Evolutionary computation techniques to assess losses in water supply networks

S. Mambretti, P. S. Martins & R. L. Moraes
School of Technology, University of Campinas, UNICAMP, Brazil

Abstract

In this work, a methodology that combines both simulation and measurements of pressures and discharges on the water supply network is applied to a case study. The demands of the model nodes are systematically changed by means of two evolutionary algorithms and the network is simulated in order to match the readings of the instruments. By comparing the measurements with the simulated values it is possible to assess the losses and estimate their locations.

Specifically, two methods have been tested and applied to a case study: the former based on Simulated Annealing (SA) and the latter on Genetic Algorithms (GAs). The simulations show that the methods based on GAs perform much better and are able to detect the different hypothesized scenarios, while the single individual used by SA risks being trapped in an unfeasible zone in its search. Moreover, the solution obtained by GAs can be further improved by means of a simple Hill Climbing procedure (HC), thus achieving satisfactory results.
Keywords: water supply networks, water losses, Genetic Algorithms, hydraulic modeling.

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) [1], 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 [2, 3]. Common objectives functions are the not-delivered water volumes or the number of customers affected by interruptions caused by pipe bursts [4].

Such condition led to the development of models that are either able to generate pipe breaks [1] or have available good databases about previous breakages [5, 6]. 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.

A new methodology has already been presented that identifies the areas where losses are mostly expected [7]. 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. 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. The optimization was based on a classic Genetic Algorithm.

This paper focus on different methods of Evolutionary Computation in order to establish the best procedure to minimize the Objective Function. Moreover, we assessed whether or not the value of the Objective Function is a reliable indicator of the goodness of the presented solution.

The remainder of this paper is organized as follows: Section 2 discusses the optimization methods used in this work. Section 3 introduces the case study (city of Castegnato). Finally, in Section 4 we present our remarks and conclusions.

2 Evolutionary computation

Complex and multi-objective optimization problems are often solved by means of Evolutionary Computation. The term Evolutionary Computation (EC) [8] represents a large spectrum of heuristic approaches to simulate evolution, including (for example) Genetic algorithms (GAs) [9, 10], Simulated Annealing (SA) [11], Particle Swarm Optimization [12], and Ant Colony Optimization [13]. In this work, two approaches have been tested (Simulated Annealing and Genetic Algorithms), as described in the following sections.

2.1 Simulated annealing (SA)

The first method used in this work is based on Simulated Annealing, where we admit the possibility, at the beginning of the optimization, of a solution that worsens the objective function (O.F.). This is in order to explore a larger space and to arrive to the best solution, while also avoiding being trapped in the local optima

as would happen with the simple Hill-Climbing techniques. For the application of SA, a random number $F \in]0,1[$ is selected and the new discharge Q is calculated by equation (1):

$$Q = \lambda \cdot [-\ln(1-F)]^{\frac{1}{k}} \qquad (1)$$

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

$$\mu = \lambda \cdot \Gamma\left(1 + \frac{1}{k}\right) \qquad (2)$$

The initial temperature $T_0 = -\frac{\delta f^+}{\ln(p_0)}$ is computed setting the initial value of the probability of acceptance of a positive scenario p_0. During the optimization, the temperature decreases following the law:

$$T_{i+1} = T_i \cdot W = T_0 \cdot W^i \qquad (3)$$

where $W = 0.99995$, having decided to perform one million simulation runs. The results of the optimization are reported in Table 4. Notice that $T = 0$ means that the implemented procedure is a Hill Climbing; increasing the initial temperature and the value of k causes the procedure to have more variability. One million simulations have been performed for each couple of parameters (k, T).

2.2 Genetic algorithm

The second optimization method used in this work is a Genetic Algorithm with mutation and crossover operators and roulette wheel selection [14].

This algorithm has been tested using one and two points for crossing over the chromosomes. As it is 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 GAs, a procedure that applies the so-called Hill Climbing is also used in order to find the best possible solution.

As for the GAs, the parameters used in the algorithm are:
- Number of individuals per population: 2500
- Number of generations: 100
- Elitism: 20%

Notice that as there are 440 nodes in the network, there is also the same number of parameters to be calibrated; the number or individuals of the population is taken as more than 5 times the number of parameters.

