Optimal reconstruction of historical water supply to a distribution system: B. Applications

M. M. Aral, J. Guan, M. L. Maslia, J. B. Sautner, R. E. Gillig, J. J. Reyes and R. C. Williams

ABSTRACT

In a recently completed case-control epidemiological study, the New Jersey Department of Health and Senior Services (NJDHSS) with support from the Agency for Toxic Substances and Disease Registry (ATSDR) documented an association between prenatal exposure to a specific contaminated community water source and leukaemia in female children. An important and necessary step in the epidemiological study was the reconstruction of the historical water supply strategy of the water distribution system serving the Dover Township area, New Jersey. The sensitivity of solutions to: (1) pressure and pattern factor constraints, (2) allowable operational extremes of water levels in the storage tanks, and (3) the non-uniqueness of the water supply solution are analysed in detail. The computational results show that the proposed approach yields satisfactory results for the complete set of monthly simulations and sensitivity analyses, providing a consistent approach for identifying the historical water supply strategy of the water distribution system. Sensitivity analyses indicated that the alternative strategy obtained from the revised objective function and the variation of constraints did not yield significantly different water supply characteristics. The overall analysis demonstrates that the progressive optimality genetic algorithm (POGA) developed to solve the optimization problem is an effective and efficient algorithm for the reconstruction of water supply strategies in water distribution systems.

Key words | exposure-dose reconstruction, genetic algorithm, water distribution system

M. M. Aral (corresponding author)
J. Guan
Multimedia Environmental Simulations Laboratory,
School of Civil and Environmental Engineering,
Georgia Institute of Technology,
Atlanta,
Georgia 30332,
USA
Tel.: +404-894-2243
E-mail: *maral@ce.gatech.edu*

M. L. Maslia
J. B. Sautner
R. E. Gillig
J. J. Reyes
R. C. Williams
Division of Health Assessment and Consultation,
Agency for Toxic Substances and Disease Registry,
Atlanta,
Georgia 30333,
USA

INTRODUCTION

In a recently completed case-control epidemiological study (NJDHSS 2003), the New Jersey Department of Health and Senior Services (NJDHSS) documented a statistically significant association and consistency in multiple measures of association between prenatal exposure to time-specific contaminated community water source (1982–1996) and leukaemia in female children of all ages (odds ratio (OR) of 5.0 and 95% confidence interval (CI) of 0.8–3.1). An important and necessary step in the epidemiological study was the reconstruction of the historical water supply strategy of the water distribution system serving the Dover Township area, New Jersey. The models and solution algorithm (POGA) used to reconstruct the historical water supply to the distribution system are described in an accompanying paper by the authors (see Part A, this issue). The analyses were based on extensive collections and aggregation of data derived from historical records of the water distribution system and on field data collected for the present-day (1998) water distribution system (Maslia *et al.* 2000*a*, *b*, 2001).

As part of the epidemiological study, exposure indices were calculated. The indices were derived by determining the percentage of water that study subjects may have received from each point of entry (i.e. wells and well fields) to the water distribution system. The percentage or proportionate contribution of water thus became a surrogate for exposure pathways and exposure intervals. This approach allowed epidemiologists to assess more accurately the association between the occurrence of childhood cancers and exposure to each of the sources of potable water entering the distribution system, including those known to have been historically contaminated. Complete details pertaining to the historical reconstruction of the water supply to the water distribution system, including historical databases, are provided in Maslia *et al*. (2001).

Using the formulation and the solution algorithm described in Part A of this paper, (this issue), water supply strategies for the distribution system were obtained for each month of the study period (January 1962–December 1996), as required by the epidemiological study protocol (NJDHSS 2003). Examples of results obtained using the progressive optimality genetic algorithm (POGA) solution strategy presented in this paper demonstrate the effectiveness of the proposed solution method. Results obtained from two different optimization models are also analysed to evaluate the effect of the non-uniqueness of the water supply strategies on the proportionate contribution of water to locations serviced by the distribution system. The sensitivity of the solution to constraints of the problem is also investigated through extensive sensitivity analyses. The results of these analyses are discussed in this section of the paper.

