

The Second SREE Conference on Engineering Modelling and Simulation (CEMS 2012)

Gene Expression Programming Algorithms for Optimization of Water Distribution Networks

Huanhuan Wang^a, Shuming Liu^{a*}, Fanlin Meng^a, Mingming Li^a

^a School of Environment, Tsinghua University, Beijing, 100084, China

Abstract

Water distribution system (WDS), as the important and expensive infrastructure, plays a crucial role in supplying water for the citizens living in the urban areas. The huge cost of the system drives researchers to seek the optimal cost design. During the last several decades, evolutionary methods such as genetic algorithms (GA), ant colony optimization algorithms (ACO) and simulated annealing (SA) have been found to explore the optimal combination of pipe diameters for design of WDS. However, in the case of complex distribution network, those methods suffer with low computational efficiency. In this paper, the gene expression programming (GEP) algorithms, a method based on the genetic algorithms, is adopted to solve the optimal design of WDS. This method is applied to the Hanoi benchmark network, the result of 6.081 million costs and 15,000 evaluations is combined with that obtained with GA and SA. From the results, it is observed that the proposed algorithm is an advanced alternative for the design of WDS, taking the computational efficiency and the ability of finding global or near-global optimal solutions into consideration.

© 2012 Published by Elsevier Ltd. Selection Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Water distribution network, water supply, Gene Expression Programming, Optimal Design

1. Introduction

Water distribution system (WDS) as the important and expensive infrastructure is one of the urban lifelines and the study of its design plays a very important role in urban management. The pipes account for the major capital cost of WDS in these components, which transfer water from one node to another

* Corresponding author. Tel.: +86-10-62787964; fax: +86-10-62787964.

E-mail address: shumingliu@tsinghua.edu.cn.

one. The optimal combination of pipe diameters should be able to provide the nodes' hydraulic heads larger than that of the minimum required values at all the demand nodes. The hydraulic head available at a special demand node depends on head-loss values connected with the supply links of the node and hydraulic head available at the upstream node. The head-loss values connect with a pipe differs from the pipe material, length, diameter, and discharge. The pipe material and length are fixed and discharge carried by the pipe depends on its diameter. So the pipe diameters become the decision variables in the optimal design problem of WDS. Heuristic methods provide the satisfactory, but not necessarily optimal solutions to the complex problems in a fast way. And the application of heuristic optimization algorithms has been retained as an active research area for optimal WDS design. In the optimization algorithms, the objective function calls for evaluation at every trial made with different values on decision variables. Based on the objective function value obtained, the heuristic search progress is achieved. One of the most promising and commonly used methods is Genetic Algorithm(GA) [1], which is an adaptive stochastic algorithm based on natural selection and genetics. Many researchers have proposed genetic algorithms for the optimal WDS design. Other meta-heuristic methods have also been recently applied to explore the optimal pipe diameters, such as Simulated annealing(SA) [2], which simulates the physical annealing process that metal gradually cools down from a very high temperature to attain the solid crystalline form, harmony search method, the ant colony optimization(ACO) [3], which is based on the foraging behavior of ants. Mohan and Babu [4] applied the method of Honey-bee Mating Optimization(HBMO) to this problem.

In this paper, the gene expression programming (GEP) [5], a search technique that evolves computer programs and an extension of GP and GA, has been proposed to optimize WDS. The analysis reveals that GEP has the capability of exploring the optimal pipe diameters from the discrete choices with relatively less number of evaluations. Findings from this study indicate that GEP is an alternative to other heuristics methods for the design of water distribution system, taking the computational efficiency and the ability of finding the global or near-global optimal solutions into consideration.

2. GEP Optimization

Gene expression programming (GEP), an extension of GP and GA, is a search technique that evolves computer programs (mathematical expressions, decision trees, polynomial constructs. The computer programs of GEP are all encoded in linear chromosomes, which are then expressed or translated into expression trees (ETs). ETs are sophisticated computer programs that are usually evolved to solve a particular problem, and are selected according to their fitness at solving that problem. Thanks to genetic modification, population of ETs will discover traits and therefore will adapt to the particular problem they are employed to solve. It means that, within enough time and setting the stage correctly, a good solution to the problem will be discovered. GEP is a full-fledged genotype to phenotype system, with the genotype totally separated from the phenotype, while in GP, genotype and phenotype are one more mess and formally, as a simple replicator system. As a result of this, the full-fledged genotype to phenotype system of GEP surpasses the old GP system by 100-60,000[5, 6].

