

# LEAK DETECTION AND CALIBRATION USING TRANSIENTS AND GENETIC ALGORITHMS

By John P. Vítkovský,<sup>1</sup> Angus R. Simpson,<sup>2</sup> Member, ASCE, and Martin F. Lambert<sup>3</sup>

**ABSTRACT:** Leak detection and calibration of pipe internal roughnesses in a water distribution network are significant issues for water authorities around the world. Computer simulation of water distribution systems to determine the location and size of leaks is emerging as an important tool. A major uncertainty in developing computer models is the condition of the interior of the pipes in the network, especially if they are old. An innovative technique for leak detection and calibration called the inverse transient technique has been recently developed. This paper uses the genetic algorithm (GA) technique in conjunction with the inverse transient method to detect leaks and friction factors in water distribution systems. A continuous variable representation has been developed for the GA coding scheme in this paper. Two new GA operators for crossover and mutation are also introduced. The inverse transient method using the GA technique is effective at finding leakage locations and magnitudes while simultaneously finding the friction factors for different transient data record lengths.

## INTRODUCTION

Large and complex pipe networks for distribution of water to cities cost many millions of dollars to construct, operate, and maintain. The least accurately known parameters in a water distribution system are often the pipe internal roughnesses or friction factors (Task 1987). Over time, the Darcy-Weisbach friction factor of a pipe increases with age due to tuberculation (the buildup of deposits on the pipe wall).

Finding leaks and calibrating for pipe roughnesses simultaneously is possible using inverse methods where the results of measurements are known but parameters (such as friction factors and leaks) of the physical system are unknown. A transient in a pipeline system occurs when a variation in pressure and velocity of the flow is caused by a change, for example, by a valve shutting or opening. Inverse methods require a large amount of data, which can be collected by frequent pressure measurements during a transient event. The inverse transient approach provides an effective procedure for determining leaks and pipe friction factors for a water distribution system by (1) initiating a transient event; (2) measuring the transient pressures at the measurement sites via a telemetry system or a data acquisition system; and (3) by using the inverse transient method (ITM) to determine leak locations and magnitudes of friction factors by minimizing the deviations between measured and numerical model pressures. Transient or water hammer waves propagate through the system for a considerable time following the change.

Liggett and Chen (1994) proposed the inverse transient method as a calibration and leak detection technique. Their method required an optimization technique to fit the measured data to the numerical model results. Liggett and Chen (1994) used the Levenberg-Marquardt method that develops derivatives generated through an adjoint analysis. However, the search space for the problem can be enormous. The Levenberg-Marquardt method is a standard optimization technique, but the solution it finds often depends on the starting point, and it cannot guarantee convergence to a global optimum. An

alternative method to the Levenberg-Marquardt method, or any derivative-based technique, is one based on the genetic algorithm (GA) technique. The GA method does not guarantee reaching the global optimal solution, but comparisons show that it seems more effective than nonlinear programming in finding optimal solutions. The GA appears to search more widely in the solution space compared to other techniques. Details of the GA technique applied to the inverse transient problem are presented in this paper. Results are presented for a case where leaks and friction factors are determined simultaneously by the ITM using GAs. The effect of altering the length of pressure record used is also investigated.

## UNSTEADY FLOW MODELING IN WATER DISTRIBUTION SYSTEMS

Unsteady flow modeling or water hammer analysis is usually implemented by a method of characteristics solution of the governing partial differential equations (Wylie and Streeter 1993). A method of characteristics solution approach is used in this paper. The leakage,  $Q_L$ , is a term that is included in an unsteady equation of conservation of mass at nodes (inflow to a node equals outflow). Leakage from an orifice in the side of a pipe may be simulated using the following equation:

$$Q_L = C_d A_0 \sqrt{2gH_L} \quad (1)$$

where  $C_d$  = orifice discharge coefficient;  $A_0$  = area of leak orifice;  $g$  = gravitational acceleration; and  $H_L$  = hydraulic grade at the leak. Leakage is assumed to occur only at nodal positions.