BACKGROUND

Water distribution system configuration

The configuration of the water distribution system serving the Dover Township area, such as the number of pipelines, wells, storage tanks and high-service and booster pumps (hereafter referred to as 'pumps'), changed each year during the historical period. For example, the 1962 water distribution system served nearly 4,300 customers from a population of about 17,200 persons (Board of Public Utilities, State of New Jersey 1962) and was characterized for modelling by (Part A, Figure 1):

- approximately 2,400 pipe segments ranging in diameter from 2 to 12 inches (5 cm to 30 cm) and comprising a total service length of 77 miles (124 km);
- three groundwater extraction wells (two well fields) with a rated capacity of 1,900 gal min^{-1} (7,192 l min^{-1});
- one elevated storage tank and standpipe with a combined rated storage capacity of 0.45 Mgal (1.70 Ml); and
- total annual production of 359 Mgal (1,359 Ml), which included the production of about 1.3 MGD (4.9 Ml day^{-1}) during the peak production month of May.

In contrast, in 1996 – the last year of the historical reconstruction period – the water distribution system served nearly 44,000 customers from a population of about 89,300 persons (Board of Public Utilities, State of New Jersey 1996) and was characterized for modelling by (Part A, Figure 2):

- more than 16,000 pipe segments ranging in diameter from 2 to 16 inches (5 cm to 40 cm) and comprising a total service length of 482 miles (776 km);
- 20 groundwater extraction wells (eight well fields) with a rated capacity of 16,550 gal min^{-1} (62,642 Ml min^{-1});
- 12 pumps;
- three elevated and six ground-level storage tanks with a combined rated storage capacity of 7.35 Mgal (27.82 Ml); and
- total annual production of 3,783 Mgal (14,319 Ml), which included the production of about 13.9 MGD (52.6 Ml day^{-1}) during the peak production month of June.

Details pertaining to the configuration of the water distribution system serving the Dover Township area, throughout the historical period (1962–1996), are provided in numerous tables and appendices in Maslia *et al*. (2001).

Hydraulic modelling

The bases of comparison for all sensitivity analysis results (obtained by applying the POGA method) were the corresponding results obtained through the manual adjustment of water supply well pattern factors (hereafter referred to as the 'manual adjustment process'). To conduct the initial (or base) historical simulations, model parameter values input to EPANET2 (Rossman 2000) required variation that reflected the change in the historical data. For example, data documenting the installation year of network pipelines were available on an annual basis (Maslia *et al*. 2001, Appendix A), and thus model parameters describing the pipeline network were modified in EPANET2 simulations on an annual basis. Data documenting water production were available on a monthly basis (Maslia *et al*. 2001, Appendix B), and thus, EPANET2 model parameters associated with production were varied for each month of the historical period simulations. Data for other model parameters such as the on-and-off cycling of wells were not available throughout the entire historical period. For these data, quantitative estimation and qualitative description methods were used to derive values required to conduct the EPANET2 simulations. A summary of model parameters, data availability and time-unit variation required to conduct the historical reconstruction simulations is provided in Table 1.

Representation of wells, storage tanks and pumps

Two methods exist for supplying water to the water distribution system serving the Dover Township area. In the first method, groundwater wells were operated to supply demand by discharging water directly into the distribution system (e.g. wells 13–15, Figure 1, Part A). In 1968 pumps were added to the distribution system. From that year forward, some wells supplied storage tanks (e.g. Holly Plant ground-level, Figure 2, Part A), then the pumps were operated based upon some predetermined operating schedule to meet distribution-system demands (e.g. wells 21–30, 37 and 39–42, Figure 2, Part A); other wells still discharged directly into the distribution system (e.g. wells 31–35 and 38, Figure 2, Part A). The representation of this latter method of supplying water to the distribution system is shown in Figure 1a. This was the method used to represent the supply of water during simulation and calibration of the present-day (1998) water distribution system (Maslia *et al*. 2000*a*, *b*), and hereafter is referred to as the 'well-storage tank-pump' or WSTP simulation method. Using this method to simulate the supply of water to the distribution system requires the following information:

- known operating schedules for water supply well on-and-off cycling;
- observed storage tank water-level variations;
- realistic pump-characterstic curves; and
- known operating schedules for the on-and-off cycling of pumps.

Because data describing this information were available for the present-day system, simulation and calibration of the 1998 water distribution system was accomplished by using the WSTP simulation method (Maslia *et al*. 2000*a*, *b*).