. When applying a GEP to the design of water distribution systems, two main difficulties are experienced, the redundant values generated when using a binary alphabet and the large computational effort associated with using a hydraulic solver for every string of a population. As with previous researchers who applied GA and ACOA to the design of water networks, in this study a gene represents the diameter of pipe to be selected from a group of finite discrete diameters.

Figure 1 shows the flow chart of a gene expression programming. The structural organization of GEP genes is better understood in terms of open reading frames. Consider, for example, the algebraic expression: $x \times y + x / y$, which can also be represented as a diagram or ET, shown as Figure 2.

This kind of diagram representation is in fact the phenotype of GEP individuals, being the genotype easily inferred from the phenotype:

{0123456} is the {+×/xyxy}, which is the straightforward reading of the ET from left to right and from top to bottom. The diameter can be designed for gene as an element of an individual.

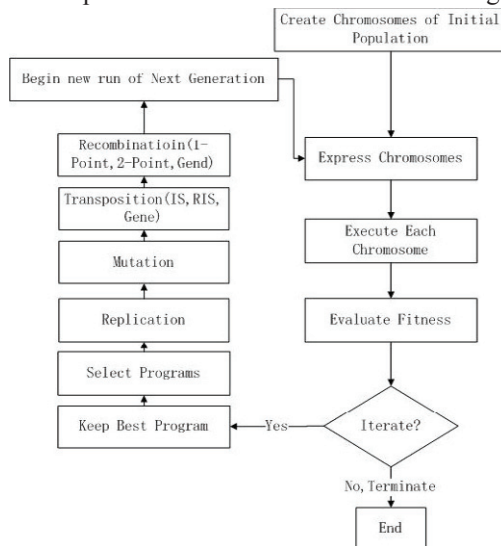


Fig. 1. The flowchart of gene expression algorithms

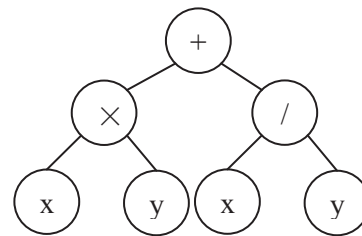


Fig.2 Schematic diagram of ET

3. Optimization Model for Optimization of WDS

In any WDS, the source head, elevation of demand nodes, demand values and pipe lengths and layout are known in advance. So the typical single-objective formulation of the optimal design of water distribution systems [7, 8] makes its best to minimize the cost with pipe diameters as the logical decision variables, while pipe layout, connectivity and demands as imposed as constraints. So the optimization problem for water distribution systems design can be stated mathematically as to

$$\text{Minimize } f_{\text{cost}}(D) = \sum_{i=1}^{NP} c_i L_i \quad (1)$$

Subject to:

$$\left\{ \begin{array}{l} \sum Aq + Q = 0 \\ lh = 0 \\ h_{\min} \leq h_i \leq h_{\max}, i \in ND \\ v_{\min} \leq v_i \leq v_{\max}, i \in NP \\ D_i \geq D_{\min}, D_i \in \{D_1, D_2, \dots, D_k\} \end{array} \right. \quad (2)$$

Where $f_{\text{cost}}(D)$ is the total cost of the network; D is the diameter of pipes; c_i is the cost of the pipe with diameter i per unit length; L_i is the total length of pipe i in the network; NP means the number of pipes; ND respects the number of nodes; h_i is the node pressure of node i and h_{\min}, h_{\max} are the

minimum and maximum value of node pressure, the same as pipe flow, v ; k is the list number of available pipes.

4. Case Study

The performance of the GEP model developed for optimization of the least-cost design of a water distribution system problem is evaluated by the optimization of the well-known Hanoi network.

4.1. Hanoi network

The well-known pipe system, Hanoi network in Vietnam [9] shown in Figure 3 is considered to be optimized with proposed method. It consists of 34 pipes, 32 nodes and 3 loops. There is a reservoir at node 1 with 100 meters Hydraulic Grade Line feeds the system. The other nodes (2 to 32) are demand nodes with minimum Hydraulic Grade Line requirement of 30 meters. For all the pipes, Hazen-William coefficient is 130. Moreover, α , β and ω , which is 10.6668 in EPANET 2, in the Hazen-Williams equation, are 1.852, 4.87 and 10.6668. The pipes can take one of six diameter options, thus the optimization gives a search space size of $6^{34} \approx 2.87 \times 10^{26}$ to evaluate the optimization of WDS. The cost function is nonlinear, $f_{\text{cost}}(D) = 1.1D_i^{1.5}L_i$, which cost is in dollars, diameter is in inches, and length is in meters

