a virtuous cycle in management of water distribution networks.

The final goal of this research is the development of a new methodology that is able not only to locate areas where losses are mostly expected, but also to help in making decisions related to the different possibilities of rehabilitation, for instance: whether the better option is either to repair or to substitute the pipe; the expectation of duration of the pipe as a function of the laying depth of the pipe, the type of traffic on the road, the material of the pipe, the pressure cycle and other parameters which are now the subject of current research.

References

Aksela K., Aksela M., Vahala R. (2009), "Leakage detection in a real distribution network using a SOM", *Urban Water Journal*, 6:4, 279-289

Almandoz J.; Cabrera, E.; Arregui, F.; Cabrera Jr., E., Cobacho R. (2005). "Leakage Assessment Through Water Distribution Network Simulation." *ASCE J. of Water Resour. Plan. Manage.* 131(6), pp.458-466

Alvisi S., Grata S., Franchini M. (2006) "Leakage detection planning in water distribution systems", in "Management of Water Networks" Proceedings of the Conference "Efficient Management of Water Networks. Design and Rehabilitaion Techniques." Bertola and Franchini (Eds), Ferrara 2006, De Angeli Editore, Milano, Italy

Back T., Fogel D., Michalewicz Z. (1997) *Handbook of evolutionary computation* IOP Publishing Ltd. and Oxford University Press, New York and Oxford

De Jong, K. A. (1975) *An analysis of the behaviour of a class of genetic adaptive systems.* Doctoral dissertation, University of Michigan. Dissertation Abstracts International 36(0), 5140B. (University Microfilms No. 76–9381).

Eberhart, R. C., Shi, Y. (2001) "Tracking and optimizing dynamic systems with particle swarms." In *Proceedings of the IEEE congress on evolutionary computation (CEC)* (pp. 94–100), Seoul, Korea. Piscataway: IEEE.

Engelhardt M.O., Skipworth P.J., Savic D.A., Saul A.J., Walters G.A. (2000) "Rehabilitation strategies for water distribution networks: a literature review with a UK perspective", *Urban Water*, 2, pp. 153-170

Giustolisi, O., Laucelli, D., Savic, D.A., (2006) "Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment", *J. of Civil Engineering and Environmental Systems*, Taylor & Francis, UK, No.3, Vo.23, pp. 175-190.

Goldberg D.E., 1989, *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley, Reading.

Halhal, D., Walters, G. A., Savic, D. A., Ouazar, D. (1999) "Scheduling of water distribution system rehabilitation using structured messy genetic algorithms" *Evolutionary Computation*, Volume 7, No. 3; MIT Press

Holland J. H. (1962) "Outline for a logical theory of adaptive systems" *Journal of the ACM*, Volume 9, Issue 3; ACM.

Holland J. H. (1975) "Adaptation in natural and artificial systems" University of Michigan Press.

Islam M.S., Sadiq R., Rodriguez M.J., Francisque A., Najjaran H., Hoorfar M. (2011), "Leakage detection and location in water distribution systems using a fuzzy-based methodology", *Urban Water Journal*, 8:6, 351-365

Kennedy J., Eberhart R., "Particle Swarm Optimization", From *Proc. IEEE Int'l. Conf. on Neural Networks* (Perth,Australia), IEEE Service Center, Piscataway, NJ, IV:1942–1948, 1995.

Kennedy, J., Mendes, R. (2002) "Population structure and particle swarm performance." In *Proceedings of the IEEE congress on evolutionary computation (CEC)* (pp. 1671–1676), Honolulu, HI. Piscataway: IEEE.

Kirkpatrick S., Gelatt C. D., Vecchi M. P. (1983). "Optimization by Simulated Annealing". *Science* 220 (4598): 671–680.

Male, J. W., Walski, T. M., and Slutsky, A. H. (1990) "Analyzing water main replacement policies." *J. Water Resour. Plan. Manage.*, 116(3), 362–374.

