

## A two-stage evolutionary optimization approach for an irrigation system design

Milan Cisty, Zbynek Bajtek and Lubomir Celar

### ABSTRACT

In this work, an optimal design of a water distribution network is proposed for large irrigation networks. The proposed approach is built upon an existing optimization method (NSGA-II), but the authors are proposing its effective application in a new two-step optimization process. The aim of the paper is to demonstrate that not only is the choice of method important for obtaining good optimization results, but also how that method is applied. The proposed methodology utilizes as its most important feature the ensemble approach, in which more optimization runs cooperate and are used together. The authors assume that the main problem in finding the optimal solution for a water distribution optimization problem is the very large size of the search space in which the optimal solution should be found. In the proposed method, a reduction of the search space is suggested, so the final solution is thus easier to find and offers greater guarantees of accuracy (closeness to the global optimum). The method has been successfully tested on a large benchmark irrigation network.

**Key words** | irrigation, NSGA-II, optimal design, water distribution system

**Milan Cisty** (corresponding author)  
**Zbynek Bajtek**  
**Lubomir Celar**  
Faculty of Civil Engineering,  
Slovak University of Technology in Bratislava,  
Radlinskeho 11,  
Bratislava 810 05,  
Slovak Republic  
E-mail: [milan.cisty@stuba.sk](mailto:milan.cisty@stuba.sk)

### INTRODUCTION

Finding effective ways to build irrigation systems which meet irrigation demands and also achieve positive environmental and economic outcomes requires, among other activities, the development of new modelling tools. Due to the high costs associated with the necessary material and the installation of an irrigation water distribution system (WDS), it is essential to optimize the design of the WDS, while the hydraulic requirements (e.g., the required pressure on irrigation machines) of the network are satisfied. In past decades, a variety of optimization methods have been proposed for this purpose. The development of an optimization model of an irrigation network started with the application of the Labye graphic method (Labye *et al.* 1988) and linear programming. Since the early 1990s, researchers have subsequently focused on so-called heuristic optimization methods such as genetic algorithms (GAs) (Savic & Walters 1997), differential evolution (Vasan &

Simonovic 2010), artificial ant colony-based optimizations (Abbasi *et al.* 2010) and other approaches (Sheikholeslami *et al.* 2016). A good overview of the various heuristic methods applied to this problem can be found in, for example, Maier *et al.* (2014) or Giustolisi *et al.* (2015). The aforementioned research attempted to find a robust method for solving the optimal design of a water distribution network. Nevertheless, it must be pointed out that there is still some uncertainty as to how close various heuristic methods can get to a global optimum of this optimization task. Even for well-known small optimization problems such as the 'Hanoi' benchmark network (Fujiwara & Khang 1990), achieving the (known) best result requires a great deal of computational effort with substantial expertise in order to set up the optimization method parameters, and the individual results are not entirely guaranteed without a lot of experience. When a large-scale irrigation network

such as the ‘Balerma’ benchmark network (which is used for testing purposes in the present paper) is to be optimally designed, the uncertainty grows, which justifies the need for further research (Reca & Martinez 2006).

For solving the optimal design of a WDS, the authors of the present paper propose a new multi-step methodology that utilizes as its most important feature the ensemble approach, in which more optimization runs cooperate and are used together. This is contrary to the usual approach, where only one (the best) solution from several computational runs is selected and considered as a result of the optimization. This cooperation among multiple runs of the calculations serves for finding narrower boundaries of the possible values of genes in the definition of a GA chromosome and thus for a reduction of the search space size. The advantages of utilizing an ensemble approach (instead of picking the best solution from more runs) have already been demonstrated in various scientific domains, including several applications in water resources management and hydrology (Lauçelli et al. 2007; Solomatine & Ostfeld 2008).

