Zheng, Feifei; Simpson, Angus Ross; Zecchin, Aaron Carlo
Dynamically expanding choice-table approach to genetic algorithm optimization of water distribution systems Journal of Water Resources Planning and Management, 2011; 137(6):547-551

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010;
posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

A Dynamically Expanding Choice Table Approach to Genetic Algorithm Optimization of Water Distribution Systems

Feifei Zheng[1], Angus R. Simpson[2] and Aaron C. Zecchin[3]

Abstract: This paper proposes a modified genetic algorithm (GA) for optimization of water distribution systems. A method of dynamically expanding pipe choice table selections and reducing the number of decision variables is introduced that occurs during a GA run. Based on the progressive selection, an initially reduced size choice table for each decision variable is allowed to dynamically expand and then the number of decision variables is gradually reduced. This process enables the GA search to concentrate on promising regions of the search space. The dynamically expanding choice table genetic algorithm (GA_{DECT}) has been applied to a benchmark case study, the New York Tunnels Problem. The results obtained show that the GA_{DECT} yields a superior performance in terms of solution quality and computational efficiency.

CE Database subject headings: Optimization; Water distribution systems; Algorithms.

INTRODUCTION

Evolutionary algorithms have been introduced over the last 15 years to seek the least-cost design of water distribution systems. Among them, genetic algorithm (GA) optimization has gained popularity in terms of optimal design of water distribution systems because of its robustness and search performance (Simpson et al. 1994; Savic and Walters 1997). Many methods have been developed by researchers to improve the

[1]PhD student, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, 5005, Australia. fzheng@civeng.adelaide.edu.au
[2]Professor, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, 5005, Australia. asimpson@civeng.adelaide.edu.au
[3]Lecturer, School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, 5005, Australia azecchin@civeng.adelaide.edu.au

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

performance of GAs. A creeping mutation operator, variable power scaling of the fitness function and Gray coding (Dandy et al. 1996) were incorporated into the GA and were shown to be more effective. Vairavamoorthy and Ali (2000) applied integer coding in GAs to avoid the problem of redundant states often found when using binary or Gray codings. Wu et al. (2001) introduced a fast messy genetic algorithm to deal with optimization of water networks, showing significant improvement in terms of efficiency and robustness. Vairavamoorthy and Ali (2005) used a pipe index method to modify GA-based pipe optimization. Other evolutionary optimisation approaches have also been developed. Eusuff and Lansey (2003) proposed a shuffled frog leaping algorithm (SFLA) which showed improvement on the convergence speed in the context of optimal design of water distribution systems. Maier et al. (2003) applied ant colony optimization approach to optimize water distribution systems. Zecchin et al. (2006) proposed a Max-Min Ant System optimization (MMAS) and compared results obtained by GAs.

THE MODIFIED GENETIC ALGORITHM

Dynamically expanding choice tables

Typically all available diameters in the complete choice table for a decision variable are considered as potential choices for each pipe of the network when a GA is applied to optimize a WDS design. All regions within the solution space are considered to be equally important in the conventional GA, and hence, much computational effort is wasted on investigating infeasible or unnecessarily high cost regions within the search space.

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

In this research, a dynamically expanding choice table method is proposed to reduce the search space so that the GA can concentrate on promising regions of the search space. Initially, all the diameters in the full choice table are sorted from the smallest to largest and each pipe is given a diameter choice table. In the new method, only a subset of pipe sizes in the full choice table for each pipe (say the 3 successive middle sizes) are used to generate solutions randomly in the GA's initial population. During the GA run, if most of the members of the population in a generation have taken on the smallest diameter for a particular pipe from its corresponding reduced size choice table, this implies that this pipe diameter potentially can be further reduced in size to further reduce the cost of the whole network. Consequently, a smaller diameter is added to the pipe's current choice table and the choice table has been dynamically expanded. The same principle can be applied to the larger diameter options in a choice table. As a result, each decision variable in terms of pipe diameter size selects its own tailored choice table in the later generations.

Reduction of the number of decision variables

If the majority of members in a population select the diameter size for a particular pipe at the extremity of the full choice table, this pipe is locked to be the selected pipe size and then removed as a decision variable (whether it is either the smallest or largest diameter options). This process is used to dynamically remove such decision variables that cannot be further evolved as they have already converged at one extremity of the choice table. Therefore, the GA is able to more effectively and efficiently search the reduced search space, and focus on regions that show promise.