Hourly operations data for the historical water distribution systems were limited and, for most of the systems, such data were not available (Table 1). Additionally, it was the historical supply of water distributed to pipeline locations that was of importance to the epidemiological study, rather than the specific hourly operation of the WSTP combination. In order to simplify the simulation of the WSTP combination and, thus, reduce data requirements for simulation, a method of idealizing the WSTP combination was developed, designated the 'supply-node-link' or SNL simulation method (Figure 1b). The SNL method eliminated the need for including the storage tank and pump combinations in the historical simulations.

To replace the WSTP method with the SNL method using EPANET2, the WSTP system was idealized as shown in Figure 1b. Ideally, if measured hourly data for the pumps were available for the historical water distribution systems, the total flow in surrogate link K over a 24-hour period (Figure 1b) would be equal to the total flow through link J over a 24-hour period from pumps P_1 and P_2 (Figure 1a). Accordingly, flow supplied to the distribution system by the supply nodes to meet demand (S_1, S_2 and S_3 in this example, Figure 1b) should be equal to the flow that would have been supplied by the pumps (P_1 and P_2 shown in Figure 1a). Using the SNL simulation method, the hourly variation in supply over a 24-hour period was simulated in EPANET2 by adjusting the pattern factors, previously described in Part A of this paper (this issue). Owing to brevity and space limitations, specific details describing the derivation of the SNL

Table 1 | Summary of model parameters, data availability and time-unit variation for historical reconstruction analysis, Dover Township area, New Jersey (from Maslia *et al*. 2001)

Model parameters	Data availability	Time unit variation for historical reconstruction analysis	Notes
Network pipeline data	1962–96[1]	Annual	Assumed operational date of 1 January for in service year
Hydraulic device in service date	1962–96[1,2]	Annual	Assumed operational date of 1 June for in service year
Pipe roughness coefficient	1998	No variation	Maslia *et al*. (2000*a*)
Pipe diameter values	1998	No variation	Maslia *et al*. (2000*a*)
Pump characteristic data	1998	No variation	Maslia *et al*. (2000*a*)
System production data	1962–96[2]	Monthly	Maslia *et al*. (2001, Appendix B)
Point demand (node) values	October 1997–April 1998	Monthly	Maslia *et al*. (2000*a*, 2001)
Pattern factors (system operations)[3], 1962–77	None	Hourly	Maslia *et al*. (2001, Table 12)
Pattern factors (system operations)[3], 1978–87	Typical peak day (summer) and non-peak day (fall) for selected years[4]	Hourly	Maslia *et al*. (2001, Table 12)
Pattern factors (system operations)[3], 1988–96	Typical peak day (summer) and non-peak day (fall) for selected years; 1996, and March and August 1998[4,5]	Hourly	Maslia *et al*. (2000*a*, 2001, Table 12)
Nodal concentrations or percentage contribution of water from specified source	March and April 1996 barium sample collection and transport simulation[5]	24-hour average	Simulated, 24-hour average of percentage contribution of water to model node from water source point of entry (well or well field)

[1]Data from George J. Flegal, United Water Toms River, Inc., written communication to Morris L. Maslia, 25 February 1997.
[2]Data from annual reports of the Board of Public Utilities, State of New Jersey (1962–96).
[3]Model parameters include supply well on-and-off cycling schedules simulated by using pattern factors in EPANET2 and starting water levels in storage tanks.
[4]Data from Richard Ottens, Jr, Production Manager, United Water Toms River, Inc., written communication to Morris L. Maslia, 1998.
[5]Refer to Maslia *et al*. (2000*a*).

simulation method are not provided herein, but can be found in Maslia *et al*. (2001). However, to demonstrate that the idealized SNL simulation method supplies the distribution system with an equivalent amount of water when compared with the 'real-world' WSTP simulation method, both simulation methods were applied to the present-day (1998) water distribution system for conditions that existed in August 1998. Measured and simulated pump flows – using the WSTP simulation method – are compared with simulated flows for the SNL method for the Holly and Parkway facilities (Figure 2).