## GA OPTIMIZATION

GAs mimic the way populations of species genetically evolve to suit their environment over many generations. Using this analogy, a process involving selection, crossover, and mutation can be used to evolve a population of potential solutions of engineering design and analysis problems toward improved solutions. These solutions will satisfy the specified constraints while minimizing or maximizing the objective function. GA optimization has been successfully applied to the optimization of water distribution system design (Murphy and Simpson 1992; Simpson et al. 1994; Dandy et al. 1996). Genetic algorithms have been previously applied to calibration and leak detection (Tang et al. 1999; Vítkovský et al. 1999). Simpson and Goldberg (1994) describe the main steps to the implementation of a GA.

The inverse transient method of Liggett and Chen (1994) involves fitting a numerically modeled pressure or hydraulic grade line (HGL) trace to a measured HGL trace by minimiz-

<sup>1</sup>PhD Student, Dept. of Civ. and Envir. Engrg., Univ. of Adelaide, Adelaide, Australia, 5005. E-mail: jvitkovs@civeng.adelaide.edu.au

<sup>2</sup>Sr. Lect., Dept. of Civ. and Envir. Engrg., Univ. of Adelaide, Adelaide, Australia, 5005.

<sup>3</sup>Sr. Lect., Dept. of Civ. and Envir. Engrg., Univ. of Adelaide, Adelaide, Australia, 5005.

Note. Discussion open until January 1, 2001. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this technical note was submitted for review and possible publication on March 10, 2000. This technical note is part of the *Journal of Water Resources Planning and Management*, Vol. 126, No. 4, July/August, 2000. ©ASCE, ISSN 0733-9496/00/0004-0262-0265/\$8.00 + \$.50 per page. Technical Note No. 21555.

ing the differences between calculated and measured values. The objective of the GA is usually formulated to maximize the fitness of a string, which may be achieved in the following equation by maximizing the negative of the sum of absolute differences:

$$\text{fitness} = - \sum_{i=1}^M |H_i^* - H_i| \quad (2)$$

where  $M$  = total number of data points;  $H_i^*$  = measured HGL, and  $H_i$  = modeled HGL.

### GA Coding Scheme

The Darcy-Weisbach pipe friction factors ( $f$ ) and lumped leak coefficients ( $C_d A_0$ ) may be found from the ITM (Liggett and Chen 1994). Both of these types of decision variables may take on any value within a range of continuous values. A lower and an upper bound of the decision variables need to be specified. Representation of decision variables within a GA string is usually by discrete values within a continuous range. For a discrete coding scheme, a lookup table relating a code (either binary or integer) with a corresponding discrete value (for example,  $f = 0.015, 0.016$ , etc.) is necessary. To suit the continuous nature of the friction factor and lumped leak coefficient values, a continuous coding scheme has been developed for implementation of the GA in this paper. The initial population of strings of unknown friction factors and leaks in the GA process is generated randomly using values across the continuous range between the selected lower and upper bounds. The ranges for the seeding of the friction factors and the lumped leak coefficients were from 0.01 to 0.05 and from 0.0 to 0.001, respectively.

### New Crossover Operator

A new crossover operator referred to as two child staggered average crossover (Vítkovský and Simpson 1997) has been developed to exploit the continuous nature of the variables being represented by the string. This operator is based upon a crossover operator first used by Savic and Walters (1995) named "one child average crossover." The schematic of this new crossover operator with three crossover points is shown in Fig. 1. This operator averages the friction factor values from a pair of strings for corresponding bits in the same positions of the parent strings while the original bits are used at the

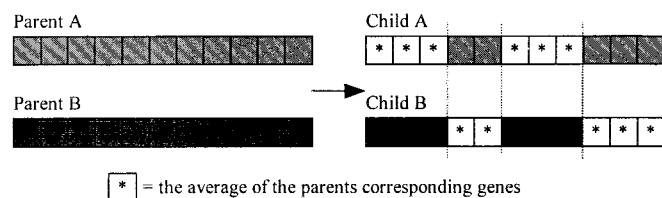


FIG. 1. Two Child Staggered Crossover Example

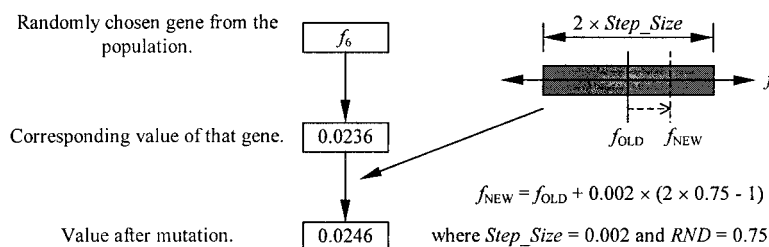


FIG. 2. Continuous Adjacency Mutation Example

other locations. This operator differs from Savic and Walters's scheme in that it maintains some of the original genetic information in each parent string.