Mambretti S., Orsi E. *Genetic Algorithms for Leak Detection in Water Supply Networks.* 1st International Conference on Urban Water, 25-27 April, 2012, New Forest, UK

Mambretti S., Martins P.S., Moraes R.L. "Evolutionary Computation Techniques to Assess Losses in Water Supply Networks" 7th International Conference on Sustainable Water Resources Management, 21 – 23 May, 2013, New Forest, UK

Pétrowski A. (1996) “A clearing procedure as a niching method for genetic algorithms, Proceedings of *IEEE International Conference on Evolutionary Computation* pp. 798-803.

Poli R., Kennedy J., Blackwell T. (2007) “Particle swarm optimization. An overview.” *Swarm Intell* 1: pp. 33–57

Puust R., Kapelan Z., Savic D.A., Koppel T. (2010), “A review of methods for leakage management in pipe networks”, *Urban Water Journal*, 7:1, 25-45

Rossman, L.A., 2000. *EPANET2 and programmer’s toolkits.* Cincinnati, OH: Risk Reduction Engineering Laboratory, U.S. Environmental Protection Agency.

Sivanandam S.N., Deepa S.N. *Introduction to Genetic Algorithms* Springer-Verlag Berlin Heidelberg 2008, 442 pp.

Sundahl A., (1996) “Using break data on water pipe systems for renewal planning” COST Action C3 workshop, 18 and 19 June 1996, Brussels.

van Laarhoven P., Aarts E. (1987) *Simulated Annealing: Theory and Applications.* Springer.

Walski T.M., Bezts W., Posluszny E.T., Weir M., Whitman B.E. (2006) “Modelling leakage reduction through pressure control” *Journal of AWWA*, 98:4, pp. 147-155

Wu Z.Y, Sage P. (2006) “Water loss detection via genetic algorithm optimization-based model calibration” ASCE 8th Annual International Symposium on Water Distribution System Ananlysis, Cincinnati, Ohio, August 27-30

Wu, Z. Y, Walski, T., Mankowski, R., Cook, J. Tryby, M. and Herrin G. (2002) “Calibrating Water Distribution Model Via Genetic Algorithms”, in *Proceedings of the AWWA IMTech Conference*, April 16-19, Kansas City, MI.

Zhang L., Zhou C., Liu X., Z M., Ma M., Liang Y. (2003) “Solving multi objective optimization problems using particle swarm optimization.” *Proceedings of IEEE Congress on Evolutionary Computation 2003* (CEC 2003), Canbella, Australia, pp. 2400–2405.

Tables

Table 1: ID of the node or link where the device is positioned and type (pressure transducers *P* are positioned on nodes; flowmeters *Q* on links).

ID	Type	Note
45	*P*	
150	*P*	
144	*P*	
191002	*P*	
41	*P*	
224	*P*	
105	*P*	
4121002	*P*	
273	*P*	
690	*P*	
2431004	*P*	
237	*P*	
363	*P*	
300	*P*	
250	*Q*	
69	*Q*	
71	*P*	
66	*Q*	
381004	*P*	
185	*Q*	These measures are carried out in correspondence of the pump stations.
167	*Q*	
177100	*P*	
680100	*P*	

Table 2 – Scenarios of water loss

N	Number of Nodes	Area of coverage	Discharge (loss) Q_i^{loss} per node (l/s)	Total Discharge (loss)
1	75	North	0.180	13.52
2	50	East	0.270	
3	378	All	0.036	
4	70	clusters	0.193	
5	40	clusters	0.338	

Table 3: Values O.F. after using Simulated Annealing with different parameters. The O.F. with the initial configuration is 5.3926. Note that when T=0 the SA reduces to HC.

O.F. Weibull Shape Parameter (k)	Temperature (T)				
	0	0.5	1	5	10
1	2.6027	4.6943	5.5604	6.9473	13.1846
10	0.3397	6.9916	7.1754	11.1221	8.2630
100	0.4848	15.1249	15.0526	14.2636	14.7489
1000	0.9979	17.2065	17.4476	17.5842	17.9049

Table 4: O.F. reached by one-point crossover GA and hill climbing.