In summary, there are five cases that may occur for a choice table as shown in Fig. 1. Assume that the full choice table is made up of pipe diameters D1 to D10 ranked from

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

the smallest to the largest diameter. An initial reduced size choice table including D5, D6 and D7 (middle column in Fig. 1) is used to randomly generate the initial population of GA.

The following threshold percentages are defined (1) for expanding the choice table (P_e) (2) for removing decision variables (P_r) and (3) for when the majority of population members select the middle size of the current choice table during the GA run (P_s). Five cases are given as follows.

Case 1: For a particular pipe, if more than P_e percent of the members in a population select the smallest size (D5) of the current choice table (middle column of Fig. 1), a smaller pipe size D4 is *added* to the choice table (the second column from left in Fig. 1). Diameters of D4 and D5 are then randomly reselected for this pipe for all the members in the GA population.

Case 2: If more than P_e percent of the members in a population select the largest size (D7) of the current choice table (middle column of Fig. 1), a larger pipe size D8 is *added* to the choice table (the second column from right in Fig. 1). Diameters of D7 and D8 are then randomly reselected for this pipe for all the members in the GA population.

Now consider the situation where the choice table has been eventually expanded to include either the smallest or largest pipe:

Case 3: If more than P_r percent of the members in a population select the smallest size (D1) of the choice table (the first column from left in Fig. 1), this pipe is *removed* as a decision variable and the diameter for this pipe is locked at the minimum pipe size (D1).

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

Case 4: If more than P_r percent of the members in a population select the largest size (D10) of the choice table (the last column on the far right in Fig. 1), this pipe is *removed* as a decision variable and the diameter for this pipe is locked at the maximum pipe size (D10).

Now consider the situation where the majority of the pipes are the pipe size from the middle of the current choice table for that pipe:

Case 5: If more than P_s percent of population members select the middle size (D6) of the current choice table for a particular pipe during the GA run, all the pipe sizes in the current choice table are randomly *reselected* for this pipe in all members of the whole population. This process is used to maintain the population diversity, as occurs with the common mutation operator. However, case 5 is quite different from the normal mutation operator in that it only occurs when most of the population members select the middle pipe size diameter from its corresponding choice table.

CASE STUDY

The dynamically expanding choice table genetic algorithm (GA_{DECT}) was developed in C++ and combined with the EPANET2 hydraulic network solver. A total of 1000 independent optimization runs based on different random number seeds have been performed for New York Tunnel Problem (NYTP). The parameters settings used in GA_{DECT} are given in Table 1. Constraint tournament selection was used in GA_{DECT} (Deb 2000).

Case Study: New York Tunnels Problem

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

The New York Tunnels Problem (NYTP) has 21 existing tunnels and 20 nodes fed by the fixed-head reservoir. Details of this network, including the layout, the head constraints, pipe choices and costs, and water demands can be found in Dandy et al. (1996). The objective is to determine which pipes should be installed in parallel with the existing pipes to minimize the cost while satisfying the minimum head requirement at all nodes. The entire choice table for the NYTP case study involved 16 choices of pipe diameters consisting of {0, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 180, 192, and 204} inches.

An initial choice table with the diameters of {48, 60, 72} inches for each pipe was used to seed the initial population in the GA_{DECT} for the NYTP case study. One requirement of the proposed GA_{DECT} is that the threshold percentages (P_e, P_r and P_s) need to be specified. As given in Table 1, the parameter settings for the NYTP case study were as follows: P_e=65% (that is, expansion of the choice table occurred if more than 65% of the members selected the largest or smallest pipe size for a pipe from its choice table); P_r=95% (that is, if more than 95% of the members for a particular pipe have selected the smallest or the largest diameter size, this pipe is locked in to be the smallest or largest diameter and then removed as a decision variable); P_s=70% (that is, if more than 70% of the members selected a particular middle size for a pipe from its choice table, all the sizes in the current choice table are randomly reselected for this pipe for the whole population). An example of the initial choice table and the final choice table for a typical GA_{DECT} run applied to the NYTP, after dynamic expansion plus the decision variable removal, are shown in Table 2.