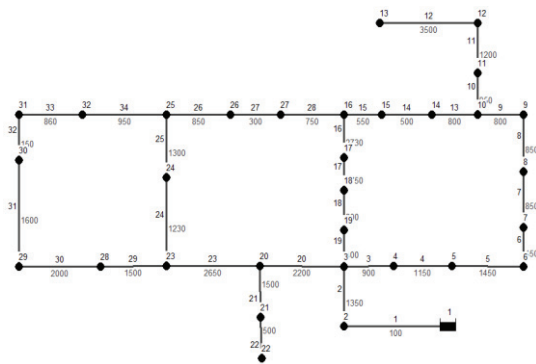


Fig. 3. Hanoi Network

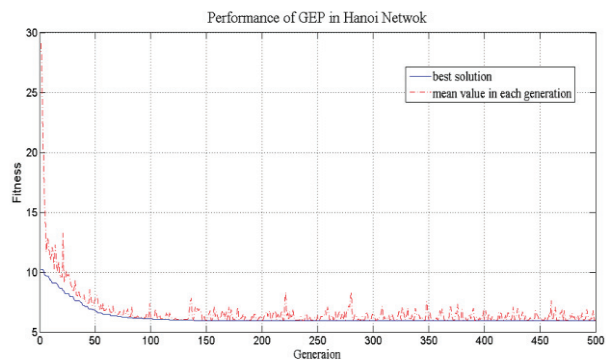


Fig. 4. Evolution process of the best solution with GEP for Hanoi Network

In order to avoid the randomness of the search process due to use of different initial solutions, the result has been obtained taking the executing three times for GEP, using different parameter configurations for the Hanoi network, with the same initial solution of maximum diameters(40 in) in each pipes. The crossover probabilities vary from 1.0 to 0.7 with each 0.05 interval and the mutation probabilities differ from 0 to 0.05 with each 0.01 interval. The best parameter values adopted are a population size of 100, 0.9 probability of crossover, 0.03 probability of mutation and the maximum number of generations is set to 500. The diameter of pipe is designed for genes in $\{+\times\times\text{aaaa}\}$ with function set of $\{+\times\}$ and three constants 0 to 2; One point and two point recombination rate are 0.3, 0.3 respectively, gene replication rate was 0.95. Fitness function has the penalty of head and flow deficit component part to get solutions.

4.2. Results

Table 1 shows the cost of optimal solutions to Hanoi problem obtained by GEP, compared with former results from other methods. It is interesting to observe that some of the most popular and well performing solutions have a consistent common part out of 34 links. Some of them don't respect completely hydraulic head bounds if the solution simulated with EPANET, which is due to the different value of ω . So the results of Vasan, Cunha and Farmani [2, 10, 11], who produced lower costs for the network, were found to be infeasible, as the pressure constraints were violated.

Table 1. Differences in diameters (in) of optimal solutions obtained for Hanoi network

PIPE	Savic & Walters(GA)	Farmani et al. Self-adaptive fitness formulation	Cunha & Sousa	Bicik	Vasan & Simonovic	Bolognesi et al.	Proposed Method
	(GA)		(SA)	(GANetXL)	(DE)	(GHEST)	(GEP)
1-9	40	40	40	40	40	40	40
10	30	30	30	30	30	30	30
11	24	24	24	24	24	24	24
12	24	24	24	24	24	24	24
13	20	20	20	20	20	20	20
14	16	12	16	16	12	16	16
15-16	12	12	12	12	12	12	12
17	16	16	16	16	16	16	16
18	20	24	20	24	20	24	24
19	20	20	20	20	20	20	20
20	40	40	40	40	40	40	40
21	20	20	20	20	20	20	20
22	12	12	12	12	12	12	12
23	40	40	40	40	40	40	40
24-25	30	30	30	30	30	30	30
26	20	20	20	20	20	20	20
27	12	12	12	12	12	12	12
28	12	12	12	12	12	12	12
29	16	16	16	16	16	16	16
30	16	12	12	12	12	12	12
31	12	12	12	12	12	12	12
32-33	16	16	16	16	16	16	16
34	20	24	24	24	24	24	24
ω	10.5088	10.6668	10.5088	10.6668	10.6668	10.6668	10.6668
Cost (10 ⁶ \$)	6.073	6.069	6.056	6.081	6.056	6.081	6.081
Calls	1,000,000	24,100	53,000	-	50,201	16,600	15,000

For example, the minimum head in DE [14] is 29.59 meters when it gets the result of 6.056 million. And the result of self-adaptive fitness formulation method is infeasible while the solution runs in EPANET for the head constraints. The best solution which is feasible with EPANET ($\omega=10.6668$) to be date found, when pressure constraints of 30 meters are 6.081 million \$.