	Sc. 1 GA	Sc. 1 HC	Sc. 2 GA	Sc. 2 HC	Sc. 3 GA	Sc. 3 HC	Sc. 4 GA	Sc. 4 HC	Sc. 5 GA	Sc. 5 HC
1	0.3965	0.1053	0.3856	0.1095	0.3848	0.1110	0.3245	0.2920	0.4462	0.2170
2	0.1619	***0.0916***	0.4252	0.1151	0.3750	0.1155	0.4020	0.2723	0.4465	0.2081
3	0.1864	0.1045	0.4139	0.1153	***0.1801***	0.1006	0.4322	0.3001	0.3561	0.1619
4	0.1633	0.0962	***0.3558***	0.1088	0.2073	0.1074	0.3304	0.2970	0.4331	0.2278
5	0.2853	0.1078	0.4839	0.1190	0.4144	0.1026	0.4319	0.2984	0.4221	0.2105
6	0.1956	0.0949	0.5118	0.1122	0.3314	0.1072	0.4135	0.2827	0.3366	0.2098
7	0.1806	0.0926	0.5251	0.1170	0.4181	***0.0947***	0.4319	0.2935	0.4461	0.2046
8	0.1876	0.0932	0.4293	***0.1084***	0.3341	0.1031	0.3211	0.2882	0.3269	0.1918
9	0.1542	0.0945	0.4056	0.1088	0.3258	0.1144	***0.3049***	***0.2482***	***0.2811***	***0.1523***
10	***0.1421***	0.1022	0.4440	0.1121	0.5384	0.1044	0.4309	0.2913	0.4466	0.2187

Table 5: O.F. reached by GA 2 point crossover and hill climbing.

	Sc. 1 GA	Sc. 1 HC	Sc. 2 GA	Sc. 2 HC	Sc. 3 GA	Sc. 3 HC	Sc. 4 GA	Sc. 4 HC	Sc. 5 GA	Sc. 5 HC
1	0.1690	0.1052	0.2749	0.1052	0.1387	0.0978	***0.2840***	0.2501	***0.1923***	***0.1548***
2	0.1578	0.1020	0.2958	0.1045	0.1433	0.1047	0.2953	0.2632	0.2897	0.2012
3	0.1574	0.1024	0.3532	0.1052	0.1791	0.1031	0.3767	***0.2463***	0.2713	0.2069
4	***0.1469***	0.0953	***0.1964***	0.1029	0.1250	0.1107	0.3889	0.2738	0.4133	0.2286
5	0.1520	***0.0913***	0.3825	0.1103	0.1261	0.1037	0.3160	0.2868	0.2724	0.1933
6	0.1637	0.0994	0.3283	0.1056	***0.1179***	***0.0970***	0.3082	0.2771	0.2780	0.1743
7	0.2616	0.0912	0.2912	***0.1005***	0.1363	0.1107	0.2991	0.2708	0.4386	0.2160
8	0.1581	0.0965	0.3331	0.1089	0.1658	0.1218	0.3200	0.2910	***0.1923***	0.1654
9	0.1701	0.1066	0.2345	0.1086	0.1532	0.1063	0.3079	0.2738	0.4140	0.2223
10	0.1473	0.0963	0.2774	0.1030	0.1708	0.1015	0.3191	0.2842	0.3029	0.1904

Table 6: O.F. reached by PSO for Scenario 1.

Initial	OF =	5.3926					
α	0.2	0.4	0.4	0.4	0.5	0.3	0.3
C_1	0.5	1	0	0	1	1	1.33
C_2	0.5	1	0	2	1	1	1.33
C_3	0.5	1	2	0	1	1	0.67
C_4	0.5	1	2	2	1	1	0.67
O.F.	0.8736	0.1700	0.1317	0.1344	0.3702	***0.1094***	0.1622
	0.8201	0.1906	0.1199	0.1739	0.1372	0.1491	0.1453
	0.7796	0.1416	0.1197	0.1934	0.2067	0.1270	0.1232
	0.7075	0.1242	0.2167	0.3189	0.2182	0.1402	0.1509
	0.8130	0.2762	0.1583	0.2904	0.1652	0.1235	0.1439

Table 7: O.F. reached by PSO for Scenario 2.