### New Mutation Operator

A new mutation operator has also been developed for continuous variables. The mutation operator alters the gene (friction factor or lumped leak coefficient) to a new value within a range of  $Step\_Size$  below or above the current value. Fig. 2 illustrates the mutation operator. The mutated gene value is given by

$$gene_{NEW} = gene_{OLD} + Step\_Size(2 \cdot RND - 1) \quad (3)$$

where  $RND$  = a uniformly distributed random number between 0.0 and 1.0; and  $Step\_Size$  = maximum increment allowed in a gene.  $Step\_Size$  values are different for different types of parameters. For example, the  $Step\_Size$  for a friction factor is about three orders of magnitude larger than that for a lumped leak coefficient. Due to the small size of the lumped leak coefficient and its sensitivity, it was necessary to search for leaks in logarithmic space. This allows strings to move efficiently between differing magnitudes of leaks and to avoid negative lumped leak coefficients.

### CASE STUDY EXAMPLE NETWORK

The example water distribution pipe network considered in this paper is shown in Fig. 3 and consists of 11 pipes and seven nodes. It is supplied from a reservoir and a constant inflow at node 7. The Darcy-Weisbach friction factors and the pipe lengths for the network are shown in Fig. 3 below each pipe. All pipes are 254 mm in diameter with a wall thickness of 1.6 mm. The network is based on a network introduced by Pudar and Liggett (1992).

The transient is initiated in the network by varying the discharge at node 4. The measurement calibration data used in this paper for the example network include the HGL variation at up to six nodes over a 40-s interval (corresponding to 72 data points per node). The example network problem is examined with an actual leak at node 2. Both the friction factors and the lumped leak coefficients are determined simultaneously using the GA search technique. For this case, the simulated leak at node 2 has a lumped leak coefficient of  $C_d A_0 = 1.0 \times 10^{-4}$ , corresponding to an approximate leak hole diameter of 13 mm (the pipe being 254 mm).

Chromosomes have been encoded in the GA for this example network such that the first set of genes in the chromosome correspond to the friction factors and the second set to the leak candidates (Fig. 4). For the example network, possible leak locations are assumed at five nodes (2, 3, 4, 5, and 6). In the GA procedure, an initial population of solutions of leaks and friction factors is randomly generated. A forward transient analysis is run for each string in the population that represents both the leaks and friction factors.

Three groups of GA runs were used to investigate the effect

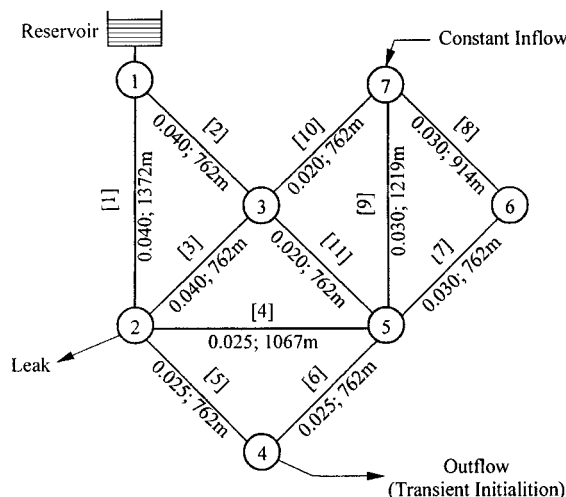


FIG. 3. Layout of Example Network

TABLE 1. Friction Factors Found by GA for Varying Lengths of Pressure Record

Pipe number (1)	Correct $f$ (2)	Friction Factor $f$		
		18 time steps (3)	36 time steps (4)	72 time steps (5)
1	0.040	0.0387	0.0389	0.0410
2	0.040	0.0413	0.0411	0.0393
3	0.040	0.0333	0.0361	0.0388
4	0.025	0.0218	0.0227	0.0253
5	0.025	0.0246	0.0239	0.0259
6	0.025	0.0293	0.0270	0.0247
7	0.030	0.0306	0.0302	0.0283
8	0.030	0.0277	0.0285	0.0268
9	0.030	0.0245	0.0272	0.030
10	0.020	0.0230	0.0212	0.0215
11	0.020	0.0232	0.0238	0.0199
$[E_f]$	—	[10.3%]	[6.99%]	[3.42%]

Note:  $E_f$  = average absolute difference in friction factors.