The authors of the present paper assume that the main problem in finding the optimal solution of this task is the size of the search space in which the optimal solution should be found, so the reduction of the search space is proposed hereinafter as the second main feature of the proposed methodology. The effectiveness of reducing the search space has already been confirmed in various works dealing with the optimization of water distribution networks; e.g., a methodology based on the critical path method is suggested to reduce the search space in Kadu et al. (2008), and a methodology based on a dynamically expanding choice-table genetic algorithm has been applied to the New York Tunnels benchmark network in Zheng et al. (2011). A stage-wise approach consisting of deconstructing a looped network into tree networks and coupling heuristic and deterministic algorithms was used for the reduction of the search space in Cisty (2010) and Zheng et al. (2013).

In the present work, a multi-step optimization approach is proposed in such a way that the optimization is accomplished in two phases. In the first phase, suboptimal solutions are searched for; in the second phase, the optimization problem is solved with a reduced search space based

on these solutions, which significantly supports the finding of an optimal solution.

This paper is structured as follows: In the ‘Methodology’ section, WDS optimization is briefly explained and defined. This section subsequently describes the methods applied, and a multi-phase methodology is described therein. The experimental data, the design of the benchmark network, and the results are presented and discussed in the ‘Results and discussion’ section. In the ‘Conclusions’ section, the authors provide a description of the paper’s basic findings.

## METHODOLOGY

### Definition of the WDS optimization

In this work, the design of water distribution networks is viewed as a least-cost optimization problem with pipe diameters as the decision variables that must be selected from a discrete set of commercially available pipe diameters. The following formula should be minimized:

$$C_{\text{total}} = \sum_i^N C_i \cdot L_i \quad (1)$$

where  $C_{\text{total}}$  is the total cost of all the links (pipes) in the network;  $L_i$  is the length of the pipe  $i$ ;  $C_i$  is the unit cost of this pipe; and  $N$  is the number of pipes in the network.

As genetic algorithms (NSGA-II) are used in this work, many variants of the network design are successively evaluated during the optimization, mainly in terms of the pressures in many variants of the progressively optimized network. The software developed by the authors of this paper accomplishes this with EPANET, a widely used water distribution network simulation model (Rossman 2000). The authors of the paper chose this software because of its compatibility with other researchers, its speedy implementation in the C language, and its easy customization with the help of its so-called ‘Toolkit version’ (DLL).

In some of the network designs evaluated during the optimization process, a problem with insufficient pressure could occur. It is common practice to add a penalty term to the cost function (1) in order to penalize this or any

other constraint violations. This approach has been used in all the works dealing with the optimization of the benchmark network used in this study (e.g., Xuewei & Potter 2015). This technique requires a penalty factor (some number) to scale the constraint violations to the approximately same magnitude as the cost. Finding the proper penalty factor is essential to a successful optimization, but it is usually a time-consuming activity, and the investigator's expertise is essential. As this tuning is an obstacle to the application of optimization methods, a different approach has been proposed in this work; here, the preservation of the minimal allowable pressure in the network is treated as another optimization criterion (in addition to (1)), and a multi-objective optimization is used (Savic et al. 2011). Searching for the penalty factor is therefore not necessary. The second objective is defined as follows:

$$\text{Minimise } H_d = \sum_{j=1}^{N_N} \max(H_{\min} - H_j; 0) \quad (2)$$

where  $H_d$  is the total head deficit;  $H_{\min}$  is the minimum required head at a node;  $H_j$  is the actual head at the node  $j$ ; and  $N_N$  is the number of nodes in the network subject to hydraulic model constraints.

Another benefit from a multi-objective approach is that it offers many optimal solutions (a Pareto front), among which, other than the cheapest solutions could also be interesting, for instance, solutions with a little higher cost but a lower pressure deficit could be useful in some cases (e.g., when irrigation machines with different demands for a minimal pressure are expected to be used in the network in the future).

## Proposed methodology

The proposed solution of the given optimization problem is a two-phase computational procedure, which uses the NSGA-II algorithm (Deb et al. 2002). This algorithm is not new in optimizing water distribution networks (e.g., Artina et al. 2012; Wang et al. 2014); the authors want to present as their contribution a methodology for its usage in the context of the proposed two-step optimization process (Figure 1).