Initial	OF =	4.9520					
α	0.2	0.4	0.4	0.4	0.5	0.3	0.3
C_1	0.5	1	0	0	1	1	1.33
C_2	0.5	1	0	2	1	1	1.33
C_3	0.5	1	2	0	1	1	0.67
C_4	0.5	1	2	2	1	1	0.67
O.F.	1.2187	0.2438	0.1667	0.1792	0.2492	***0.1108***	0.1455
	1.1101	0.1182	0.1568	0.2326	0.2158	0.1456	0.1268
	1.1982	0.2480	0.1549	0.2364	0.2432	0.1431	0.1420
	1.0873	0.1532	0.1577	0.2575	0.2449	0.1277	0.1157
	1.0574	0.2183	0.1463	0.2549	0.3199	0.1284	0.1788

Table 8: O.F. reached by PSO for Scenario 3.

Initial	OF =	6.5345					
α	0.2	0.4	0.4	0.4	0.5	0.3	0.3
C_1	0.5	1	0	0	1	1	1.33
C_2	0.5	1	0	2	1	1	1.33
C_3	0.5	1	2	0	1	1	0.67
C_4	0.5	1	2	2	1	1	0.67
O.F.	0.6966	0.1199	0.1277	0.1993	0.1717	***0.0992***	0.1167
	0.7336	0.1378	0.1524	0.1774	0.1944	0.1192	0.1171
	0.6825	0.1229	0.1232	0.1391	0.1357	0.1191	0.1236
	0.7432	0.1062	0.1371	0.1583	0.1403	0.1101	0.1263
	0.6885	0.1096	0.1307	0.1491	0.1407	0.1209	0.1067

Table 9: O.F. reached by PSO for Scenario 4.

Initial	OF =	6.1170					
α	0.2	0.4	0.4	0.4	0.5	0.3	0.3
C_1	0.5	1	0	0	1	1	1.33
C_2	0.5	1	0	2	1	1	1.33
C_3	0.5	1	2	0	1	1	0.67
C_4	0.5	1	2	2	1	1	0.67
O.F.	1.1012	0.1957	0.1426	0.2267	0.2187	0.1503	0.1184
	1.0382	0.1692	0.1585	0.2114	0.3652	0.1357	0.1982
	1.0968	0.1687	0.1810	0.2138	0.2205	0.1281	0.1189
	1.0669	0.1365	0.2023	0.2165	0.3832	0.1373	0.1669
	1.0168	0.2087	0.1808	0.2741	0.3977	***0.1096***	0.1513

Table 10: O.F. reached by PSO for Scenario 5.

Initial	OF =	6.3482					
α	0.2	0.4	0.4	0.4	0.5	0.3	0.3
C_1	0.5	1	0	0	1	1	1.33
C_2	0.5	1	0	2	1	1	1.33
C_3	0.5	1	2	0	1	1	0.67
C_4	0.5	1	2	2	1	1	0.67
O.F.	1.3572	0.1415	0.1815	0.2072	0.2328	0.1340	0.1940
	1.4369	0.1789	0.1747	0.2625	0.2592	0.1788	0.1324
	1.4287	0.1700	0.1809	0.2551	0.1629	0.1294	0.1293
	1.3921	0.1847	***0.1262***	0.2081	0.2112	0.1286	0.1950
	1.2872	0.2592	0.1714	0.2494	0.3121	0.1645	0.1538

Table 11: Instruments left for the test with reduced devices

45	*P*	30.04
150	*P*	37.38
144	*P*	39.77
191002	*P*	35.26
41	*P*	35.58
224	*P*	38.71
105	*P*	36.06
4121002	*P*	19.28

Table 12: Results for the optimization with reduced instruments.

O.F.	Distance from the best individual	O.F.	Distance from the best individual
0.338	0.556	***0.094***	***0.570***
0.314	0.543	0.102	0.590
0.314	0.671	***0.094***	***0.581***
0.182	0.745	0.301	0.558
0.146	0.616	0.107	0.579
0.146	0.616	***0.092***	***0.000***
0.095	0.576	0.105	0.599
0.301	0.558	***0.094***	***0.587***
0.093	***0.589***	***0.094***	***0.592***
0.094	***0.624***	0.242	0.564
0.106	0.495	---	---

Figure Captions:

Figure 1. Activity diagram of the method (approach) employed.