The results obtained using both the WSTP and the SNL methods produce nearly identical simulated flows. Total simulated supply to the distribution system from the

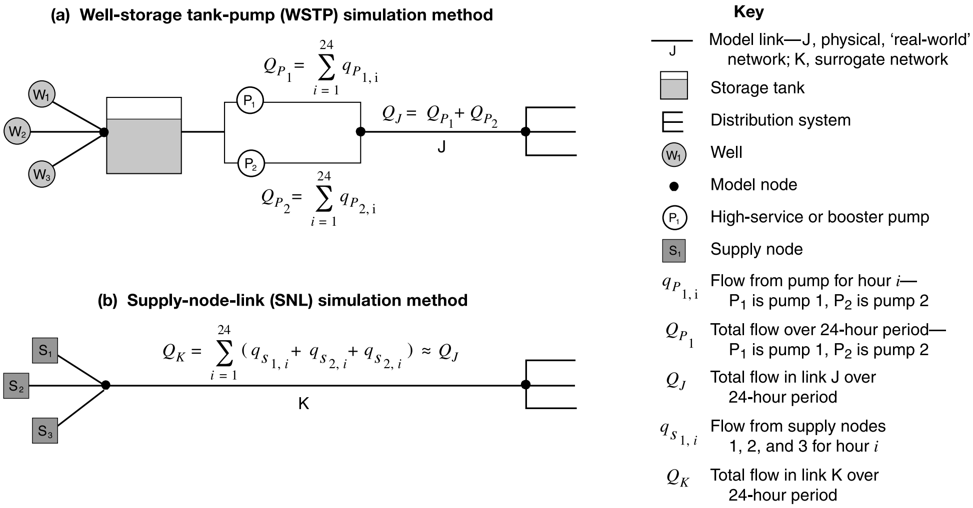

Figure 1 | Distribution system representation of supply well, storage tank and pump combination for (a) physical 'real-world' network and (b) model network used for historical reconstruction analysis (from Maslia *et al*. 2001).

Holly facility over a 48-hour period using the SNL method was 5.62 Mgal (21.27 Ml), which is nearly identical to the measured supply of 5.63 Mgal (21.31 Ml). For the Parkway facility, simulated supply using the SNL method was 8.53 Mgal (32.29 Ml) whereas the measured supply was 8.32 Mgal (31.49 Ml). The difference between the two results is less than 3%. These results confirm the appropriateness of representing the 'real-world' WSTP distribution system (Figure 1a) with the surrogate SNL distribution system (Figure 1b). Thus, the Holly, Parkway and Windsor facilities (Figure 2, Part A) were represented in the historical water distribution system simulations using the SNL method.

Water quality modelling

EPANET2 has the ability to compute the percentage of water reaching any point in the distribution system over time from a specified location (source) in the network: the 'proportionate contribution' of water from a specified source. To estimate the proportionate contribution of water, a source location is assigned a value of 100 units. The resulting solution provided by the water quality simulator in EPANET2 then becomes the percentage of water supplied at any location in the distribution system contributed by the source location of interest.

For the historical reconstruction analyses, a source trace analysis was conducted for every month of the historical period. The list of source node identifications assigned to points of entry for the source trace analyses (manually adjusted pattern factor simulations and the POGA simulations) is provided in Maslia *et al*. (2001, Appendix F). These source nodes were assigned a value of 100 units in order to estimate the percentage of water supplied to locations in the historical water distribution system. Because initial conditions must be 'flushed out' of the distribution system before retrieving the proportionate contribution results, the monthly historical net work models were run for simulation periods of approximately 1,200 hours to reach a state of stationary water quality dynamics ('dynamic equilibrium'). The results of the source trace analyses reported herein represent the last 24 hours of the 1,200 hours of the simulation period. In these runs, hydraulic time steps of 1 hour and water quality time steps of 5 minutes were used. For some monthly simulations in the 1980s, the water quality time steps were reduced to 1 minute. These smaller water quality time steps were necessary to ensure that the mass balance summed to 100%.