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010;
posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

As can be seen from Table 2, the second column is the initial choice table of {48, 60, 72} inches for diameters for each pipe and the third column is the final choice table for each pipe at the end of GA run. The final column is the least-cost solution found by the GA_{DECT} with a cost of $38.64 million (the current best known-least-cost solution). It is observed from Table 2 that choice tables for individual pipes were expanded differently during the GA run, despite the fact that they all started with the same initially reduced size choice table. The pipes labeled with a hash were removed as decision variables, as a pipe size of zero was selected during the GA run. From column 3 of Table 2, the total search space covered by the GA_{DECT} is given by $5^9 \times 7^9 \times 9^2 \times 11 \approx 7.0224 \times 10^{16}$, which is only a small fraction (3.62×10^{-9}%) of the size of the original solution space.

As can be seen from Table 2, some pipes (such as pipe 4, 6, 10, 11, 12, 13, 14, 15 and 20) moved towards the smaller pipe sizes during the GA run and finally were dropped as decision variables with a pipe size of zero, indicating that it was not economic for these pipes to be duplicated. However, several pipes (such as pipe 7, 16, 17, 18, 19, 21) were assigned larger sizes within the GA process, implying that these pipes were the potential candidates for duplication. It is noted that choice tables of some pipes (such as pipe 1, 2, 3, 5, 8, 9) expanded to larger diameters at the beginning and then to smaller diameters afterwards, showing that these pipes were indentified to be potential duplicates initially, but were eliminated from consideration in the later generations of the GA.

The dynamic reduction of the number of the decision variables for a typical GA_{DECT} run is shown in Fig. 2. At stage A in Fig. 2, there were 21 decision variables. After 16 generations (at stage B), pipe 11 was the first pipe dropped out as a decision variable with

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

a size of zero. The following sequence of pipes involving 4, 10, 12, 13, 14, 15 and 20 were consecutively eliminated. Thus, only 13 decision variables were left at stage C after 44 generations. Subsequently, pipes 2, 3, 5, 6 and 8 were removed as decision variables from stage C to D. After 176 generations (at stage E), only six decision variables were left, which were pipes 7, 16, 17, 18, 19 and 21. In the final stage, GA_{DECT} dealt with a reduced search space size and hence worked more efficiently.

RESULTS AND DISCUSSION

For the NYTP cases study, the current best known solution with a value of $38.64 million was first found by Maier et al. (2003) and this solution has been also found by the proposed GA_{DECT}. Fig. 3 gives a summary of a range of different sets of threshold values for GA_{DECT} applied to the NYTP case study. The GA_{DECT} program with each set of threshold values was performed for 1000 runs using different random number seeds. As can be seen from Fig. 3, GA_{DECT} with relatively high threshold percentages is able to find the best known solution with higher frequency, but at the expense of increased computational overhead. It was found that GA_{DECT} with P_e=65%, P_r=95% and P_s=70% exhibited overall well with an appropriate balance between performance in terms of frequency that the best solution was found and computational efficiency based on 1000 different runs.

The results for GA_{DECT} (P_e=65%, P_r=95% and P_s=70%) runs are given in Table 3. In order to enable a comparison of performance, the results of other optimization techniques that have previously applied to the NYTP case study are also included in Table 3. The best solution found by Improved GA (Dandy et al. 1996) and Messy GA (Wu and Simpson 2001) was $38.80 million, which deviates 0.414% from the best known solution.

In terms of efficiency, the proposed GA_{DECT} outperformed the other optimization techniques, but had slightly more average evaluations than the ACO (Maier et al. 2003). However, it is highlighted that there were only three different ACO runs used, whilst a total of 1000 different GA_{DECT} runs were performed in this study. The average cost solution produced by GA_{DECT}, based on 1000 different runs, is $39.06 million, which only deviates 1.087% from the known-least-cost solution. Even though the average cost solution provided by MMAS (Zecchin et al. 2006) and particle swarm optimization (PSO) (Dandy et al. 2010) are slightly lower than that of GA_{DECT}, the number of random number seeds are only 20 and 30 respectively, The GA_{DECT} was able to locate the current best solution 479 times out of a total of 1000 different runs, a higher frequency in finding optimal solutions than the PSO but slightly lower than that found by DE (Dandy et al. 2010).

CONCLUSION

A dynamically expanding choice table approach has been developed to enhance the performance of GA optimization for water distribution systems. The proposed approach provides a guide for the GA search to focus within regions of good fitness values. Thus, the search time is reduced and the optimal solution is more likely to be found. It is noted that, from the results of NYTP case study, the GA_{DECT} performed better than, or at least as good as, other optimization techniques.

REFERENCES

Dandy, G. C., Simpson, A. R., and Murphy, L. J. (1996). "An improved genetic algorithm for pipe network optimization." *Water Resources Research*, 32(2), 449-457.