3 Case study

The case study considered 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 [7]. As over the years the board of water supply managers recorded more than 50% of water losses, a number

of transducers have been installed throughout the network. Fig. 1 illustrates the transducers which are uniquely identified by an integer ID and represented by a circle. The dots represent nodes and the edges the pipes. These elements of the graph are further detailed in Table 1.
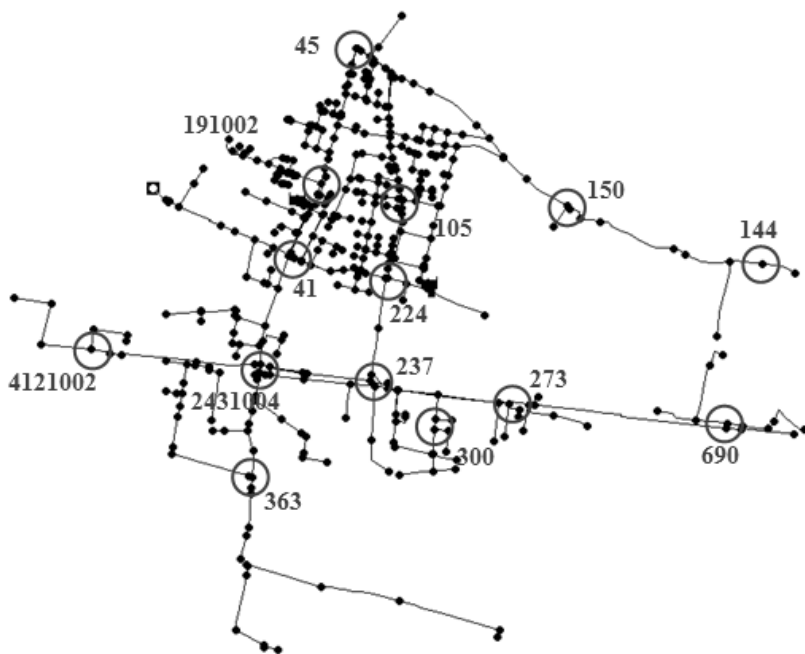

Figure 1: ID and position of measurement devices (Castegnato).

In order to understand whether the number and position of devices are appropriate or not to locate the leakages, five scenarios have been simulated. These scenarios impose losses in the different areas of the town. They allow us to verify whether or not they can be reconstructed by the algorithms mentioned in Section 2. The O.F to be minimized is given by equation 4:

$$
\text{O.F} = \min \left(\sum_{i=1}^{\text{\# control nodes}} \frac{|h_{\text{meas}} - h_{\text{comp}}|}{|h_{\text{meas}}|} \cdot W_H \right.
$$

$$
+ \sum_{i=1}^{\text{\# control links}} \frac{|q_{\text{meas}} - q_{\text{comp}}|}{|q_{\text{meas}}|} \cdot W_Q +
$$

$$
\left. \frac{|q_{\text{globallyExpected}} - q_{\text{globallyComputed}}|}{|q_{\text{globallyExpected}}|} \cdot W_{GE} \right) \quad (4)
$$

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

N	N/L	Type	ID	N	N/L	Type	ID
01	N	P	45	13	N	P	363
02	N	P	150	14	N	P	300
03	N	P	144	15	L	Q	250
04	N	P	191002	16	L	Q	69
05	N	P	41	17	N	P	71
06	N	P	224	18	L	Q	66
07	N	P	105	19	N	P	381004
08	N	P	4121002	20	L	Q	185
09	N	P	8	21	L	Q	167
10	N	P	690	22	N	P	177100
11	N	P	2431004	23	N	P	680100
12	N	P	237	24	–	–	–

where W are weights, which are all set to 1 for the theoretical scenarios. Clearly, their actual value ultimately depends on the expected precision of the real devices.

