

## Dealing with uncertainties in losses assessment in water supply networks: preliminary results

G. Becciu<sup>1</sup>, S. Mambretti<sup>1</sup>, P. S. Martins<sup>2</sup> & R. L. Moraes<sup>2</sup>

<sup>1</sup>*DICA, Politecnico di Milano, Italy*

<sup>2</sup>*School of Technology, UNICAMP, Brazil*

### Abstract

In many countries water losses can be much larger than 50%, with great economic losses because of the energy required for pumping and for the primary treatments that very often are required.

Methodologies have been developed in order to identify the areas where losses are most expected, limiting the excavations as far as possible.

To this end, the authors developed a methodology which requires the installation of a number of instruments on the network, in order to measure pressures and discharges; then, the demands at the nodes are changed by means of optimization methods and the network is simulated with a computer program, in order to match the readings of the instruments; losses are higher where the demands have been most increased.

In previous papers, the authors tested different evolutionary methods to identify the areas where losses are most expected, comparing results from theoretical networks that were, from all the other aspects, exactly identical. In the real world, unfortunately, differences from the simulated the real networks are not limited to the discharge demands, but other uncertainties are present.

The most evident uncertainty is related to pipe roughness, which can be considered a parameter of the model; sometimes even the diameters are listed wrong. In the paper, the robustness of the proposed method is tested using networks where roughness parameters and diameters are different, changing the roughness parameter in order to determine the influence of the errors in the evaluation of the areas where losses are most expected.

*Keywords: water supply networks, water losses, evolutionary computation, hydraulic modelling, models uncertainties.*



## 1 Introduction

The operators of public water systems are often challenged with the maintenance of system infrastructure that can deliver safe (and clean) drinking water to customers.

In many countries, and among them Italy and Brazil, this infrastructure has been in operation for many years and can be a significant source of water loss (e.g. loss through theft or unauthorized consumption, failure, administrative errors, inaccurate measurements, data handling or a combination of these).

Lost or unaccounted-for revenue water also represents a serious damage to the environment, and moreover sources are becoming more scarce due to pollution and the increase in demand.

Methodologies have been developed in order to identify the areas where losses are most expected, limiting the excavations as far as possible.

In particular, managers need to know where and how to intervene, (e.g. repair or substitution of a pipe) [1]. This issue is usually dealt with multi-objective optimization problem, where the objective function (OF) 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 the Authors to develop models that are either able to evaluate 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. The limited funds available constrain the invested annual budget and increase the importance of interventions time scheduling.

A new methodology has already been presented [7, 8] that identifies the areas where losses are mostly expected, by means of 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 computer program. 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: in the model, the volume of water losses is not inserted, because they are not known whereas in the actual network they are present.

The main hypothesis in the papers presented previously by the Authors was that the simulated network matches perfectly the real one; in the real world, uncertainties are present. The most obvious, and studied, is related to the pipe roughness, which is surely a parameter of the model, tied to the material but also to the age of the pipe, and therefore not known. Moreover, it may happen that, excavating, the actual pipe diameter differs from that listed in the records of the Company that manages the network. This latter problem is reducing while the networks are better studied, but still it is a problem to be considered as possible in the modelling.

This paper focuses on different methods of Evolutionary Computation in order to establish the best procedure to minimize the Objective Function when these uncertainties are present.

## 2 Evolutionary Computation

Complex and multi-objective optimization problem are often solved by means of Evolutionary Computation. The term Evolutionary Computation (EC) [9] represents a large spectrum of heuristic approaches to simulate evolution, which include: Genetic Algorithms (GA) [10, 11], Simulated Annealing, Particle Swarm Optimization, Artificial Neural Network, and others.

Following the previous jobs [7, 8], Genetic Algorithms [12] and Particle Swarm Optimization [13] have been used, followed by the Hill Climbing procedure.