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

Dandy, G. C., Wilkins, A., and Rohrlach, H. (2010). "A methodology for comparing evolutionary algorithms for optimising water distribution systems." Proceedings of the 12th *Water Distribution System Analysis Symposium*, Tucson, Arizona.

Deb, K. (2000). "An efficient constraint handling method for genetic algorithms." *J Comput. Methods Appl. Mech. Engrg*., 186,311-338.

Eusuff. M. M., and Lansey. K. E. (2003). "Optimisation of water distribution network design using shuffled frog leaping algorithm." *J. Water Resources Planning and Management*, 129(3), 210-225.

Maier, H. R., et al. (2003). "Ant colony optimization for the design of water distribution systems." *J. Water Resources Planning and Management*, 129(3), 200-209.

Simpson, A. R., Dandy, G. C., and Murphy, L. J. (1994). "Genetic Algorithms compared to other techniques for pipe optimization." *J. Water Resources Planning and Management*, 120(4), 423-443.

Savic, D. A., and Waters G.A. (1997). "Genetic Algorithms for least-cost design of water distribution networks." *J. Water Resource Planning and Management*, 123(2), 67-77.

Vairavamoorthy, K., and Ali, M. E. (2000). "Optimal Design of water distribution systems using genetic algorithms." *Computer-aided Civil and Infrastructure Engineering*, 15(2), 374-382.

Vairavamoorthy, K., and Ali, M. E. (2005). "Pipe index vector: A method to improve genetic-algorithm-based pipe optimization." *J. Hydraul. Eng.*, 131(12), 1117-1125.

Wu, Z. Y., Boulos, P. F., Orr, C. H., and Ro, J.J. (2001). "Using genetic algorithms to rehabilitate distribution systems." *J. Am. Water Works Assoc*., 93(11),74-85.

Wu, Z. Y., and Simpson, A. R. (2001). "Competent genetic-evolutionary optimization of water distribution systems." *Journal of Computing in Civil Engineering,* 15(2), 89-101.

Zecchin, A. C., Simpson, A. R., and Maier, H. R. (2006). "Application of two ant colony optimization algorithms to water distribution system optimization." *J. Mathematical and Computer Modelling*, 44,451-668.

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

Figure 1 An example of expanding of a choice table and reduction of decision variables

Figure 2 An example of dynamic reduction of number of decision variables

Figure 3 Results of GA_{DECT} with different sets of threshold values applied to the NYTP case study

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010;
posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

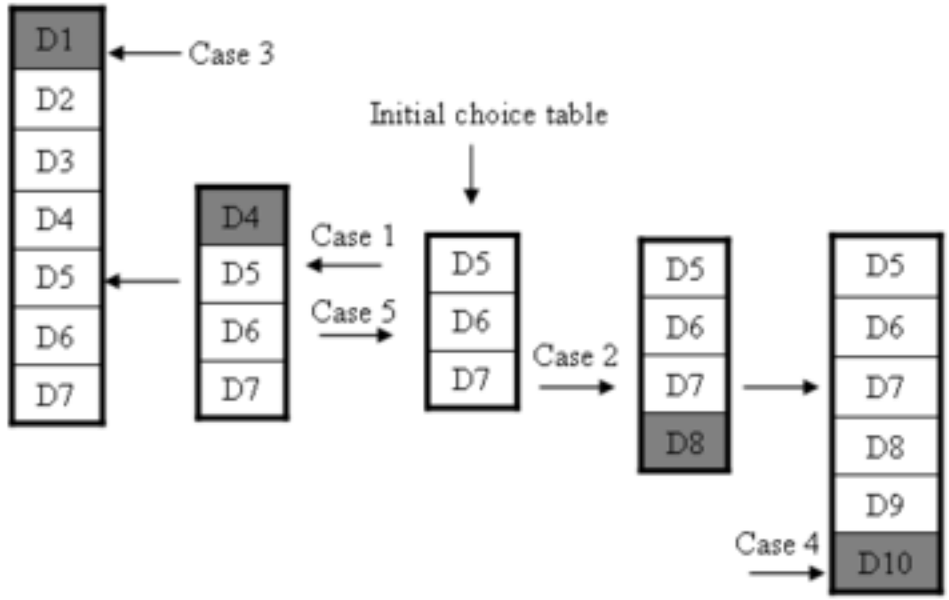

Figure 1 An example of expanding of a choice table and reduction of decision variables

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010;
posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