The final design obtained here is similar to that of Bolognesi and Bicik in that the all links have the same diameters. Table 1 indicates that the algorithms described here found good, if not optimum, solutions for Hanoi network with a little fewer calls of EPANET in evaluation than the GANetXL and GHEST (about 16,600 calls). Although their idea of different penalization of solutions depending on the distance from the boundary of feasible and infeasible solutions is a positive move, the way it has been implemented can't improve the result. This could be due to the use of a constant penalty weight and implementation of penalty without inclusion of head deficiency for infeasible solutions with negative deficiency. Figure 4 shows the performance of GEP with the increasing generation which gets the best solution and average solution fitness. During the evaluation, the method shows unsteady to get the best

solution, as the parameter is so crucial. Nevertheless, the trend, we can see, is significant to get the best solution. The best solution of 6.081 million is gotten about 15,000 calls after 150 generations. Then, the best solution after 15,000 calls is the same as that at 15,000 calls though the average values in each generation, which is convergent during the evolution, are fluctuated sometimes.

In general, the GEP algorithms achieved results. It is shown that the algorithm was able to find the optimum or near-optimum solution with considerably a little less computational effort. The main advantage of the method is that it uses these function set and evolves a little more quickly to the optimum, and may be more robust with more lists of pipes. It simply uses the characteristic of the search space in each generation and doesn't need the initial feasible solution. The ability to find the feasible solution as well as the optimum solution represents a little improvement in method performance.

5. Conclusion

Based on the result obtained in this research, in which GEP was applied to benchmark WDS problems, the GEP is an attractive alternative to GA, SA and DE for the optimal design of WDS. For the Hanoi benchmark, the proposed method found a feasible solution of \$ 6.081 million, which is the least cost.

Moreover, the GEP algorithms found to be a little more computationally efficient compared with that evaluation for the GA, SA and DE. According to the number of evaluations, the proposed method produces the same good results in fewer calls than other precious method. Although an improved optimal design with less number of evaluations was obtained by the proposed method, it is not possible to state categorically that this method will be always given the best solutions as the problem of optimal WDS design is a nonlinear, nonconvex and discrete in nature NP-hard problem.

References

- [1].Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*. 1989, Boston,Mass: Addison-Wesley.
- [2].Cunha, M.D. and J. Sousa, *Water distribution network design optimization: Simulated annealing approach*. Journal of Water Resources Planning and Management-Asce, 1999. **125**(4): p. 215-221.
- [3].Maier, H.R., et al., *Ant colony optimization distribution for design of water systems*. Journal of Water Resources Planning and Management-Asce, 2003. **129**(3): p. 200-209.
- [4].Mohan, S. and K.S.J. Babu, *Optimal Water Distribution Network Design with Honey-Bee Mating Optimization*. Journal of Computing in Civil Engineering, 2010. **24**(1): p. 117-126.
- [5].Ferreira, C., *Gene expression programming: a new adaptive algorithm for solving problems*. Complex Systems, 2001. **13**(2): p. 87-129.
- [6].Guven, A., *Linear genetic programming for time-series modelling of daily flow rate*. Journal of Earth System Science, 2009. **118**(2): p. 137-146.
- [7].Reca, J. and J. Martinez, *Genetic algorithms for the design of looped irrigation water distribution networks*. Water Resources Research, 2006. **42**(5).
- [8].Montesinos, P., A. Garcia-Guzman, and J.L. Ayuso, *Water distribution network optimization using a modified genetic algorithm*. Water Resources Research, 1999. **35**(11): p. 3467-3473.
- [9].Fujiwara, O. and D.B. Khang, *A 2-PHASE DECOMPOSITION METHOD FOR OPTIMAL-DESIGN OF LOOPED WATER DISTRIBUTION NETWORKS*. Water Resources Research, 1990. **26**(4): p. 539-549.
- [10].Vasan, A. and S.P. Simonovic, *Optimization of Water Distribution Network Design Using Differential Evolution*. Journal of Water Resources Planning and Management-Asce, 2010. **136**(2): p. 279-287.
- [11].Farmani, R., et al., *Self-adaptive fitness formulation for evolutionary constrained optimization of water systems*. Journal of Computing in Civil Engineering, 2005. **19**(2): p. 212-216.