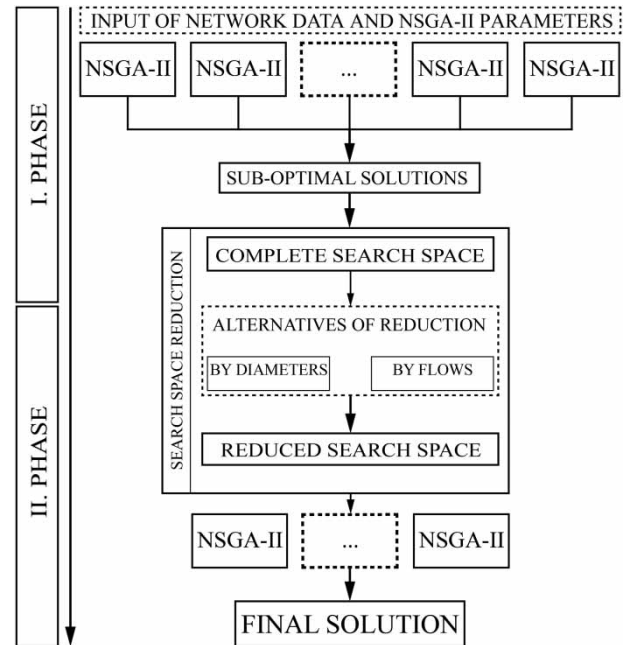


Figure 1 | Proposed methodology scheme.

The *first phase* of the optimization consists of several runs of the NSGA-II, which is applied in this phase by varying its parameters for every run, i.e., changing the population size, the number of generations, and the crossover and mutation parameters. This is done with the aim of obtaining different sub-optimal solutions which have a relatively low cost.

These sub-optimal solutions are subsequently used in the *second phase* of the proposed methodology, in which the final optimization run is built on sub-optimal solutions from the previous phase. The main idea in this phase is a reduction of the search space in which NSGA-II is searching for a solution. The search space is defined by the overall number of possible diameters for each link and their combinations from which the optimal solution could be selected. In the first phase for all the links, all the available diameters are considered. For the proposed benchmark network used in this paper (the Balerma irrigation network), there are 454 pipes, which have to be designed using a set of ten pipe diameters. This means that  $10^{454}$  alternatives are possible, which is a rather large search space for a heuristic algorithm.

Two alternative approaches were tested in the second phase of the proposed methodology to overcome this

problem. In the first alternative, diameters from the first phase's sub-optimal solutions are used to enhance the optimization run by reducing the search space in the second phase. In the second alternative, flows in the pipes from the sub-optimal solutions are used.

The outcome taken from the first phase for the purpose of the *second phase's first alternative* is a  $D_{ij}$  matrix of the diameters of all the suboptimal networks from the first phase, where  $i$  is a number of a suboptimal alternative, and  $j$  is a pipe number.

As is well-known, the solution in a genetic algorithm approach is coded as a chromosome. In this work, integer coding of the chromosome is used. These basic details have been described many times in the scientific literature; an interested reader can find them, for instance, in [Reca & Martinez \(2006\)](#). The chromosome of the second phase of the optimization consists of genes which represent pipes, where in each gene its lower and upper bounds are specified (the maximum and minimum allowable diameters for the corresponding pipe). The diameters are identified therein by the codes, which are the orders of the given diameter within a list of predefined potential diameters (e.g., in a price list). The lower bound for all the pipes (genes) is determined as follows:

$$D_1^{\min} = \min(D_{i,1}), D_2^{\min} = \min(D_{i,2}), \dots, D_n^{\min} = \min(D_{i,n}) \quad (3)$$

where range for  $i$  is  $(1, 2, \dots, n)$  and  $n$  is the total number of pipes. The upper bound of the chromosome is similar:

$$D_1^{\max} = \max(D_{i,1}), D_2^{\max} = \max(D_{i,2}), \dots, D_n^{\max} = \max(D_{i,n}) \quad (4)$$

where  $D_i^{\min}$  and  $D_i^{\max}$  represent the minimal and maximal allowable diameters for a gene (pipe)  $i$ . These two variables define new limit values for each gene of the chromosome; for example, with their help, a smaller search space is defined than in the previous phase, where for all the pipes (genes in a chromosome), the minimal and maximal diameters from the whole original set of available predefined diameters are used.