The OF to be minimized is:

$$\min \left( \sum_{i=1}^{\text{Number of Control Nodes}} \frac{|H_{meas} - H_{comp}|}{|H_{meas}|} \cdot W_H + \sum_{i=1}^{\text{Number of Control Links}} \frac{|Q_{meas} - Q_{comp}|}{|Q_{meas}|} \cdot W_Q + \frac{|Q_{Globally Expected} - Q_{Globally Computed}|}{|Q_{Globally Expected}|} \cdot W_{GE} \right)$$

where  $W$  are weights, set equal to 1 for the theoretical scenarios, but that should depend on the expected precision of the real devices.

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

- Number of individuals per population: 2500;
- Number of generations: 100;
- Elitism: 20%.

Note that as the nodes are 440, this is also the number of the parameters to calibrate; the number of individuals of the population is taken as more than 5 times the parameters.

As for the PSO, the simulations are run with the following parameter values:

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

While in the previous papers the variables to be changed were only the discharges at the nodes, in the present paper the possibility to change pipe roughness is also allowed.

### 3 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 [7] and Mambretti *et al.* [8].

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; they are shown in figure 1 and detailed in table 1.

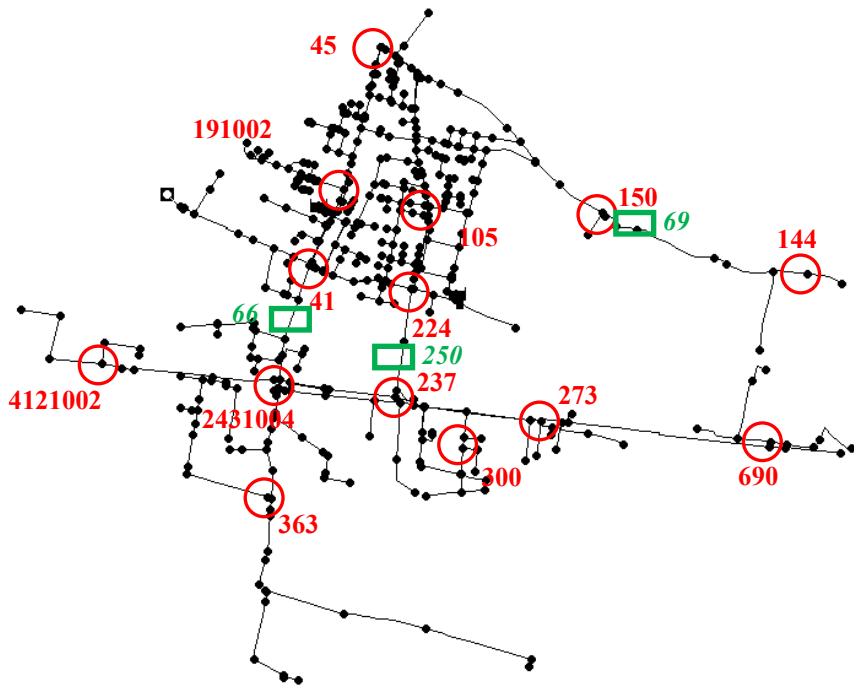


Figure 1: Position of the devices, and their ID.

In order to understand whether their number and position are appropriate to locate the leakages, different scenarios have been simulated to check whether they can be reconstructed by the algorithms mentioned in Section 2, finding that the scenarios can be reconstructed both with GA and PSO, if appropriately followed by HC.

For the present study, only one scenario of losses has been used, but under the conditions that pipe roughness in the model and in the real network:

1. are exactly the same;
2. are slightly different, and the roughness in the model cannot change;
3. are very different, and the roughness in the model cannot change;

4. are very different and the roughness in the model can change, without limits;
5. are very different and the roughness in the model can change, but there are penalties in the OF if the roughness falls outside than certain limits.

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).

Type	ID	Type	ID	Type	ID
P	45	P	273	P	71
P	150	P	690	Q	66
P	144	P	2431004	P	381004
P	191002	P	237	Q	185
P	41	P	363	Q	167
P	224	P	300	P	177100
P	105	Q	250	P	680100
P	4121002	Q	69	---	---

## 4 Results

### 4.1 Results for identical roughness values

In this case, under the hypothesis that the roughness in the real network matches perfectly that inserted in the mathematical model, allowing the changes of the roughness parameters worsen the final result.