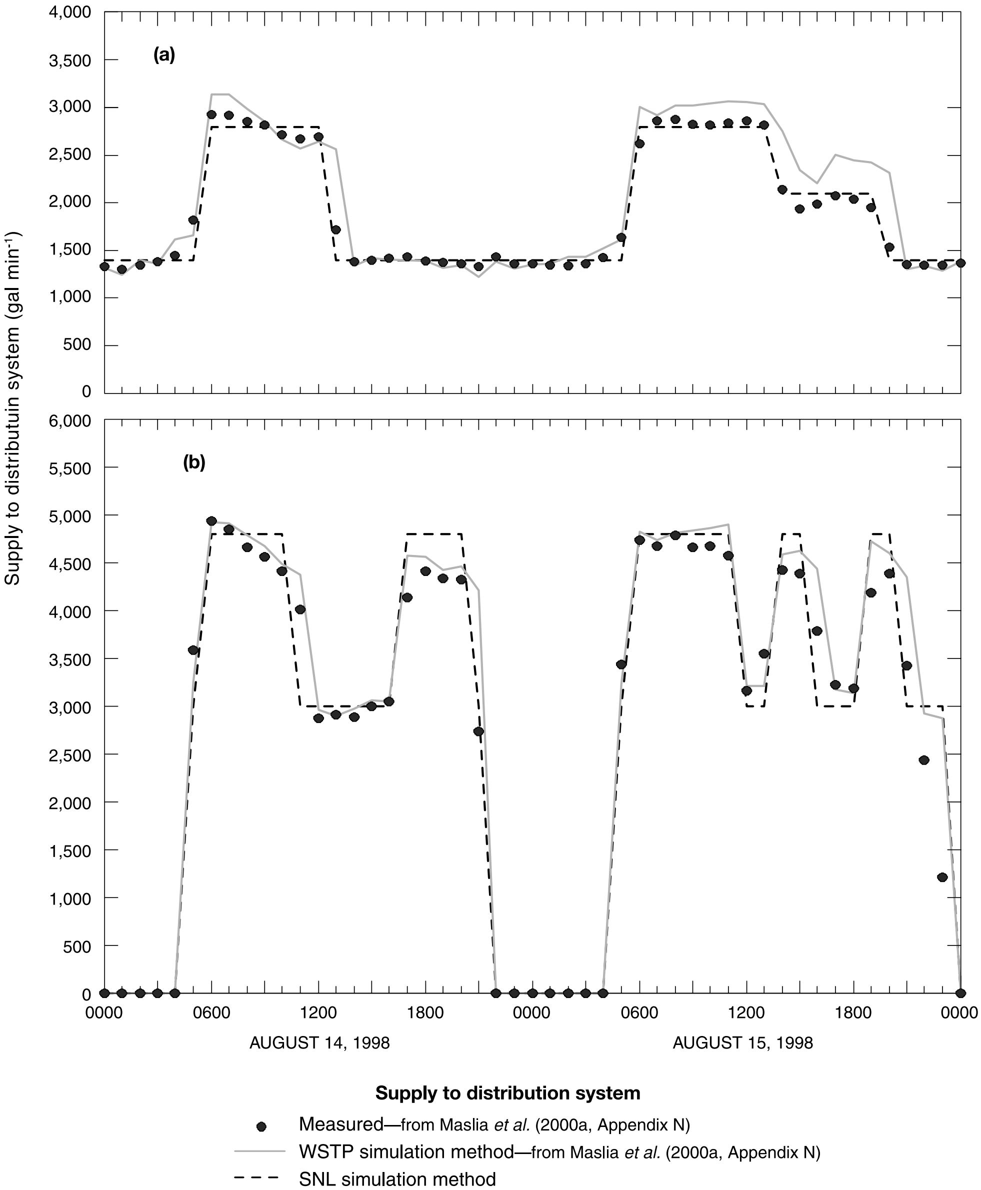

Figure 2 | Measured and simulated flows using the well-storage tank-pump (WSTP) and supply-node-link (SNL) simulation methods, Dover Township area, New Jersey, August 1998 at (a) Holly Plant facility and (b) Parkway facility (from Maslia *et al*. 2001).

HISTORICAL RECONSTRUCTION ANALYSIS USING MANUALLY ADJUSTED PATTERN FACTORS (BASE CONDITIONS)

Through an extensive field investigation and data collection effort, initial EPANET2 input files were developed for each month of every year of the study period (Maslia *et al.* 2001). In these input files, the water supply strategy for system operations was developed by manually adjusting pattern factors of the supply wells and surrogate supply nodes. The specific pressure and storage tank water level constraints used in these data files were also recorded and presented in a text file to be used as constraints in the optimization algorithm described below. One example of the simulation results obtained by the manual adjustment process is presented below. Additional results and analyses of the results are discussed in depth in Maslia *et al.* (2001).