Figure 2: Computational model.

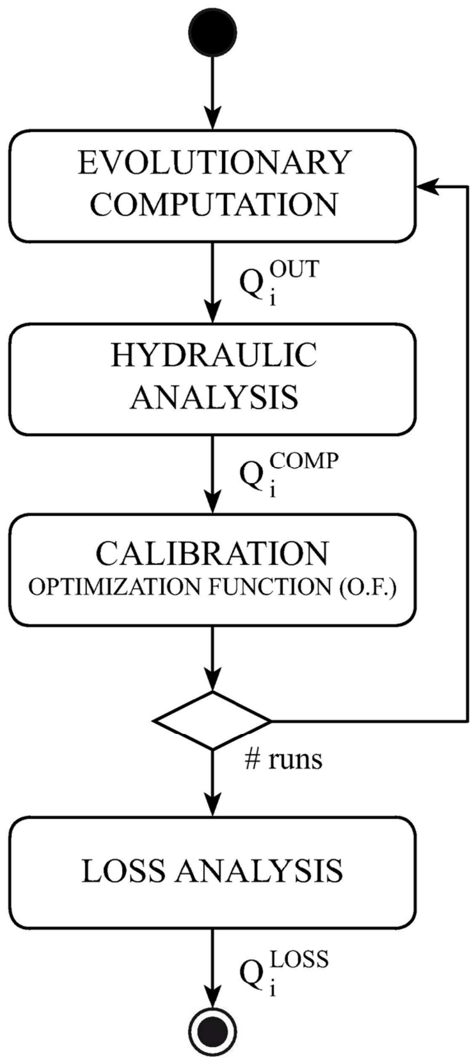
Figure 3: Position of the devices, and their ID

Figure 4: Position of the losses in the different scenarios

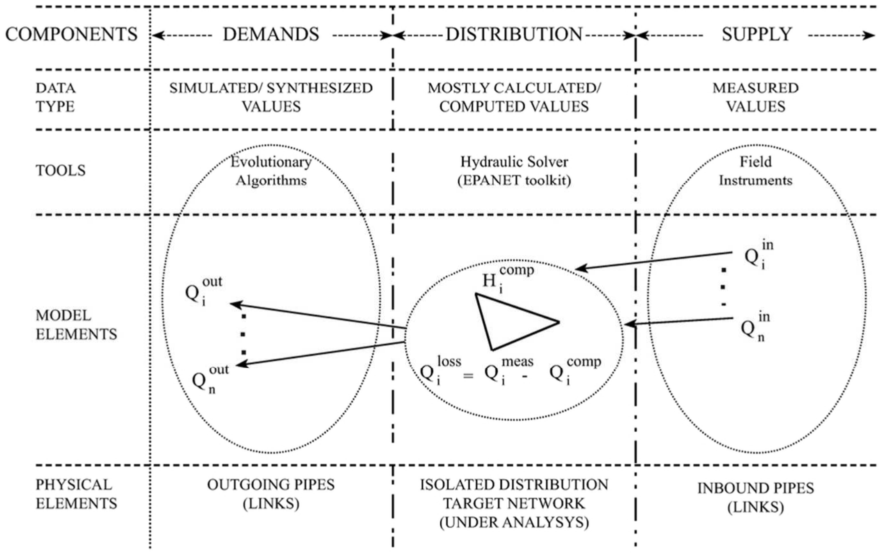
Figure 5: Path of the potential solution using the SA method: in grey the unacceptable solutions are reported, in black those acceptable. As can be seen, the initial OF is much lower. Simulations performed with k=10 and T=5. Above: results of 1 million simulations. Below: trend of the first 100 simulations.

Figure 6: Losses in Castegnato according to theoretical scenarios: (A): Scenario 1, (B) Scenario 5. The best reconstruction of Scenario 1 (C) and Scenario 5 (D), performed minimizing the O.F. with two-point crossover GA followed by HC.