Actually, this was expected, because the EC methods are not able to find the maximum of the OF, but only to get close to it. Therefore, changing the parameters (which, under the mentioned hypothesis, are already set to the optimum) prevents the algorithm from finding a solution as good as that found with the roughness parameters already set at their best value.

### 4.2 Results for different roughness values for the real and the modelled networks, using GA

For GA, results carried out under the mentioned scenarios are reported in Table 2. As usual [7, 8] ten runs have been performed, due to the random characteristics of the method. In order to test the sensitivity of the model to the roughness, the first simulations have been performed with small changes (values of  $n_{Manning}$  have been changed from standard [14] 0.011 to 0.010 and from 0.012 to 0.011, i.e. implying that in the model the pipes are considered smoother, which is usual as “manuals’ values” are normally referred to new pipes) and then larger changes have been made (values of  $n_{Manning}$  have been changed from 0.011 to 0.008 and from 0.012 to 0.009). After the first simulations, performed without allowing the program to modify the roughness to find the best solution, roughness has been considered a parameter to be modified and in the first

instance it was left without any constraints; after that, simulations have been performed inserting a penalty in the evaluation of the OF, trying penalty functions as the following:

$$\text{if Roughness} < 0.010 \text{ then Penalty} = W_p * (0.010 - \text{Roughness}) \quad (1)$$

$$\text{if Roughness} > 0.013 \text{ then Penalty} = W_p * (\text{Roughness} - 0.013) \quad (2)$$

In the OF function, the weight  $W_p$  allows changes to the constraints. In this paper, this parameter has been changed from 1 to 5 (results are presented with  $W_p = 2$ ); the value of the single Penalty might be very low, but considering that the network is composed by 440 links, the influence of this function is very high. Obviously, in order to compare the results, the values of the OF presented in Table 2 are reported without the added penalty. In Tables 2 and 3 the values reported in bold and italics are the best ones within the series of ten runs.

Table 2: Values of the OF for different model configurations: (a) Exact roughness; (b) Wrong roughness not modified (little changes); (c) Wrong roughness not modified (large changes); (d) Wrong roughness modified with GA without constraints (starting with large changes); (e) Wrong roughness modified with GA with constraints (starting with large changes). The best value for the ten runs is reported in bold and italics.

	(a)	(b)	(c)	(d)	(e)
1	0.3965	0.3268	0.6353	<b><i>0.6111</i></b>	0.4044
2	0.1619	0.3417	0.5925	0.9341	<b><i>0.3962</i></b>
3	0.1864	0.3181	<b><i>0.5733</i></b>	1.1265	0.5632
4	0.1633	<b><i>0.3024</i></b>	0.6442	1.1076	0.6382
5	0.2853	0.3159	0.6074	0.6598	0.4653
6	0.1956	0.3311	0.5971	0.7019	0.6209
7	0.1806	0.3209	0.6561	0.9143	0.5832
8	0.1876	0.3352	0.6355	0.6739	0.7253
9	0.1542	0.3158	0.5937	1.0280	0.4398
10	<b><i>0.1421</i></b>	0.3277	0.7333	0.8278	0.5793

It is possible to reach some considerations from these results.

First, changing the roughness does significantly affect the final results. This occurrence is very important especially when considering that the roughness parameters are estimated empirically, and the values inserted into the model never match the real ones (i.e. that could be measured only through experimental tests) – which would allow the calibration of the model but are too expensive to be performed.

The use of GA for calibrating these parameters actually lowers the value of the OF; however, not necessarily the “calibrated” roughness parameters match the real values after the calibration. That means the number of calibration points

(the measuring devices) is too limited to determine a single maximum, and therefore many maxima are possible and the algorithm finds one of them.