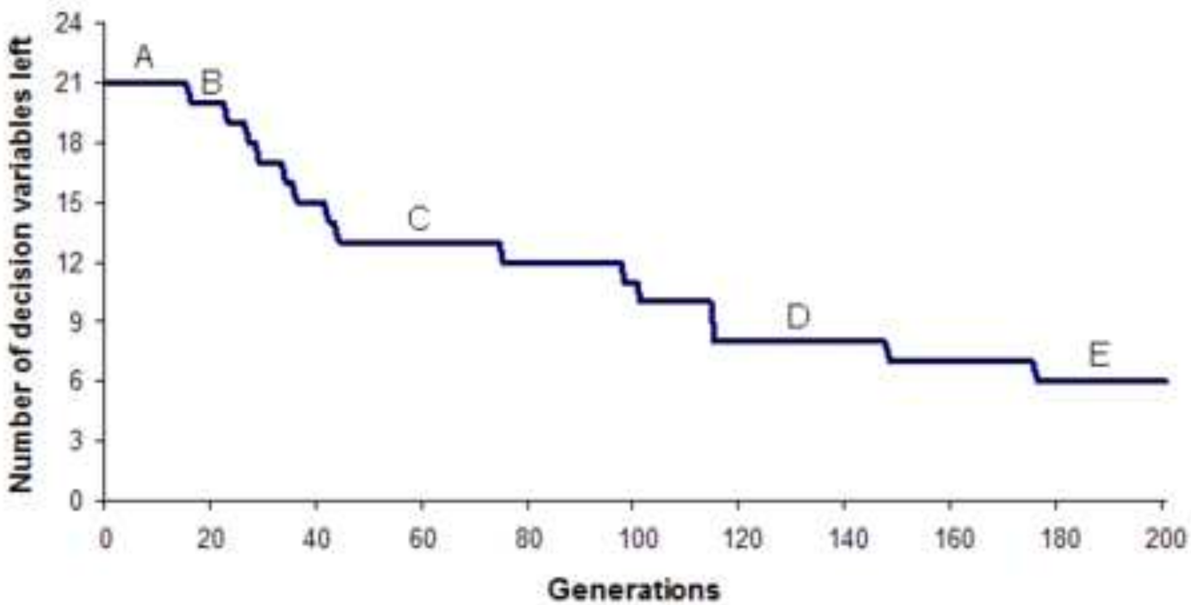

Figure 2 An example of dynamic reduction of number of decision variables

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010;
posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