TABLE 2. Lumped Leak Coefficients Found by GA for Varying Lengths of Pressure Record

Node number (1)	Correct $C_d A_0$ (2)	Lumped Leak Coefficient $C_d A_0$		
		18 time steps (3)	36 time steps (4)	72 time steps (5)
2	$1.0 \times 10^{-4}$	$9.31 \times 10^{-5}$	$9.12 \times 10^{-5}$	$9.95 \times 10^{-5}$
3	0.0	$1.22 \times 10^{-6}$	$8.44 \times 10^{-9}$	$1.19 \times 10^{-7}$
4	0.0	$1.34 \times 10^{-6}$	$2.02 \times 10^{-7}$	$1.33 \times 10^{-7}$
5	0.0	$2.93 \times 10^{-7}$	$3.19 \times 10^{-7}$	$1.89 \times 10^{-7}$
6	0.0	$5.44 \times 10^{-10}$	$3.37 \times 10^{-7}$	$1.83 \times 10^{-7}$
$[E_L]$	—	[6.90%]	[8.79%]	[0.50%]

Note:  $E_L$  = average difference in lumped leak coefficient (only at node 2).

of changing the length of pressure record. The first group uses 18 time steps of pressure data for each measurement node, the second uses 36 time steps, and the third uses 72 time steps, corresponding to simulation times of 10, 20, and 40 s, respectively. All of the runs performed in this section use three measurement nodes (nodes 2, 3, 6).

Results in Tables 1 and 2 show the following trend: the more pressure data used in the analysis, the smaller the average error in the friction factors and lumped leak coefficients. The results in Table 2 show false leaks at nodes with no actual leaks, but the false leaks are approximately three orders of magnitude smaller than the actual leaks at nodes where a leak was present. The effects of these false leaks on the transient results were negligible and correspond to a hole in the pipe of about 0.5 mm in diameter (compared to a pipe diameter of 254 mm).

A comprehensive laboratory experimental testing program to detect leaks and friction factors is under way at the University of Adelaide for verifying the effectiveness of both the inverse transient technique and the GA technique. A single pipeline is currently being tested as part of ongoing research. In addition, a laboratory-based looped pipe network is currently under construction, and test results for detecting leaks in networks should be available in late 2000.

## CONCLUSIONS

The ITM is a powerful technique for leak detection and calibration of friction factors in water distribution systems. Transient techniques have a distinct advantage over traditional steady state calibration techniques in that they provide much more data. The previous implementation of the ITM by Liggett and Chen (1994) used derivative methods of optimization. Such methods may fail to converge or may converge to a local minimum rather than the global minimum. An alternative to derivative methods is a GA search method. The focus of this paper has been to implement the ITM using a GA search method. There is a trade-off between the fast speed of a derivative method of optimization and the slower but more robust and comprehensive GA search. The GA has the ability to bound its search in the parameter space. This is a distinct advantage over gradient-based methods where bounding can cause the optimization method to fail.

Results have been presented in an example case study for the ITM using GAs where both friction factors and leaks have been found. A continuous variable representation for use in the GA has been proposed in this paper. A new crossover operator and a new continuous mutation operator have been introduced. The effect of varying and lengths of record has been investigated.

## ACKNOWLEDGMENTS

The financial assistance for this research from the Australian Research Council and an Australian Postgraduate Award scholarship is gratefully acknowledged.