The addition of a penalty function improves the results. However, two problems have still to be solved. The former, inserting some limits implies the assumption of prior knowledge of the roughness value, which is not necessarily true. The latter, the best result carried out considering the penalty function is not related to the best simulation carried out without considering the penalty function. For instance, the best result reported in table 2, without considering the penalty function, is equal to 0.3962 (column *e*, table 2); if the penalty is added, the total value is equal to 4.1631. If we consider the best value of the total OF, this is equal to 1.8675, but it is related to a value (without the penalty) equal to 0.6389. As the best correspondence between the values read and simulated in the measuring devices is given by the OF without penalties, the value of 0.3962 is selected, even if this simulation has been performed using roughness parameters which have been selected outside the allowed range. In this way, the penalty is used to guide the algorithms towards an acceptable result, but it is not used for the selection of the best individual.

#### 4.3 Results for different roughness values for the real and the modelled networks, using PSO

Simulations have been performed also with PSO, allowing the model to change the roughness values and either imposing or not imposing constraints. Results are reported in Table 3.

In general, the use of PSO is much more complex than the use of GA. In the latter case, in fact, no calibration is required. The increase of the number of individuals and/or iterations theoretically increases the accuracy of the solution – or, at least, it does not make any difference.

Using PSO is completely different: the parameters used to find the new individual have to be accurately calibrated. Shortly, and referring to the literature [13] for a better description of the method, the individuals have a position and a velocity, and their subsequent position is computed using equations (3) and (4):

$$v_i(t+1) = \alpha \cdot v_i(t) + c_1 \cdot rand \cdot [x_{pbest}(t) - x_i(t)] + c_2 \cdot rand \cdot [x_{gbest}(t) - x_i(t)] + c_3 \cdot rand \cdot [x_{pbest}(\forall t) - x_i(t)] + c_4 \cdot rand \cdot [x_{gbest}(\forall t) - x_i(t)] \quad (3)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (4)$$

In the equation (3), we have:

- $\alpha$  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; in the paper value of 0.4 has been used;
- $c_1, c_2, c_3, c_4$  are acceleration parameters; in the paper values of 0.5 have been used for all  $c$  parameters;
- $rand$  is a random number  $\in [0,1]$ ;



$x_{pbest}$  is the individual with the local best (among neighbours);  
 $x_{gbest}$  is the individual with the global best.

For instance, a too-low value of the parameter  $\alpha$  does not allow the exploration of the whole space and it causes a premature convergence of the method towards a value away from the optimum; while a too-high value of  $\alpha$  produces instabilities and, again, the solution cannot be found.

First results, shown in Table 3, are worse than those produced with the GA. So far, however, this is the best combination of parameters found, while other values of  $\alpha$ ,  $c_1$ ,  $c_2$ ,  $c_3$ ,  $c_4$  brought final values of OF even worse. Therefore, as the EC are quite empirical methods, the parameters' values cannot be known in advance and it is possible that significant improvements can still be achieved. Again we have the problem that the best value with constraints is normally not the solution which values recorded and simulated are closest.

It is also to be noted that, in the present version of the method, an "individual" is stored in an array where the first  $N_{Nodes}$  values are the discharges at nodes, while the following  $N_{Links}$  values are the roughness at links. Unfortunately, with this implementation of the method, the coefficients are the same for both sets of parameters (discharge and roughness), and this is probably inappropriate due to their very different nature. Development of further research will include this variation.

Table 3: Values of the OF for different model configurations: (a) Wrong roughness modified with PSO without constraints; (b) Wrong roughness modified with PSO with constraints. The best value for the ten runs is reported in bold and italics.

	(a)	(b)
1	0.8987	0.7324
2	1.2176	0.8301
3	<b><i>0.7632</i></b>	1.0238
4	0.9637	0.6711
5	0.8528	0.7580
6	1.1485	0.8644
7	1.5098	<b><i>0.5391</i></b>
8	0.9257	0.5949
9	0.7952	1.1265
10	1.0474	0.6826

Despite the aforementioned challenges, PSO has an additional capability with respect to the GA, i.e. the possibility to calibrate the parameter in order to try to keep the populations at least partially separated, thus allowing the identification of multiple optima. This feature is important for three reasons: 1) to decide whether or not the number of devices is appropriate (if multiple optima are present when changing the discharges only, devices are not sufficient to identify the scenario of leakages); 2) to identify different scenarios of roughness



distribution, and 3) to investigate the most likely among the different solutions. Research will continue by coupling these models with a regression model that is able to select constraints based on more specific equations than (1) and (2), keeping in due consideration the age and the material of the pipes in order to select different limits for different pipes.