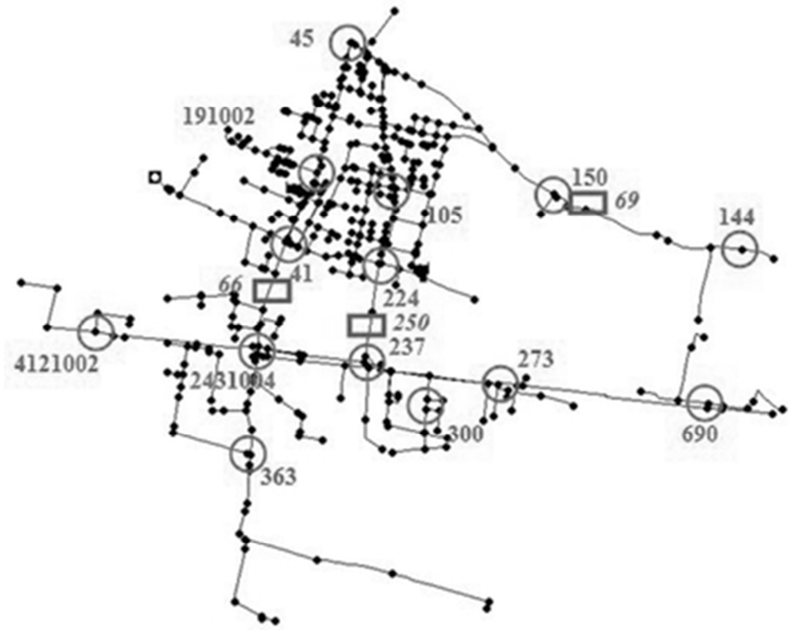
Figure 7: O.F. trend as a function of PSO parameters



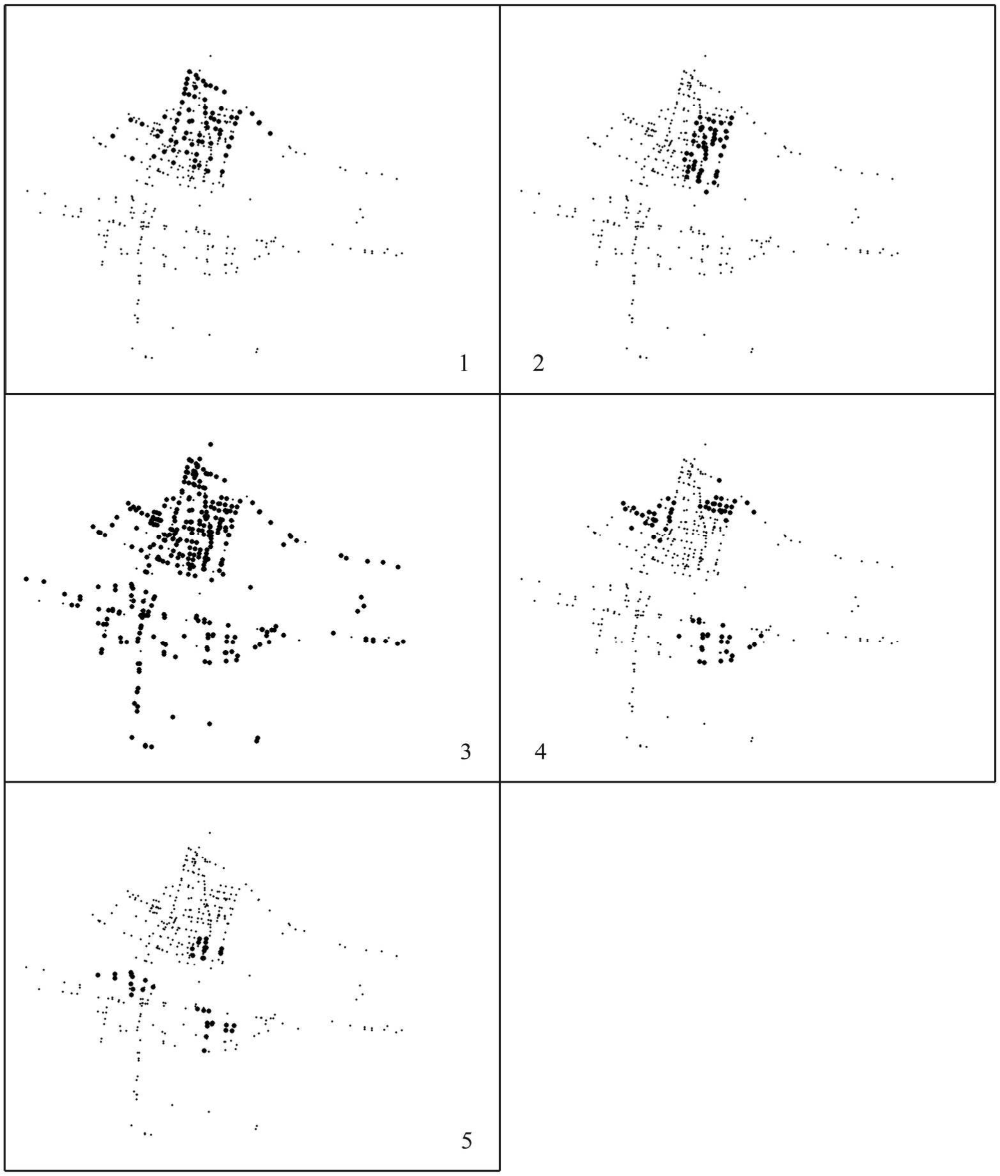
119x180mm (300 x 300 DPI)



