Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 186 (2017) 341 – 348

Procedia Engineering

www.elsevier.com/locate/procedia

XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

Population size influence on the efficiency of evolutionary algorithms to design water networks

Daniel Mora-Melià[a]*, F. Javier Martínez-Solano[b], Pedro L. Iglesias-Rey[b], Jimmy H. Gutiérrez-Bahamondes[a]

[a] *Universidad de Talca, Dpto. Ingeniería y Gestión de la Construcción, Curicó, 3340000, Chile.*
[b]*Universidad Politècnica de Valencia, Dpto. Ingeniería Hidráulica y Medio Ambiente, Valencia, 46022, Spain.*

Abstract

The optimal sizing in water distribution networks (WDN) is of great interest because it allows the selection of alternative economical solutions that ensure design requirements at nodes (demands and pressure) and at lines (velocities). Among all the available design methodologies, this work analyzes those based on evolutionary algorithms (EAs).

EAs are a combination of deterministic and random approaches, and the performance of the algorithm depends on the searching process. Each EA features specific parameters, and a proper calibration helps to reduce the randomness factor and improves the effectiveness of the search for minima. More specifically, the only common parameter to all techniques is the initial size of the random population (P). It is well known that population size should be large enough to guarantee the diversity of solutions and must grow with the number of decision variables. However, the larger the population size, the slower the convergence process.

This work attempts to determine the population size that yields better solutions in less time. In order to get that, the work applies a method based on the concept of efficiency (E) of an algorithm. This efficiency relates the quality of the obtained solution with the computational effort that every EA requires to find the final design solution. This ratio E also represents an objective indicator to compare the performance of different algorithms applied to WDN optimization.

The proposed methodology is applied to the pipe-sizing problem of three medium-sized benchmark networks, such as Hanoi, New York Tunnel and GoYang networks. Thus, from the currently available algorithms, this work includes evolutionary methodologies based on a Pseudo-Genetic Algorithm (PGA), Particle Swarm Optimization (PSO) and Harmony Search (HS).

First, the different algorithm parameters for each network are calibrated. The values used for every EA are those that have been calculated in previous works. Secondly, specific parameters remain constant and the population size is modified. After more than 500,000 simulations, the influence of the population size is statistically analyzed in the final solutions. Finally, the efficiency was analyzed for each network and algorithm. The results ensure the best possible configuration based on the quality of the solutions and the convergence speed of the algorithm, depending of the population size.

* Corresponding author
E-mail address: damora@utalca.cl

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems
doi:10.1016/j.proeng.2017.03.209

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems

Keywords: Water Distribution Systems; Design; Efficiency; Evolutionary Algorithms.

1. Introduction

The optimal lay-out, design and operation of water distribution networks (WDN) is of central importance to water industries and governments due to the vast capital investments associated with these tasks. While these problems are not independent of each other, they can be formulated and solved independently. This work is focused on the design, namely in the optimal sizing of pipes in WDN, ensuring requirements of demands and pressure at the nodes, and velocities at the lines.

This problem is normally interpreted as a NP-hard [1] problem mainly due to two reasons: nonlinear equations and discrete-valued diameters. A large variety of optimization methodologies have been developed for this problem, including approaches that reduce the complexity of the original nonlinear problem to facilitate the use of linear and nonlinear programming [2].

More recently, intelligent optimization techniques as Evolutionary Algorithms (EAs) have found applications in this area. The generic term "evolutionary computation" refers to a broad set of metaheuristic techniques to solve complex problems that base their performance on a similar mechanism to the processes of natural evolution. Besides, EAs have the advantage of being insensitive to the characteristics of the problem, so that since their introduction and subsequent popularization, EAs have been frequently used as an alternative optimization tool to conventional methods and have been successfully applied in a variety of areas, including optimization of WDN [3–6].

These techniques are based on the natural processes of evolution, using mechanisms to select the best combinations of the decision variables and to generate new solutions by recombination. It is because of that EAs require of setting the values of several algorithm components and parameters. The calibration of these parameters has a key influence on performance and efficacy of the algorithm [7,8].