## 5 Conclusions

In previous papers a methodology that was developed to identify the areas where losses are mostly expected is 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 by means of a computer program.

In this paper, the methodologies used to minimize the OF (Genetic Algorithms and Particle Swarm Optimization) have been tested in presence of a further uncertainty, i.e. the roughness.

First results have been presented. Clearly, when the roughness in both the model and in the real network are exactly the same, results are better. Moreover, as EC methods cannot find the exact optimum, but they are able only to get close to it, the results carried out starting from the perfect value of the roughness to find the best leakage scenario are worsened.

However, even if the results are not the as good as expected, they are considered encouraging, taking into account the need to improve the selected methods.

Among the two methods, GA are much easier to use and lead to better results. On the other hand, PSO seems to be more promising and with improved possibilities, even if their calibration is much more demanding, and further research efforts have to be applied to achieve better results.

For instance, in this paper parameters to be calibrated such as discharges and roughness have all been treated in the same way. This is probably inappropriate because of their different nature and the different sensitivity of the hydraulic model to these parameters.

Nevertheless, despite the difficulties highlighted, PSO seems to be superior than GA due to the possibility of finding multiple optima, which is important in all cases of uncertainties presented in this work.

Further research will therefore try to perform a better calibration of the PSO models, allowing differences between parameters of different nature. It should also allow coupling the models with those able to select a feasible range of roughness, keeping in due account the age and the material of the pipes, and imposing constraints which may be different across all the conduits.

Future developments will also comprise the analysis of real data collected on the network, and the improvement of the computer program. 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. (2006) "Leakage detection planning in water distribution systems", in "Management of Water Networks" Proceedings of the Conference "Efficient Management of Water Networks. Design and Rehabilitation Techniques." Bertola and Franchini (Eds), Ferrara 2006, De Angeli Editore, Milano, Italy.
- [2] 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.
- [3] 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, Vol. 23, pp. 175-190.
- [4] 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.
- [5] 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.
- [6] Sundahl A., (1996) "Using break data on water pipe systems for renewal planning" COST Action C3 workshop, 18 and 19 June 1996, Brussels.
- [7] Mambretti S., Orsi E. *Genetic Algorithms for Leak Detection in Water Supply Networks*. 1<sup>st</sup> International Conference on Urban Water, 25-27 April, 2012, New Forest, UK.
- [8] Mambretti S., Martins P.S., Moraes R.L. (2013) *Evolutionary Computation Techniques to Assess Losses in Water Supply Networks* 7<sup>th</sup> International Conference on Sustainable Water Resources Management, 21 – 23 May, New Forest, UK.
- [9] Back T., Fogel D., Michalewicz Z. (1997) *Handbook of evolutionary computation* IOP Publishing Ltd. and Oxford University Press, New York and Oxford.
- [10] Holland J. H. (1962) "Outline for a logical theory of adaptive systems" *Journal of the ACM*, Volume 9, Issue 3; ACM.
- [11] Holland J. H. (1975) "Adaptation in natural and artificial systems" University of Michigan Press.
- [12] Goldberg D.E. (1989) *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley, Reading.
- [13] Kennedy J., Eberhart R. (1995) *Particle Swarm Optimization*, Proc. IEEE Int. Conf. on Neural Networks (Perth, Australia), IEEE Service Center, Piscataway, NJ, IV:1942–1948.
- [14] Rossman L.A. (2000) *Epanet 2. Users Manual*. EPA Office of Research and Development U.S., Environmental Protection Agency, Cincinnati.