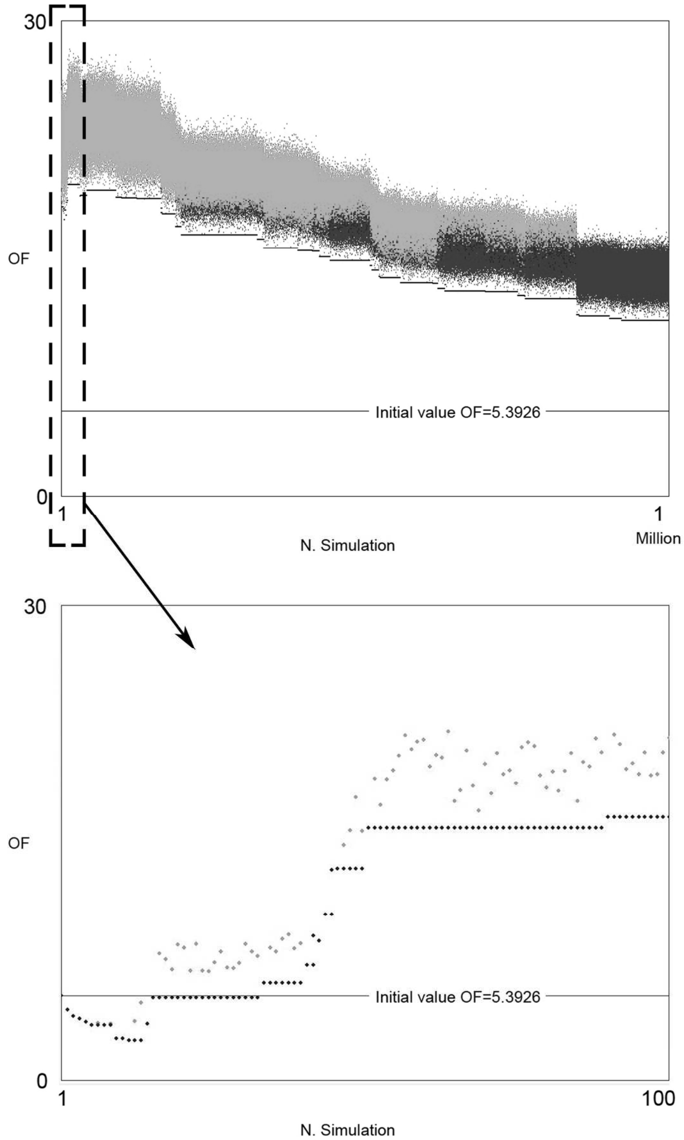
99x55mm (300 x 300 DPI)