The form and operation of some EAs are extensively studied, and one of the main conclusions is that the optimal value for these parameters is not universal for all problems, but it depends on the optimization problem and on the computational effort that will be spend in solving the problem. All EAs share some basic principles, but each has its own parameters, which guide the search algorithm to the best possible solution. Among all the parameters, the initial population size is the only one that can be considered common to all these techniques, and probably it has been one of the important topics to consider in evolutionary computation.

In this regard, it is possible to find different results in the literature on what may be the appropriate size of population in an optimization process [9,10]. Since the exploration capacity of the algorithm depends on the population diversity, researchers usually argue that a small population size could guide the algorithm to poor solutions, while a large population size could lead the algorithm to a computational time too high in finding a final solution. For this reason, in choosing the population size a trade-off between solution quality and search time should be done.

This paper applies an efficiency rate (E) to the initial population size of the algorithm. E relates the quality of the solution obtained to the computational effort involved to reach that solution. E has been calculated for the results obtained by three EAs for the pipe-sizing problem of three benchmarking networks. The selected algorithms include a PseudoGenetic Algorithm (PGA), a modified Particle Swarm Optimization Algorithm (PSO) and a Harmony Search Algorithm (HS). The different stages of the optimization process and the complete description of each EAs can be found in [7,5,11].

The presented method identify the most efficient population size for each of the analysed algorithms, and it can be applied to many other EAs and networks that have not been considered in this work.

2. Methods

This paper proposes a method to determine the most efficient population size of optimization algorithms for WDN pipe sizing. The proposed methodology includes two steps. First, the different algorithm parameters for each network are calibrated. Note that the parameters can be divided into two groups: specific parameters of each algorithm and common parameters for all algorithms. Specifically, the only parameter common to all EAs is the initial size of the random population (*P*).

This work considers only the calibration of *P*, whereas the remaining parameters are fixed in each algorithm. The constant values of these parameters were obtained from author´s previous work [8] and correspond with the most efficient combination for each of the algorithms included in this paper.

Regarding calibration of *P*, Eiben [12] distinguishes between two forms of parameter optimization: parameter control and parameter tuning. In parameter control, the initial parameter values are changing during the run, and it can be classified into three types depending on how the parameter change is made: deterministic, adaptative and self-adaptative. Despite parameter control have obvious advantages, the needed computation overhead leads to less efficiency on some problems.

On the other hand, in parameter tuning the parameters remain fixed during the optimization process, and maybe it is the most common approach to choose the best parameter calibration before starting the simulation. Some of the best known techniques are racing or sequential parameter optimization. In this sense, this paper adopts the methodology proposed by McClymont [13]. This methodology applies several algorithm trial runs and tunes the operators (*P* in this case), analyzing statistically the obtained results. After the parameters are calibrated, the second step of the method begins and efficiency rates have to be calculated, according to equations (1) and (2):

$$\eta_{\text{quality}} = \frac{N_{\text{successful}}}{N_{\text{sim}}} \quad (1)$$

$$E = \frac{\eta_{\text{quality}}}{\eta_{\text{convergence}}} \quad (2)$$

In these equations, $\eta_{quality}$ represents the quality of the solution related to a specific WDN sizing problem, so as to satisfy constraints imposed on the objective function. $N_{successful}$ is the number of “lowest” cost solutions and N_{sim} is total number of simulations performed. On the other hand, the term $\eta_{convergence}$ is related to the velocity of convergence it takes the algorithm to reach the final solution.

In this work, the term $\eta_{convergence}$ refers to the number of objective function (O.F) evaluations performed by the algorithm before finding the final solution of the problem. Finally, the ratio of $\eta_{quality}$ and $\eta_{convergence}$ defines *E* and gives an idea of the performance of the algorithm. Literally, *E* reports the number of “successful” evaluations obtained per call of the O.F. Similarly, 1/E represents the number of evaluations of the O.F for finding a successful solution of the sizing problem.