The *second alternative of the second optimization phase* is based on the minimal and maximal flows in each pipe and in each suboptimal alternative from the first phase. The outcome of the first phase for this purpose is represented by a

$Q_{ij}$  matrix of flows in all the suboptimal networks, where  $i$  is again the number of the suboptimal alternative, and  $j$  is the pipe number (as well as the gene number in a chromosome).

Based on these flows and the minimal and maximal flow velocities allowable in the pipe network ( $v_{\min}$  and  $v_{\max}$ ) are the minimal and maximal required diameters defined by the following formulas.

The minimal diameter:

$$D_1^{\min} = \sqrt{\frac{4 * \min(Q_{i,1})}{(\pi * v_{\max})}}, D_2^{\min} = \sqrt{\frac{4 * \min(Q_{i,2})}{(\pi * v_{\max})}}, \dots, D_n^{\min} = \sqrt{\frac{4 * \min(Q_{i,n})}{(\pi * v_{\max})}} \quad (5)$$

The maximal diameter:

$$D_1^{\max} = \sqrt{\frac{4 * \max(Q_{i,1})}{(\pi * v_{\min})}}, D_2^{\max} = \sqrt{\frac{4 * \max(Q_{i,2})}{(\pi * v_{\min})}}, \dots, D_n^{\max} = \sqrt{\frac{4 * \max(Q_{i,n})}{(\pi * v_{\min})}} \quad (6)$$

where  $Q_{i,n}$  represents the flow in pipe  $i$ , and  $n$  is the total number of pipes. The actual minimal and maximal diameters  $D_i^{\min}$  and  $D_i^{\max}$  are obtained by rounding them to the nearest existing diameter in the set of predefined allowable diameters. Again, due to the reduction of the possible gene values (by the limitations of their lower and upper bounds), a smaller search space is obtained, and the best search results can be expected.

## RESULTS AND DISCUSSION

The Balerma irrigation network ([Figure 2](#)) was chosen for verification of the proposed methodology because it is a relatively large, real-life network. Details about it can be found in [Reca & Martinez \(2006\)](#). It has a total of 443 demand nodes supplied by four source nodes (reservoirs). There are 454 pipes, arranged in eight loops, which are to be designed using a set of 10 PVC pipes with known unit

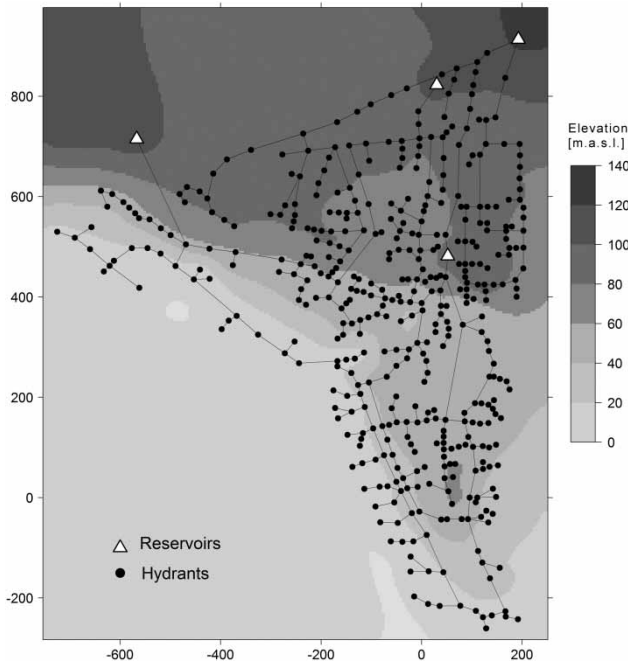


Figure 2 | Balerma irrigation project.

costs and diameters between 125 and 600 mm. The minimum required pressure head is 20 m for each demand node.