The percentage of water contributed by every well and well field for any given time, can be viewed at selected pipeline locations using a 'stacked' column graph. This method of presentation uses one column to represent each of five selected pipeline locations: A through E (pipeline locations A–E are shown in Figure 2, Part A). The contribution of water, in percentage, from each operating well or well field for the time of interest is 'stacked' one on top of the other within each column. Figure 3 is an example of simulation results using this method of presentation for the maximum demand month of July 1988. Inspection of the graph in Figure 3 indicates that residences at location A would have received their water supply from just two well fields, Holly and South Toms River, contributing 40% and 60% of the supply, respectively. Residences at location E would have received their water supply from a combination of six wells and well fields with the contribution of supply varying from 2% to 55%.

The sum of the proportionate contribution of water from all wells and well fields to any pipeline location should be 100%. Because of numerical approximation and round-off, however, the total contribution from all wells and well fields may sum to slightly less or slightly more than 100% at some locations. Such results are expected when using numerical simulation techniques. In the historical reconstruction analysis conducted for the water distribution system serving the Dover Township area, the sum of the proportionate contribution results at any location, for all historical simulations, ranges from 98% to 101%.

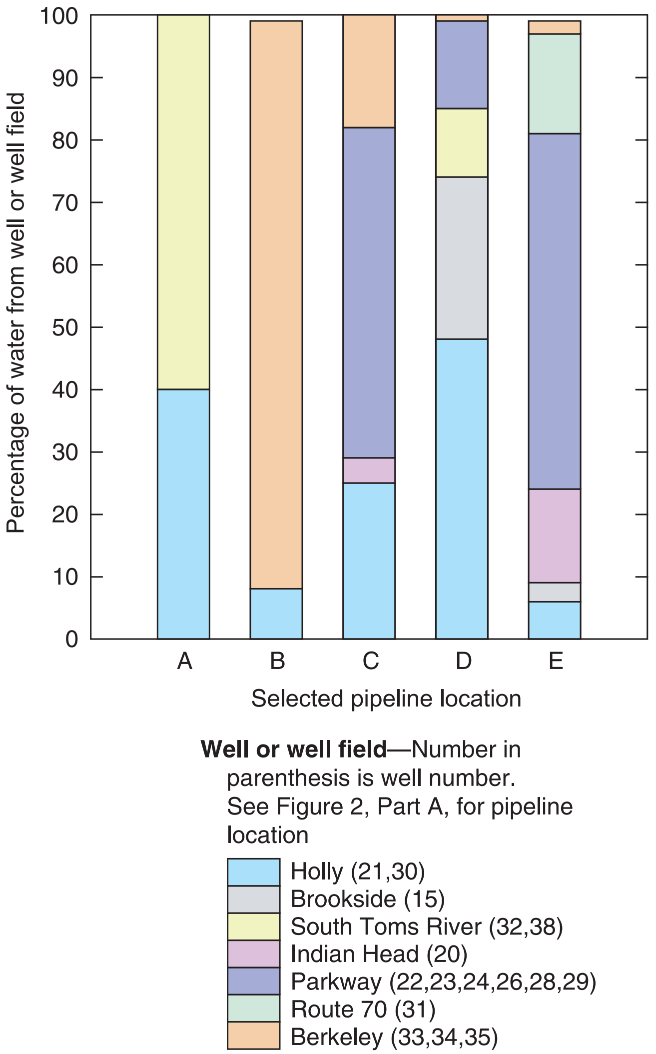

Figure 3 | Simulated proportionate contribution of water from wells and well fields to selected locations using the manually adjusted supply-well and supply-node pattern factors, Dover Township area, New Jersey, July 1988 (from Maslia *et al.* 2001).