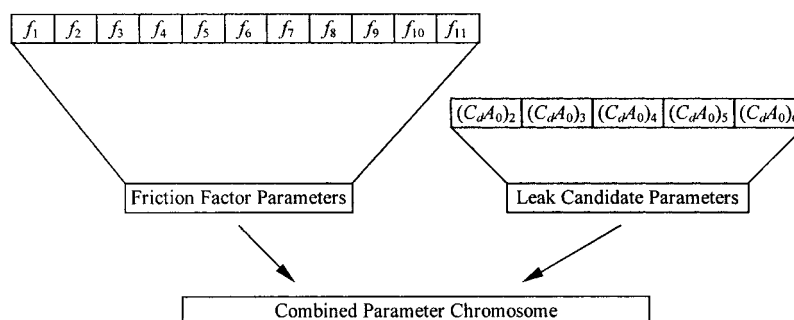


FIG. 4. Chromosome Representation for Example Network for Friction Factors and Leaks

## APPENDIX I. REFERENCES

- Dandy, G. C., Simpson, A. R., and Murphy, L. J. (1996). "An improved genetic algorithm for pipe network optimisation." *Water Resour. Res.*, 32(2), 449–458.
- Liggett, J. A., and Chen, L.-C. (1994). "Inverse transient analysis in pipe networks." *J. Hydr. Engrg.*, ASCE, 120(8), 934–955.
- Murphy, L. J., and Simpson, A. R. (1992). "Pipe optimisation using genetic algorithms." *Res. Rep. No. R93*, Dept. of Civ. Engrg., University of Adelaide, Adelaide, Australia.
- Pudar, R. S., and Liggett, J. A. (1992). "Leaks in pipe networks." *J. Hydr. Engrg.*, ASCE, 118(7), 1031–1046.
- Savic, D. A., and Walters, G. A. (1995). "Genetic algorithm techniques for calibrating network models." *Rep. No. 95/12*, Ctr. for Sys. and Control Engrg., University of Exeter, Devon, U.K.
- Simpson, A. R., Dandy, G. C., and Murphy, L. J. (1994). "Genetic algorithms compared to other techniques for pipe optimization." *J. Water Resour. Plng. and Mgmt.*, ASCE, 120(4), 423–443.
- Simpson, A. R., and Goldberg, D. E. (1994). "Pipeline optimization via genetic algorithms: from theory to practice." *Proc., 2nd Int. Conf. on Water Pipeline Sys.*, Edinburgh, Scotland, 309–320.
- Tang, K., Karney, B., Pendlebury, M., and Zhang, F. (1999). "Inverse transient calibration of water distribution systems using genetic algorithms." *Proc., Conf. on Computing and Control for the Water Industry (CCWI'99)*, Exeter, U.K.
- Task Committee on Water Supply Rehabilitation Systems. (1987). *Water supply system rehabilitation*, T. M. Walski, ed., ASCE, New York.
- Vítkovský, J. P., and Simpson, A. R. (1997). "Calibration and leak detection in pipe networks using inverse transient analysis and genetic

algorithms." *Res. Rep. No. R157*, Dept. of Civ. and Envir. Engrg., University of Adelaide, Adelaide, Australia.

- Vítkovský, J. P., Simpson, A. R., and Lambert, M. F. (1999). "Leak detection and calibration of water distribution systems using transients and genetic algorithms." *Proc., Water Distribution Sys. Mini-Symp.*, Div. of Water Resour. Plng. and Mgmt., ASCE, Reston, Va.
- Wylie, E. B., and Streeter, V. L. (1993). *Fluid transients in systems*, Prentice-Hall, Englewood Cliffs, N.J.

## APPENDIX II. NOTATION

*The following symbols are used in this paper:*

- $A$  = cross-sectional pipe area;  
 $C_d A_0$  = lumped leak coefficient;  
 $E_F$  = average absolute difference in friction factors (%);  
 $E_L$  = average difference in lumped leak coefficient (at leak) (%);  
 $f$  = Darcy-Weisbach friction factor;  
 $g$  = gravitational acceleration;  
 $gene$  = individual gene in genetic algorithm string;  
 $H$  = hydraulic grade line (HGL);  
 $H_L$  = hydraulic grade line at leak;  
 $H^*$  = measured hydraulic grade line;  
 $M$  = total number of data points;  
 $Q_L$  = nodal leak flow;  
 $RND$  = randomly generated number; and  
 $Step\_Size$  = continuous mutation step size.