In the first phase, as described in the Methodology section, ten computational runs were accomplished, with varying standard NSGA-II settings (crossover probability: 0.93–0.98; mutation probability: 0.001–0.05). Based on information from the literature and based on our own experience, we can assert that these parameters are rarely outside these ranges, e.g., Pantil & Pawar (2015). Moreover, the aim was not to seek optimal settings, but to generate different sub-optimal solutions in the first stage of the calculations. The population size was also altered, and the number of allowable iterations (evaluations of the fitness function) was set to  $20 \times 10^6$ , which is sufficient in order to guarantee the best achievable solution; at the same time, the algorithm is not running unnecessarily long after its convergence. This value is easy to find by executing one very long computational run in the beginning. In all the runs, all ten allowable diameters were considered for all the pipes. These runs were conducted with the intention of producing a set of sub-optimal solutions of the first phase. The cost of the sub-optimal solutions obtained was from 1,965,341 to 1,997,940€.

The purpose of the second phase was to improve the results of the first phase by searching through the reduced search space. This reduction was implemented in two ways. In the first case, the reduction was based on the minimum and maximum diameters for each pipe from all the networks from the first stage. In this phase, NSGA-II did not consider diameters which were outside of this range. The reduction of the search space from  $10^{454}$  to  $10^{200}$  potential solutions was thus achieved, and, after the NSGA-II second phase computations, the best result published so far for the Balerma benchmark network was achieved in the presented work (1,921,400€). For testing purposes a total of ten optimization runs were conducted in this way (the mentioned solution is the best result from all of them), and the highest cost obtained was 1,925,082€, which is still among the best solutions published for this network.

The second alternative of the second optimization phase was based on the flows in the sub-optimal networks from the first phase. The minimum and maximum flow in each pipe was determined, and the minimum and maximum diameter for each pipe (or gene) was designated according to formulas (5) and (6). Based on this principle, the reduction of the search space from  $10^{454}$  to  $10^{138}$  was reached. For this alternative with ten optimization runs (with various settings of crossover and mutation), the best obtained result for the Balerma network was 1,927,758€ and the worst 1,940,799€. Although this alternative gives slightly worse results than the previous one, the authors also ran a variant of the second alternative, where all the networks obtained in the first stage were considered *individually*. The main idea of this computational experiment is that the presented methodology can also be applied with fewer optimization runs in the first phase, i.e., even only with one run, where the minimum and maximum flow is the same, and the minimal and maximal diameters are based only on the minimal and maximal velocities ( $0.3$  and  $2.5 \text{ m s}^{-1}$ ) and this flow. This reduction was determined with the help of the assumption of a significant closeness between the suboptimal and global flows. This alternative was run with each of the ten suboptimal solutions from the first phase individually with the best obtained result of 1,929,863€ and the worst of 1,942,019€. All the results are summarized in Table 1. In the works of the authors cited in the table, the number of



**Table 1** | Comparison of the best results of the Balerma network's optimization task published by various authors

Algorithm	Cost of the optimized network (€)		Number of evaluations	Source
	Minimum/Maximum	Average		
NSGA-II/First phase	1,965,341/1,997,940	1,988,887	10,000,500	This study
NSGA-II/Second phase alt.1	1,921,400 <sup>a</sup> /1,925,082	1,923,399	10,640,000	This study
NSGA-II/Second phase alt.2a	1,933,550/1,940,799	1,933,771	9,487,750	This study
NSGA-II/Second phase alt.2b	1,929,863/1,942,019	1,935,059	9,750,536	This study
Harmony search	2,018,000	–	10,000,000	Geem (2006)
HD-DDS	1,940,923	2,165,000	30,000,000	Tolson et al. (2009)
NLP-DE	1,923,000	1,927,000	1,427,850	Zheng et al. (2011)
PEDPSO	1,921,428	1,942,231	217,400	Xuewei et al. (2015)