4 Results

Results for SA are reported in Table 2. As it can be seen, they are quite disappointing, as the final value of the O.F. is often larger than the initial one after one million simulations. The reason of the problematic performance of the SA is probably given by the presence of non-physical potential solutions which have been tested: these are given by a distribution of discharges that would induce situations in the network where pumps either cannot deliver enough flow or head, or the system has negative pressures. As can be seen in Fig. 3, once the potential solution travels in a field where the solution is not acceptable, it remains trapped for long time before being able to escape; only then the O.F. start again to diminish (in gray we report the unacceptable solutions and in black those that are deemed acceptable). As can be seen, the initial O.F is much lower. The simulations reported in Fig. 3 were performed with $k = 10$ and $T = 5$.

The results for GAs are reported in Table 3 (one point crossover) and 4 (two-points crossover). As GAs have random components, simulations have been run 10 times for each scenario and for each method. It is to be noted that the results can be further improved by repeatedly using the Hill Climbing procedure. In the mentioned tables, the HC has been carried out performing one million simulations;

Table 2: O.F. values after using Simulated Annealing with different parameters. The O.F. with the initial configuration is 5.3926.

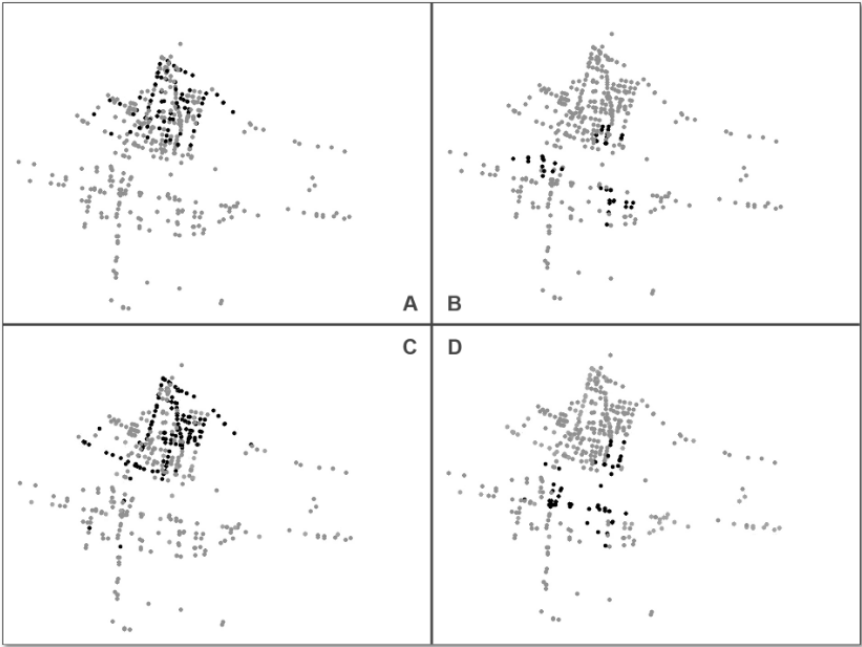| N	$T = 0$	$T = 0.5$	$T = 1$	$T = 5$	$T = 10$
1	2.6027	4.6943	5.5604	6.9473	13.1846
10	0.3397	6.9916	7.1754	11.1221	8.2630
100	0.4848	15.1249	5.0526	14.2636	14.7489
1000	0.9979	17.2065	17.4476	17.5842	17.9049

Figure 2: Losses in Castegnato according to theoretical scenarios.

however, the results can be further improved. For example, the value obtained with the 2-point-crossover GA improved with HC (Table 4, scenario 5 and simulation 1) with O.F = 0.1548. Hill Climbing was further reapplied several times, each time for one million simulation runs, obtaining subsequently O.F = 0.1309; O.F = 0.0988; and O.F = 0.0986.