E represents a strategy to relate the quality of the solutions with the computational effort that requires each algorithm to find the final solution. Statistical analysis of *E* rates will determine which are the most efficient *P* for each algorithm and tested network.

3. Case studies.

The benchmark networks provide a common set of problems on which different researchers can quickly test their algorithms. Related to this, [14] classifies these benchmark problems into four groups (small, medium, intermediate and large), according to the size of search space. In this work, three medium-sized benchmarking networks are used to analyze the population size influence in EAs behavior: New York tunnel [15], Hanoi [16] and GoYang [17]. Table 1 summarizes the most relevant information of these benchmark problems, including the number of decision variables (DV), the range of diameters plus other information:

Table 1. Relevant information.for selected benchmarking networks.

Network	Number of pipes (DV)	Available commercial diameters	Search space	Best known solution ($x10^6$ um)
New York Tunnel	21	36; 48; 60; 72; 84; 96; 108; 120; 132; 144; 156; 168; 180; 192; 204 (inches)	$1.93x10^{25}$	38.642
Hanoi	34	304.8; 406.4; 508; 609.6; 762; 1016 (mm)	$2.87x10^{26}$	6.081
GoYang	30	80;100;125;150;200;250;300;350 (mm)	$1.24x10^{27}$	177.010[a]

In all benchmarking networks, EAs implemented reach the best solution available to date in literature. Note that the complexity of the problem is related not only to the size of the solution space but also to the number of local minima in each problem. Accordingly, the most complex problem between the analysed corresponds to Hanoi network. More details and a full analysis of the complexity of networks can be found in [8,11].

4. Results

As already mentioned above, the EAs require of setting some algorithm parameters. This calibration has great impact on performance and efficiency of the algorithm. Furthermore, it is possible to differentiate between specific parameters and common evolutionary algorithm parameters (*P*) for each algorithm. In previous work [11], a wide sweep for each specific parameter of each algorithm was made. Table 2 shows the specific parameters considered, and the optimal parameter calibration for each network and algorithm considering efficiency of the algorithm.

Table 2. Optimal specific parameters calibration for considered networks.

Algorithm/Network	Hanoi	New York	Go-Yang
PGA			
Crossover frequency (P_c)	No influence	>60%	No influence
*Mutation frequency (P_m)	3÷4%	4÷5%	2÷3%
PSO			
*Velocity limit (V_{lim})	20%	20÷30%	10%
*Confusion probability (P_{conf})	10÷20%	10%	10÷20%
Learning factor (C_1)	2%	2%	2%
Learning factor (C_2)	2%	2%	2%
HS			
*Harmony Memory Considering Rate (*HMCR*)	90÷95%	85÷90%	85÷95%
*Pitch Adjustment Rate (*PAR*)	10%	10÷40%	10÷20%
*More sensitive parameters of the algorithm for better performance.			

The data in the table above have been obtained from more than 500,000 simulations, in order to have a representative sample. Note that not all the parameters have the same importance in enhancing the search process. Thereby, certain parameters are extremely value-sensitive, while others do not influence too much on the performance of the algorithm.

Once it was obtained the best possible configuration for each specific parameter and according to calibration protocol, 200 simulations were performed for each initial population size (*P*) considered, keeping fixed all remaining specific parameters. This calibration protocol has been applied to the three EAs studying this work (PGA, PSO and HS).

For example, for the PGA algorithm, Fig. 1 reflects, on the left, the importance of setting parameters in these algorithms, as the optimal combination of parameters allows to significantly increasing the number of successful solutions. On the other hand, the right side shows the percentage of successful solutions when the initial population size is varied, keeping fixed all other specific parameters (crossover and mutation frequencies in this particular case). The repeatability of obtaining minimal solutions is related to $\eta_{quality}$ in equation (1).