HISTORICAL RECONSTRUCTION ANALYSIS USING POGA

Simulations using POGA

To evaluate the models and the algorithm proposed in Part A of this paper(this issue), the integrated system of POGA and EPANET2 was applied to the water distribution system serving the Dover Township area to reconstruct the historical monthly water supply strategy. After the application of POGA and EPANET2 to the monthly data, an optimized EPANET2 input data file (the EPANET2 '.inp' file) was generated for the purpose of subsequent analyses. The optimized EPANET2 input data files represent the optimal water supply strategy for each month. The optimized water supply strategy results in much smaller objective function values when compared with the objective function value of the initial water supply strategy developed using the manual adjustment process. In this paper, due to space limitations, we will discuss results of optimal patterns and the alternative patterns generated from the application of POGA, for two months (May 1962 and October 1996) as two typical cases. We also present selected results from sensitivity analyses. The numerical results obtained for the complete set of historical simulations and comprehensive sensitivity analyses are published in the final report (Maslia *et al*. 2001).

Parameter setting

The EPANET2 and POGA integrated system requires two categories of parameters: water distribution system parameters and POGA parameters. According to the pressure requirements for the Dover Township water distribution system, the minimum pressure limit, $P_{\min}$, was selected as 15 pounds per square inch (psi) (10 N cm^{-2}), and the maximum pressure limitation, $P_{\max}$, was selected as 110 psi (75.8 N cm^{-2}). The penalty coefficient for water level violation in a tank in Equation (9) (see Part A of this paper, this issue), a_T, was selected to be 1.0 and the penalty coefficient for the initial water level violation in a tank in Equation (11) (see Part A of this paper, this issue), a_{TI}, is selected to be 5,000. The selection of these penalty coefficients may affect the optimization results because the objective function is dependent on these coefficients. The larger the penalty coefficients a_T and a_{TI} are, the more the water level violation in the tanks weighs in the objective function. Other GA parameters used in POGA are listed in Table 2.

Table 2 | Parameters used in the progressive optimality genetic algorithm (POGA)

Parameter	Value
Population size	5
String length for each variable	10
Maximum generation	2
Crossover ratio	0.9
Direct selection ratio	0.1
Mutation ratio	0.1
a_T; a_{TI}	1; 5,000
Beta (β)	100
Delta (δ)	0

Optimal patterns

The optimal patterns are calculated using two data sets. The first data set represents the water distribution system serving the Dover Township in May 1962, which contained 2,272 junctions, 2,403 pipe links, one storage tank and three water supply wells (Figure 1, Part A). The optimization calculation takes four iterations. The initial and optimal operation patterns of water supply wells are shown in Figure 4. The optimal patterns are obtained through slight adjustment based on the initial patterns using POGA. However, junction pressure violation cost is largely reduced. The relationship between objective function, time step and iterations is shown in Figure 5. In this case, the tank violation cost is very small, and initially, junction violation cost is 3,436. The first iteration reduces junction violation cost from 3,436.8 to 470.25. Consecutively, these values are reduced to 61.19, 8.91 and finally to 0.21. The objective function is convergent and monotonically reduces throughout the computation. The results imply that the water pressure of each junction in the system would stay in the allowed water pressure range (15–110 psi) (10–75.8 N cm^{-2}) if the optimized water supply strategy were employed in the system operation.

The water pressure as a function of time in a 1-day period for junction #3,399 and #4,627 clearly shows that the optimized dataset is much better than the initial dataset in terms of water pressure violation as shown in Figures 6 and 7. Using the initial patterns, the water pressure at junction #3,399 is lower than the lower pressure limit (15 psi, 10 N cm^{-2}) from 0800 to 1200 hours and from 2200 to 2400 hours. However, using the optimal patterns, the water pressure at that junction is always higher than the lower pressure limit (15 psi, 10 N cm^{-2}).