For the purpose of this research, the GAs seem to have a better performance as they work in a group of individuals, and therefore if some of them fall in a field where the solution is not allowed, it is simply discarded in the following population without affecting too much the final results.
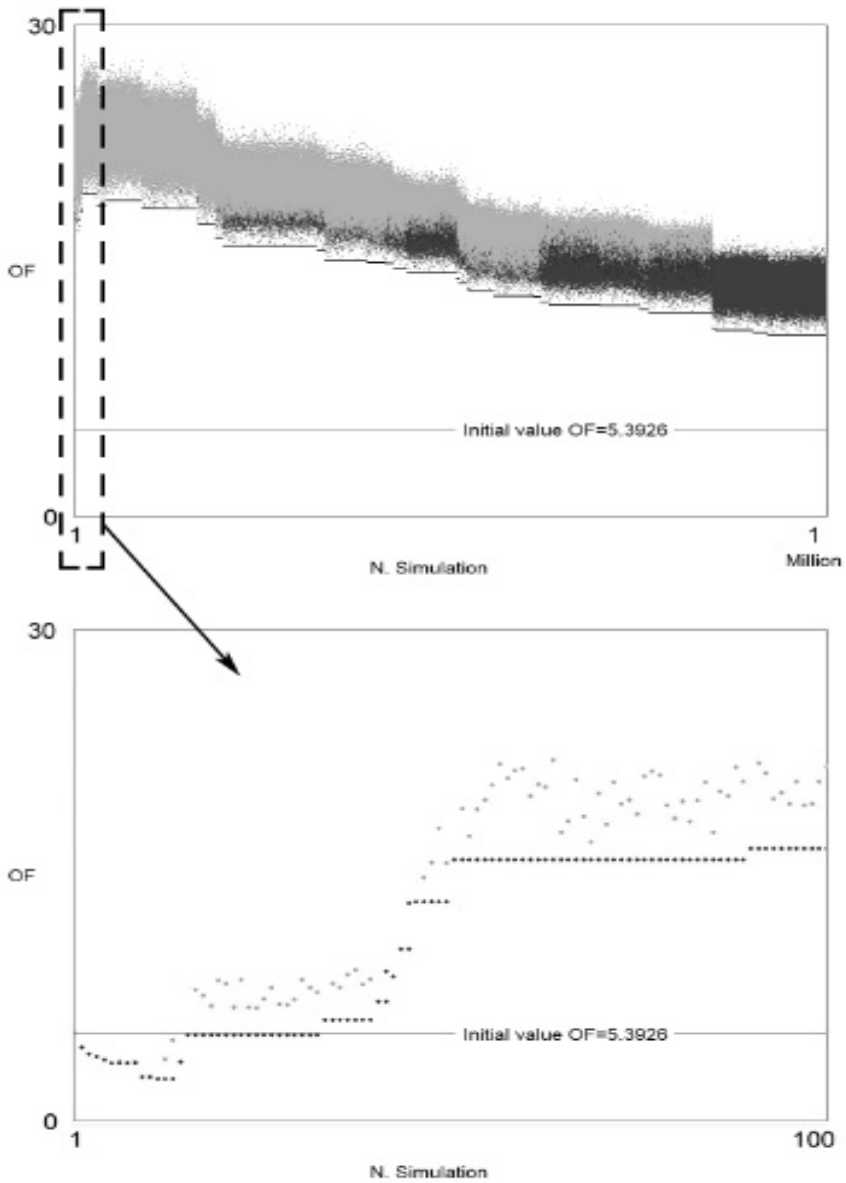
Figure 3: Path of the potential solution using SA method: results of 1 million simulations.

Table 3: O.F. reached by GA1 point crossover and Hill Climbing.

·	scenario 1		scenario 2		scenario 3		scenario 4		scenario 5	
N	GA	HC	GA	HC	GA	HC	GA	HC	GA	HC
01	0.3965	0.1053	0.3856	0.1095	0.3848	0.1110	0.3245	0.2920	0.4462	0.2170
02	0.1619	0.0916	0.4252	0.1151	0.3750	0.1155	0.4020	0.2723	0.4465	0.2081
03	0.1864	0.1045	0.4139	0.1153	0.1801	0.1006	0.4322	0.3001	0.3561	0.1619
04	0.1633	0.0962	0.3558	0.1088	0.2073	0.1074	0.3304	0.2970	0.4331	0.2278
05	0.2853	0.1078	0.4839	0.1190	0.4144	0.1026	0.4319	0.2984	0.4221	0.2105
06	0.1956	0.0949	0.5118	0.1122	0.3314	0.1072	0.4135	0.2827	0.3366	0.2098
07	0.1806	0.0926	0.5251	0.1170	0.4181	0.0947	0.4319	0.2935	0.4461	0.2046
08	0.1876	0.0932	0.4293	0.1084	0.3341	0.1031	0.3211	0.2882	0.3269	0.1918
09	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 4: O.F. reached by GA2 point crossover and Hill Climbing.