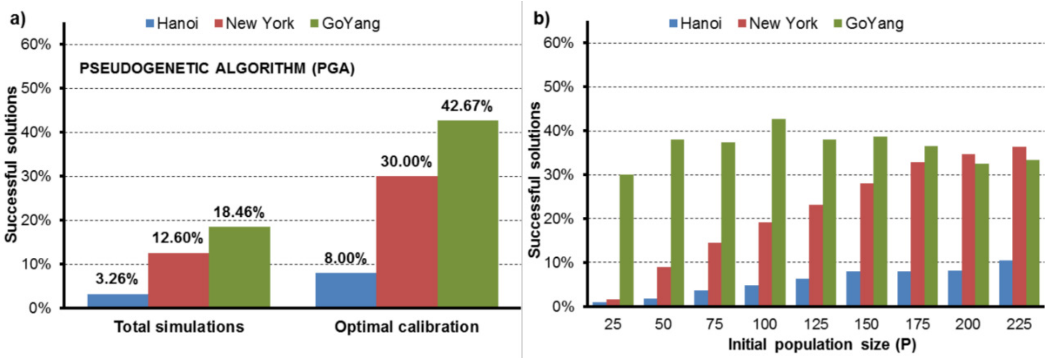

Fig.1. (a) Calibration of specific parameters, keeping fixed P; (b) Successful solutions varying P, keeping fixed remaining specific parameters.

As expected, the number of successful solutions is larger the greater is the initial population size of the algorithm. However, this improvement is not linear and all tested EAs reach a certain population size from which stabilizes and hardly improved in obtaining successful solutions. In the case of the PGA, the algorithm follows the general trend, obtaining most successful solutions to larger populations. However, it is noteworthy in the case of GoYang network no improvements are obtained for $P > 100$. This is logical, since this network is the simplest of all studied (fewer local minima), so that the probability of obtaining the best solution about 40% define the limit of the algorithm for this network.

Regarding convergence speed of each algorithm ($\eta_{convergence}$), it is important to consider several influence elements. On one hand, the specific calibration of parameters plays a key role in the speed of convergence to the final solution. For example, Fig. 2a) shows, for HS algorithm, the average number of O.F evaluations based on specific parameters *HMCR* and *PAR* for Hanoi network. The figure clearly shows that there is a relationship between the convergence speed and the calibration of specific parameters. Concretely, the higher the value of *HMCR* parameter is, the faster the final solution is found. On the other hand, the optimization process of each algorithm is an important differentiating factor in the eventual degree of convergence. Therefore, in general terms, the HS algorithm is the one that takes fewer iterations to reach the final solution, as shown in Fig. 2b). This pattern is repeated in all analyzed networks, so that the conclusions derived from the analysis of the Hanoi network can be extended to other networks.

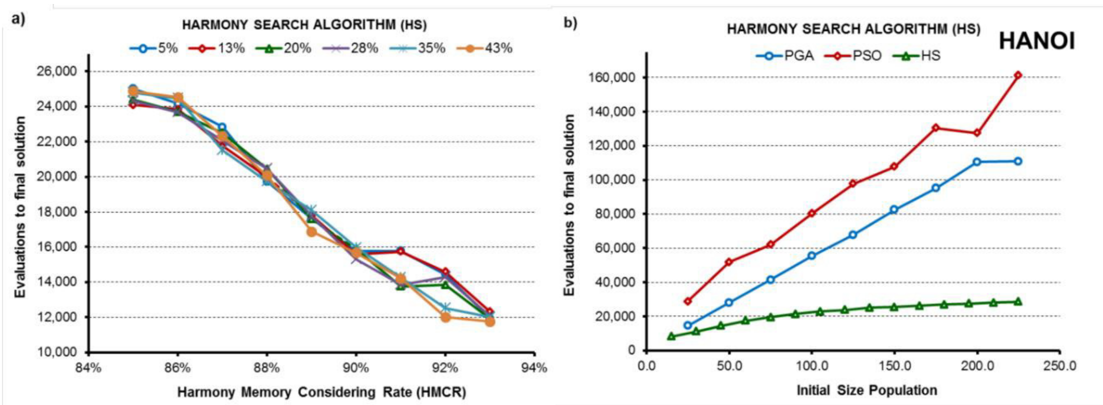

Fig.2. (a) Number of O.F evaluations based on PAR and $HMCR$; (b) Number of O.F evaluations to reach final solution based on P.