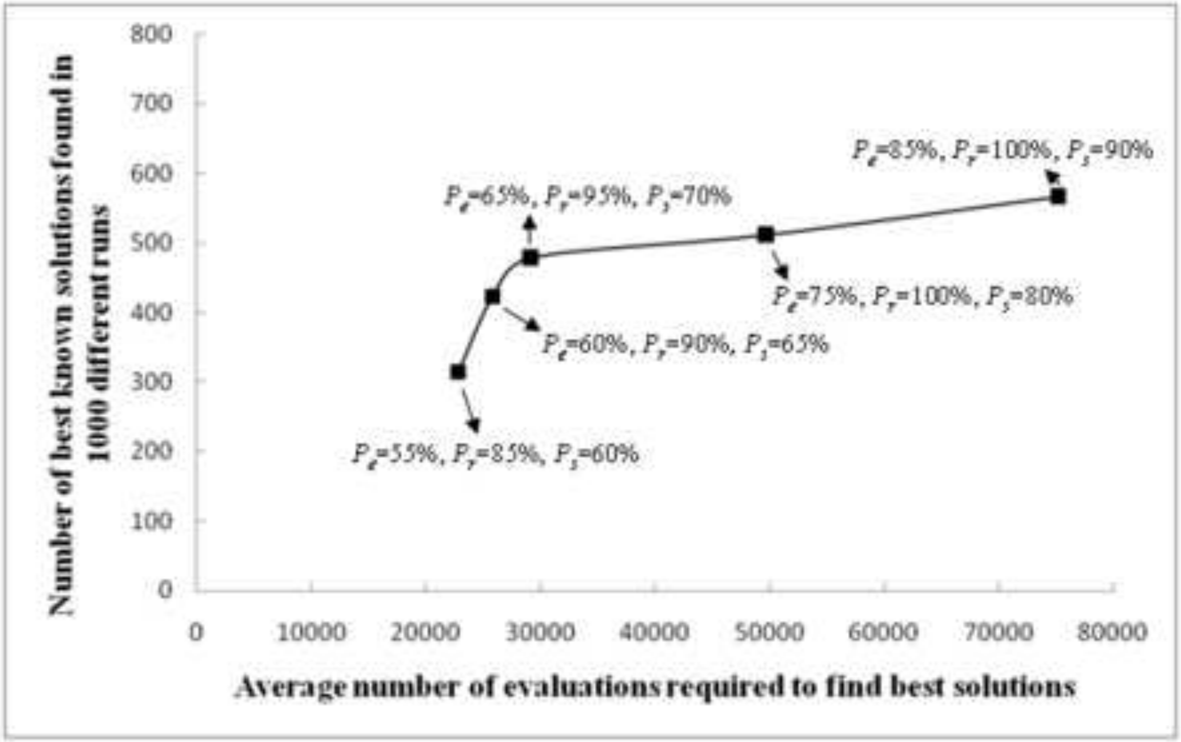

Figure 3 Results of GA_{DECT} with different sets of threshold values applied to the NYTP case study

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

Table 1 GA_{DECT} **parameter values for the NYTP case study**

Parameter	Value
Population size (N)	100
Maximum number of evaluations	100,000
Probability of crossover (P_c)	0.9
Probability of bitwise mutation (P_m)	0.0
Threshold percentage for expanding the choice table (P_e)	65%
Threshold percentage for removing decision variables (P_r)	95%
Threshold percentage for reselection (P_s)	70%

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

Table 2 An example of the expansion of choice tables and removal of decision variables during the GA_{DECT} process applied to the NYTP case study

Links	Choice table for pipe diameters (inches)		Final solution (inches)
	Initial choices	Choices at end of the run	
1#	48, 60, 72	0, 36, 48, 60, 72, 84, 96	0
2#	48, 60, 72	0, 36, 48, 60, 72, 84, 96	0
3#	48, 60, 72	0, 36, 48, 60, 72, 84, 96	0
4#	48, 60, 72	0, 36, 48, 60, 72	0
5#	48, 60, 72	0, 36, 48, 60, 72, 84, 96	0
6#	48, 60, 72	0, 36, 48, 60, 72	0
7	48, 60, 72	48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168	144
8#	48, 60, 72	0, 36, 48, 60, 72, 84, 96	0
9#	48, 60, 72	0, 36, 48, 60, 72, 84, 96	0
10#	48, 60, 72	0, 36, 48, 60, 72	0
11#	48, 60, 72	0, 36, 48, 60, 72	0
12#	48, 60, 72	0, 36, 48, 60, 72	0
13#	48, 60, 72	0, 36, 48, 60, 72	0
14#	48, 60, 72	0, 36, 48, 60, 72	0
15#	48, 60, 72	0, 36, 48, 60, 72	0
16	48, 60, 72	48, 60, 72, 84, 96, 108, 120, 132, 144	96
17	48, 60, 72	48, 60, 72, 84, 96, 108, 120, 132, 144	96
18	48, 60, 72	48, 60, 72, 84, 96, 108, 120	84
19	48, 60, 72	48, 60, 72, 84, 96, 108, 120	72
20#	48, 60, 72	0, 36, 48, 60, 72	0
21	48, 60, 72	48, 60, 72, 84, 96, 108, 120	72
Cost ($M)			38.64

Pipe was locked in at zero size and eliminated as a decision variable during the GA_{DECT} process.

Journal of Water Resources Planning and Management. Submitted July 1, 2010; accepted December 27, 2010; posted ahead of print March 14, 2011. doi:10.1061/(ASCE)WR.1943-5452.0000153

Table 3 Comparison of algorithmic performance applied to the NYTP case study

Algorithm	No. of runs	Best solution ($M)	Average cost ($M)	No. of average evaluations	No. of best solution found
GA_{DECT} (P_e=65%, P_r=95% and P_s=70%)	1000	38.64	39.06	29,101	479
Improved GA[1]	5	38.80	38.98	143,790	NA
MMAS[2]	20	38.64	38.84	30,711	NA
ACO[3]	3	38.64	NA	13,928	NA
Messy GA[4]	5	38.80	39.09	48,427	NA
PSO[5]	30	38.64	38.93	NA	10
DE[5]	30	38.64	40.33	NA	22

[1]Dandy et. al (1996). [2]Zecchin et. al (2006). [3]Maier et al. (2003). [4]Wu and Simpson (2001). [5]Dandy et. al (2010). NA means "not available"