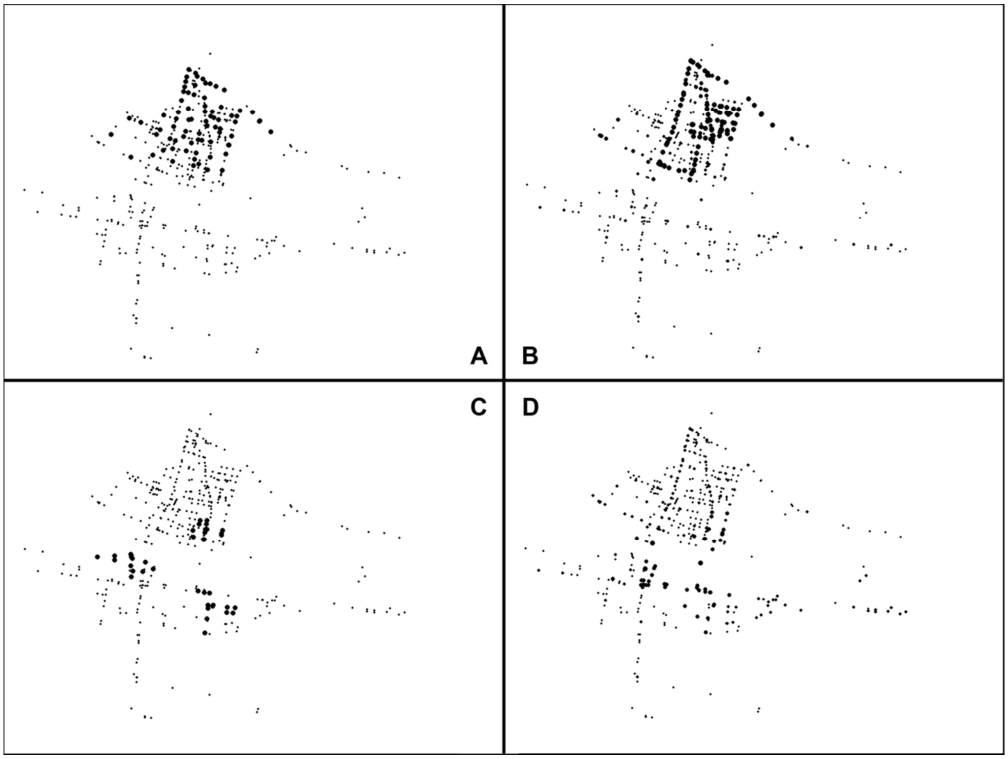
36x27mm (300 x 300 DPI)



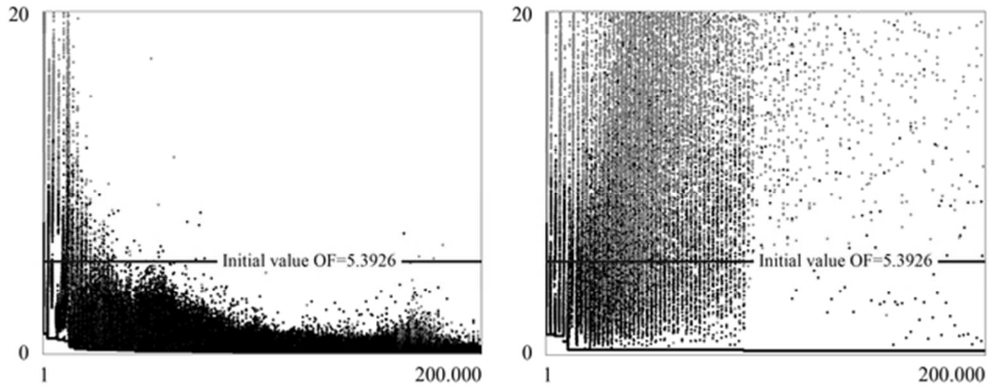
237x279mm (300 x 300 DPI)



160x256mm (300 x 300 DPI)



119x89mm (300 x 300 DPI)



54x21mm (300 x 300 DPI)