Finally, the rate between $\eta_{quality}$ and $\eta_{convergence}$ allows to define the efficiency rates according to equation (2). Fig. 3, Fig. 4 and Fig. 5 show *E* for obtaining successful solutions for Hanoi, New York and GoYang networks respectively. As mentioned above, *E* describes the level of performance of the algorithm while considering both the ability to obtain a minimum (successful) solution and the number of O.F evaluations required to reach this solution.

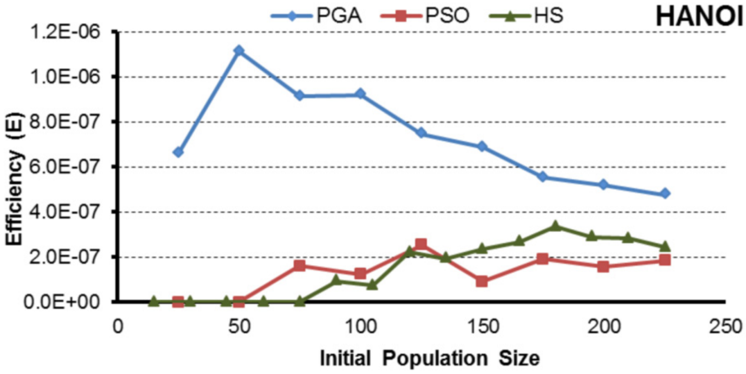

Fig.3. Efficiency of EAs in obtaining successful solutions for Hanoi network.

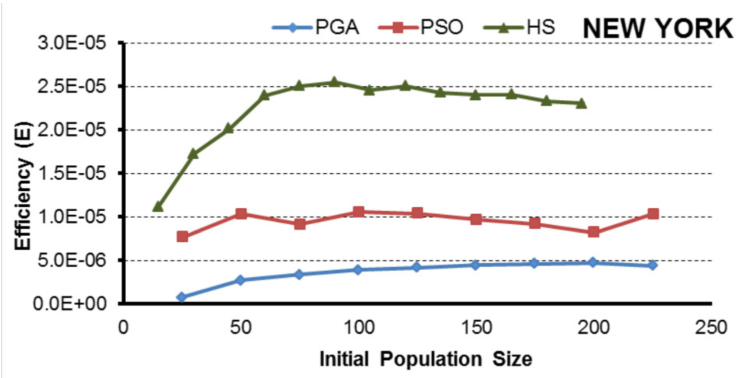

Fig.4. Efficiency of EAs in obtaining successful solutions for New York network.

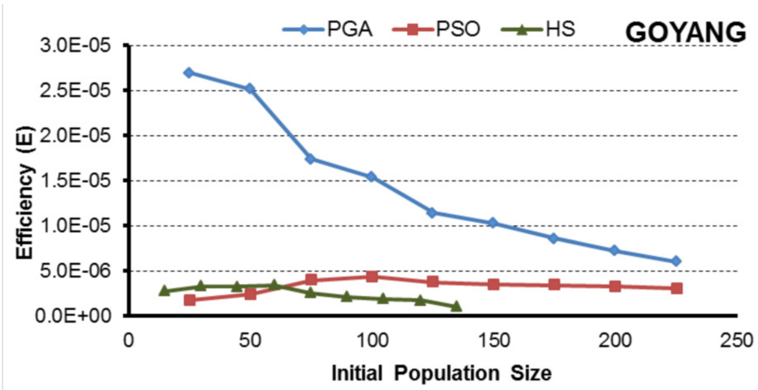

Fig.5. Efficiency of EAs in obtaining successful solutions for GoYang network.