<sup>a</sup>Best result.

necessary iterations needed to find the penalty parameter and other settings of the algorithm was not published, so the number of iterations is not evaluated in the table. When one is reporting only the number of iterations which are necessary for the final optimization, many iterations remain hidden, i.e., those which were necessary to find the parameters of the optimization solver, e.g., the penalty parameter. This tuning is not necessary in the work presented due to using a multi-objective approach, which is one of the greatest advantages of this approach. This means that although more computational runs are necessary because of the two-step character of the proposed methodology, this does not mean that the proposed methodology is more computationally demanding.

## CONCLUSIONS

The authors of this work have proposed a new two-step methodology for the optimal design of a WDS, in which more optimization runs (obtained in the first phase of the algorithm) cooperate and are used together in the second phase. Both phases of the optimization are accomplished by the NSGA-II algorithm. In this paper, the hypothesis that a better solution could be obtained in the case of the optimization of large water distribution networks was tested by incorporating a reduction of the search space into the optimization process. Two alternatives of this search space reduction were proposed and compared. If

it is desired to obtain as minimal a cost as possible, then it is recommended to use the first phase with approximately ten alternatives and apply a reduction of the search space on the basis of the pipe diameters, as is described in the paper. On the other hand, if a solution with a smaller computational time is sought, then only one optimization run in the first phase is enough, and an alternative based on the flows in this sub-optimal solution could be chosen.

A slightly better performance of the proposed methodology in comparison with the various published results is documented in Table 1. The improvement lies not only in achieving the lowest cost, but also in accelerating the optimization process through the proposed method (basically, no tuning of the optimization algorithm is necessary; it is only necessary to know the interval at which the usually appropriate parameters of the optimization method lie).

The knowledge gained from these computational experiments lies not in offering a new advanced heuristic or hybrid optimization method of a water distribution network, but in the fact that it is possible to obtain very similar results with simple, known methods if they are properly used methodologically. For the practical application of such advanced algorithms, one obstacle could be that they are sometimes difficult to obtain (considering the copyright protection of the author); on the contrary, our approach is based on the simple use of the easily obtainable NSGA-II, although the result is not guaranteed without a sufficient number of iterations. Thus, we are offering our

methodology as an option from this point of view. Its application requires no reprogramming of the original NSGA-II. The method is suitable for the design of large irrigation systems. It cannot be applied to drinking WDSs that contain various specific elements or when it is necessary to take diurnal demand patterns into account during their design, so its use for drinking water systems is currently somewhat limited. There was no room in this short article to verify the proposed ideas on more networks, which should be done in the future. Also, it is advisable to try in the future different (other than NSGA-II), and maybe more effective, multi-objective optimization methods as basic solvers in the context of this methodology, which could lead to a less computationally demanding optimization from the point of view of the number of iterations, on which we placed less emphasis in this work.

## ACKNOWLEDGEMENTS

This work was supported by the Slovak Research and Development Agency under Contract No. APVV-15-0489 and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, Grant No. 1/0665/15.