·	scenario 1		scenario 2		scenario 3		scenario 4		scenario 5	
N	GA	HC	GA	HC	GA	HC	GA	HC	GA	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

5 Remarks and conclusion

Water distribution networks are an essential component of water supply systems and represent a critical infrastructure asset to the society. As such, they require effective and efficient energy saving management. In some countries Unaccounted for Water or Non-Revenue Water (NRW) is up to 70% of the water volume supplied, which is a serious economic damage.

In this paper, a methodology to identify the areas where losses are mostly expected was verified; the procedure requires data collection (discharge and pressure) from instruments positioned on the water supply network, and successive comparison of the data collected with those simulated.

Different methods to minimize the O.F. have been tested: those based on Genetic Algorithms and those based on Simulated Annealing. In order to test the methodologies, a case study has been identified, where devices are to be installed and data collected. The position of the devices and the validity of the developed methodology were checked by means of a number of scenarios before their installation.

It is found that the areas where the largest losses are imposed are properly identified; this allowed us to confirm that the positions of the devices placed within the network are correct.

So far it has been found that the procedure based on GAs, 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 Fig. 3, scenarios 1 and 5 are reproduced, together with their best reconstruction. As it can be seen, the reduction of the value of O.F. actually brings the solution towards the correct scenario reconstruction.

The results achieved show that the scenarios are properly reconstructed, even in case errors are obviously present; the goal to identify the areas where losses are concentrated seems to be reached. Among the selected methods, the GAs have shown a better performance, probably because they work with a number of individuals, while SA employs only one individual that can become trapped in a field where solutions are not allowed.

In the next phase of this work the instruments will actually be positioned. The procedure has the potential to be improved when real (actual) data becomes available. Undergoing research is related to the development of algorithms able to further minimize the O.F, as it is shown that this is a good indicator of the goodness of the solution. Moreover, a needed improvement is the development of a method able to identify whether or not more than one optimum is present. This would help in the decision about the minimum number of devices required to be installed.

Future developments will comprise the analysis of real data collected on the network, and the improvement of the computer program (i.e. simulator). Moreover, data will also be collected by means of a portable flowmeter. 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 the improvement of the existing indicators of water supply management.

References

[1] Alvisi, S., Grata, S. & Franchini, M., Leakage detection planning in water distribution systems. *Management of Water Networks - Proceedings of the Conference Efficient Management of Water Networks. Design and Rehabilitation Techniques*, De Angeli Editore, Milano, Italy, 2006.
[2] Halhal, D., Walters, G.A., Savic, D.A. & Ouazar, D., Scheduling of water distribution system rehabilitation using structured messy genetic algorithms. *Evolutionary Computation*, **7(3)**, 1999.

[3] Giustolisi, O., Laucelli, D. & Savic, D., Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment. *J of Civil Engineering and Environmental Systems*, **23(3)**, pp. 175–190, 2006.

[4] Engelhardt, M., Skipworth, P., Savic, D., Saul, A. & Walters, G., Rehabilitation strategies for water distribution networks: a literature review with a uk perspective. *Urban Water Journal*, **2**, pp. 153–170, 2000.

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

[6] Sundahl, A., Using break data on water pipe systems for renewal planning. *COST Action C3 workshop*, Brussels, 1996.

[7] Mambretti, S. & Orsi, E., Genetic algorithms for leak detection in water supply networks. *1st International Conference on Urban Water*, New Forest, UK, pp. 25–27, 2012.

[8] Back, T., Fogel, D. & Michalewicz, Z., *Handbook of evolutionary computation*. IOP Publishing Ltd. and Oxford University Press, New York and Oxford, 1997.

[9] Holland, J., Outline for a logical theory of adaptive systems. *Journal of the ACM*, **9(3)**, 1962.

[10] Holland, J., *Adaptation in natural and artificial systems*. University of Michigan Press, 1975.

[11] van Laarhoven, P. & Aarts, E., *Simulated Annealing: Theory and Applications*. Springer, 1987.

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

[13] Dorigo, M. & Stützle, T., *Ant Colony Optimization*. The MIT Press, p. 319, 2004.

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