In general terms, [11] demonstrates that PGA is the most efficient algorithm for sizing more complex networks. Among the networks tested in this work, Hanoi network is the most complex, because it is the one that has a greater number of local minima. In this case, numerically the *E* of PGA for the best population size is 4.3 and 5.0 times larger that of PSO and HS respectively. This means that for each successful solution obtained by HS, PGA obtains 5. The same results were repeated in GoYang network. In this case, the *E* value of PGA is 6.2 and 7.8 times higher than those of the PSO and HS algorithms, respectively. The main reason for these results is the great difficulty for PSO and HS to find minimal solutions. This penalizes them in terms of efficiency.

However, HS algorithm significantly improves for less complex networks, such as the New York network. In this case, all techniques identified minimal solutions easily, but its faster convergence favors the HS algorithm in terms of efficiency. Specifically, the efficiency of the optimally population size HS was 2.4 and 5.3 fold higher than those of the PSO and PGA algorithms.

Regarding the influence of initial size population, the results show how, in general terms, small population sizes are more efficient than large populations. For the PGA algorithm, the results show that for $P > 50$, the efficiency decreases as the size of population in Hanoi and GoYang networks. For New York network, no significant differences between the tested population sizes are appreciated for $P > 75$.

For the HS algorithm, the behavior is the same in all networks. *E* improves with increasing initial population sizes, up to a limit where *E* remains stable. This size population "limit" is different in each network, being greater the greater number of local minima has the analyzed network. Thus, the most efficient population size is 180, 150 and 60 for Hanoi, New York and GoYang respectively.

Regarding the PSO algorithm, this shows no significant difference in terms of *E* throughout the range of initial population size tested. In fact, PSO have some difficulties finding optimal solutions and besides the computational effort of this algorithm is not lower than those of other methodologies. Because of this, in any of the networks analyzed it has remarkable results in terms of *E*.

5. Conclusions

Nowadays, there are numerous methodologies based on Evolutionary Algorithms that can be applied to the design of WDN. The performance of these techniques depends largely on the setting for different calibration parameters. Agree with this, population sizing is one of the important topics to consider in parameter optimization.

This work applies a methodology for comparing algorithms based on an efficiency rate (*E*) and it is applied here to determine the influence of the initial population size with the quality of the solution and with the computational effort required to reach it. Both concepts are related by *E*.

Among all evolutionary algorithms available, three have been chosen: a PseudoGenetic Algorithm (PGA), a Particle Swarm Optimization algorithm (PSO) and a Harmony Search algorithm (HS). The methodology was tested on three medium-sized benchmark networks: Hanoi, New York Tunnel and GoYang. Thus, based on the statistical analysis of the results, it is possible to highlight the following conclusions:

- In terms of *E*, PGA is best for complex pipe-sizing problems, with a greater number of local minima, as Hanoi network. In this kind of problems, PGA is more likely to identify the lowest-cost solution. Meanwhile, HS and PSO hardly found minimal solutions, which severely decrease their efficiency.
- HS improves their performance in simpler pipe-sizing problems (New York network), since all algorithms are able to identify more frequently minimal solutions, but HS required fewer O.F evaluations to reach final solution.
- Overall, large initial population sizes are not more efficient than small populations in finding the best solution. In the case of PGA algorithm, populations around 50 individuals have been shown as the most efficient.
- Normally, there is a "limit" of population size from which the algorithms do not improve more in obtaining successful solutions. For the HS algorithm, this population size is greater the greater the complexity of the network.
- Finally, for the PSO algorithm not large variations are seen in the efficiency of the algorithm associated with the initial population size.

Acknowledgements

This work was supported by the Program Initiation into research (Project 11140128) of the Comision Nacional de Investigación Científica y Tecnológica (Conicyt), Chile.