## REFERENCES

- Abbasi, H., Afshar, A. & Jalali, M. R. 2010 Ant-colony-based simulation-optimization modeling for the design of a forced water pipeline system considering the effects of dynamic pressures. *Journal of Hydroinformatics* **12** (2), 212–224.
- Artina, S., Bragalli, C., Erbacci, G., Marchi, A. & Rivi, M. 2012 Contribution of parallel NSGA-II in optimal design of water distribution networks. *Journal of Hydroinformatics* **14** (2), 310–323.
- Cisty, M. 2010 Hybrid genetic algorithm and linear programming method for least-cost design of water distribution systems. *Water Resources Management* **24** (1), 1–24.
- Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **6** (2), 182–197.
- Fujiwara, O. & Khang, D. 1990 A two-phase decomposition method for the optimal design of looped water distribution networks. *Water Resources Research* **26** (4), 539–549.
- Geem, Z. W. 2006 Optimal cost design of water distribution networks using harmony search. *Engineering Optimization* **38** (3), 259–280.
- Giustolisi, O., Berardi, L., Laucelli, D., Savic, D. & Kapelan, Z. 2015 Operational and tactical management of water and energy resources in pressurized systems: competition at WDSA 2014. *Journal of Water Resources Planning and Management* **142**, C4015002.
- Kadu, M. S., Gupta, R. & Bhawe, P. R. 2008 Optimal design of water networks using a modified genetic algorithm with reduction in search space. *Journal of Water Resources Planning and Management* **134** (2), 147–160.
- Labye, Y., Olson, M. A., Galand, A. & Tsiourtis, N. 1988 *Design and Optimization of Irrigation Distribution Networks*. FAO Irrigation and Drainage Paper, 44. FAO, Rome, Italy.
- Laucelli, D., Giustolisi, O., Babovic, V. & Keijzer, M. 2007 Ensemble modeling approach for rainfall/groundwater balancing. *Journal of Hydroinformatics* **9** (2), 95–106.
- Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A. & Ostfeld, A. 2014 Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions. *Environmental Modelling & Software* **62**, 271–299.
- Patil, V. P. & Pawar, D. D. 2015 The optimal crossover or mutation rates in genetic algorithm: a review. *International Journal of Applied Engineering and Technology* **5** (3), 38–41.
- Reca, J. & Martinez, J. 2006 Genetic algorithms for the design of looped irrigation water distribution networks. *Water Resources Research* **42**, W05416.
- Rossman, L. A. 2000 *EPANET 2 User's Manual*. US Environmental Protection Agency, Cincinnati, OH, USA.
- Savic, D. A. & Walters, G. A. 1997 Genetic algorithms for least-cost design of water distribution networks. *Journal of Water Resources Planning and Management* **123** (2), 67–77.
- Savic, D. A., Bicik, J. & Morley, M. S. 2011 A DSS generator for multiobjective optimisation of spreadsheet-based models. *Environmental Modelling & Software* **26** (5), 551–561.
- Sheikholeslami, R., Zecchin, A. C., Zheng, F. & Talatahari, S. 2016 A hybrid cuckoo-harmony search algorithm for optimal design of water distribution systems. *Journal of Hydroinformatics* **18** (3), 544–563.
- Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics* **10** (1), 3–22.
- Tolson, B. A., Asadzadeh, M., Maier, H. R. & Zecchin, A. C. 2009 Hybrid discrete dynamically dimensioned search (HD-DDS) algorithm for water distribution system design optimization. *Water Resources Research* **45** (12), W12416.
- Vasan, A. & Simonovic, S. 2010 Optimization of water distribution network design using differential evolution. *Journal of Water Resources Planning and Management* **136** (2), 279–287.
- Wang, Q., Savić, D. A. & Kapelan, Z. 2014 Hybrid metaheuristics for multi-objective design of water distribution systems. *Journal of Hydroinformatics* **16** (1), 165–177.

Xuwei, Q., Li, K. & Potter, W. D. 2015 Estimation of distribution algorithm enhanced particle swarm optimization for water distribution network optimization. *Frontiers of Environmental Science & Engineering* **10** (2), 341–351.

Zheng, F., Simpson, A. R. & Zecchin, A. C. 2011 [A combined NLP differential evolution algorithm approach for the](#)

[optimization of looped water distribution systems.](#) *Water Resources Research* **47** (8), W08531.

Zheng, F., Simpson, A. R. & Zecchin, A. C. 2013 [Coupled binary linear programming–differential evolution algorithm approach for water distribution system optimization.](#) *Journal of Water Resources Planning and Management* **140** (5), 585–597.

First received 18 March 2016; accepted in revised form 30 August 2016. Available online 25 October 2016