References

[1] D.F. Yates, A.B. Templeman, T.B. Boffey, The computational complexity of the problem of determining least capital cost designs for water supply networks, Eng. Optim. 7 (1984) 143–155. doi:10.1080/03052158408960635.
[2] E. Alperovits, U. Shamir, Design of optimal water distribution systems, Water Resour. Res. 13 (1977) 885–900. doi:10.1029/WR013i006p00885.
[3] M. Eusuff, K. Lansey, F. Pasha, Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization, Eng. Optim. 38 (2006) 129–154. doi:10.1080/03052150500384759.
[4] A. Afshar, F. Massoumi, A. Afshar, M.A. Mariño, State of the Art Review of Ant Colony Optimization Applications in Water Resource Management, Water Resour. Manag. 29 (2015) 3891–3904. doi:10.1007/s11269-015-1016-9.
[5] D. Mora-Melia, P.. Iglesias-Rey, V.. Fuertes-Miquel, F.. Martinez-Solano, Application of the harmony search algorithm to water distribution networks design, Taylor & Francis Group, 2010. doi:10.1201/b10999-67.
[6] D. Mora-Melia, P.. Iglesias-Rey, V.. Fuertes-Miquel, F.. Martinez-Solano, Comparison of evolutionary algorithms for design of sewer systems, Taylor & Francis Group, 2010. doi:10.1201/b10999-66.
[7] D. Mora-Melia, P.L. Iglesias-Rey, F.J. Martinez-Solano, V.S. Fuertes-Miquel, Design of Water Distribution Networks using a Pseudo-Genetic Algorithm and Sensitivity of Genetic Operators, Water Resour. Manag. 27 (2013) 4149–4162. doi:10.1007/s11269-013-0400-6.
[8] D. Mora-Melia, P. Iglesias-Rey, F. Martínez-Solano, P. Muñoz-Velasco, The Efficiency of Setting Parameters in a Modified Shuffled Frog Leaping Algorithm Applied to Optimizing Water Distribution Networks, Water. 8 (2016) 182. doi:10.3390/w8050182.
[9] F.G. Lobo, D.E. Goldberg, The parameter-less genetic algorithm in practice, Inf. Sci. (Ny). 167 (2004) 217–232. doi:10.1016/j.ins.2003.03.029.
[10] A. Piszcz, T. Soule, Genetic programming, in: Proc. 8th Annu. Conf. Genet. Evol. Comput. - GECCO '06, ACM Press, New York, USA, 2006: p. 953. doi:10.1145/1143997.1144166.
[11] D. Mora-Melia, P.L. Iglesias-Rey, F.J. Martinez-Solano, P. Ballesteros-Pérez, Efficiency of Evolutionary Algorithms in Water Network Pipe Sizing, Water Resour. Manag. 29 (2015) 4817–4831. doi:10.1007/s11269-015-1092-x.
[12] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. doi:10.1007/978-3-662-05094-1.
[13] K. McClymont, E. Keedwell, D. Savic, An analysis of the interface between evolutionary algorithm operators and problem features for water resources problems. A case study in water distribution network design, Environ. Model. Softw. 69 (2015) 414–424. doi:10.1016/j.envsoft.2014.12.023.
[14] Q. Wang, M. Guidolin, D. Savic, Z. Kapelan, Two-Objective Design of Benchmark Problems of a Water Distribution System via MOEAs: Towards the Best-Known Approximation of the True Pareto Front, J. Water Resour. Plan. Manag. 141 (2015) 04014060. doi:10.1061/(ASCE)WR.1943-5452.0000460.
[15] J. Schaake, F.H. Lai, Linear programming and dynamic programming application to water distribution network design, M.I.T. Hydrodynamics Laboratory, Cambridge, 1969.
[16] O. Fujiwara, D.B. Khang, A two-phase decomposition method for optimal design of looped water distribution networks, Water Resour. Res. 26 (1990) 539–549. doi:10.1029/WR026i004p00539.
[17] J.H. Kim, T.G. Kim, J.H. Kim, Y.N. Yoon, A study on the pipe network system design using non-linear programming, J. Korean Water Resour. Assoc. 27 (1994) 59–67.