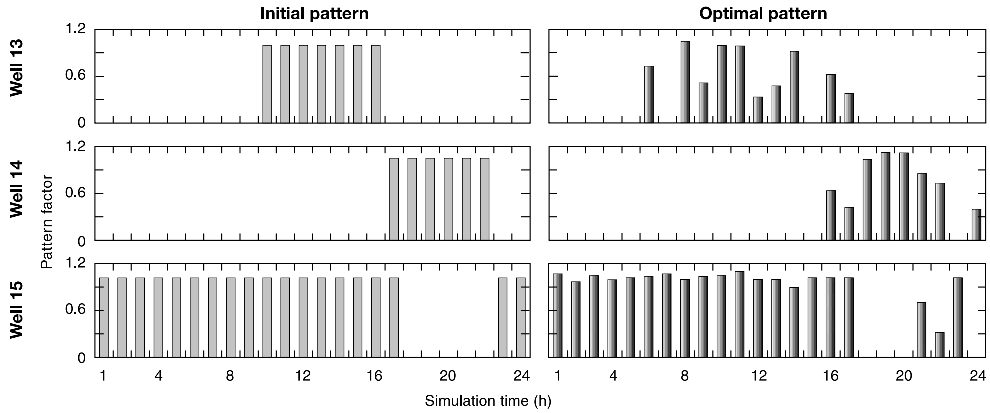

Figure 4 | Initial and optimal operation patterns of water supply wells, May 1962 (see Figure 1, Part A for well locations) (from Maslia *et al*. 2001).

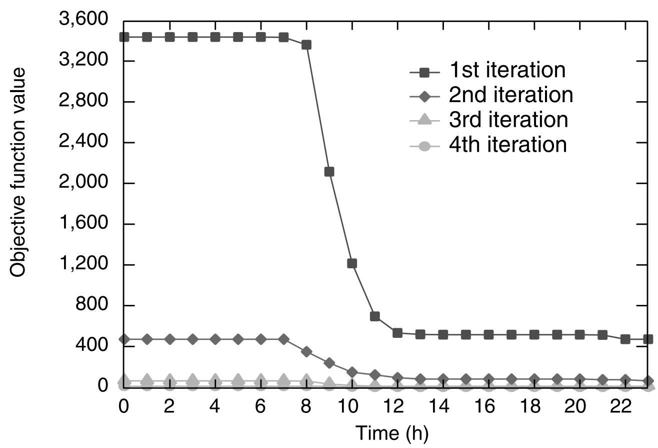

Figure 5 | Convergence of objective function as function of iteration number and time step, May 1962 (from Maslia *et al*. 2001).

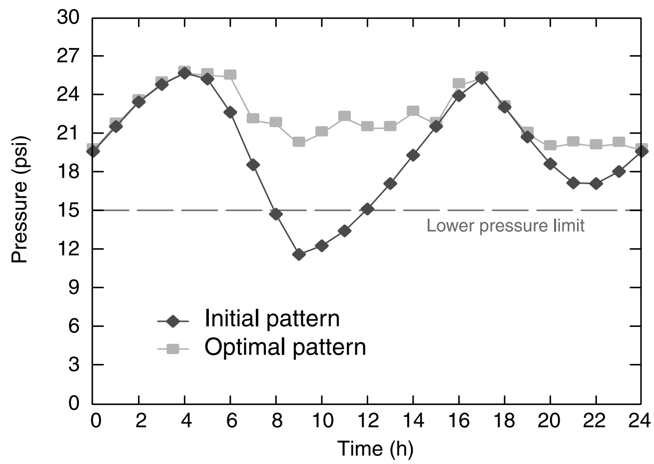

Figure 6 | Water pressure at junction #3,399 using two different patterns, May 1962 (from Maslia *et al*. 2001).

Similarly, the water pressure at junction #4,627 is lower than 15 psi (10 N cm^{-2}) from 0600 to 1400 hours using the initial patterns, while the water pressure is higher than 15 psi (10 N cm^{-2}) all the time using the optimal patterns. For both the initial and the optimal patterns, the water level in the Horton Street tank and standpipe stays within its lower level and upper level limits at all times, and the water level in the tank at the beginning of the day equals that at the end of the day for this case as shown in Figure 8. However, the tank operates with a higher water level using the optimal patterns than that using the initial patterns.

The second dataset represents the water distribution system serving the Dover Township area in October 1996. This network shows significant development and now contains 14,965 junctions, 16,048 pipe links, nine storage tanks and 20 water supply wells (Figure 2, Part A). For this case, the optimization calculation also takes four iterations. The initial and optimal operation patterns of three of the water supply wells are shown in Figure 9. Because there is a large decision space in this case, the optimal patterns have significantly changed when compared with the initial patterns. In this case, junction violation, tank violation costs and objective value are initially 42,728.6, 1,585.72 and 44,314.3, respectively. Through the

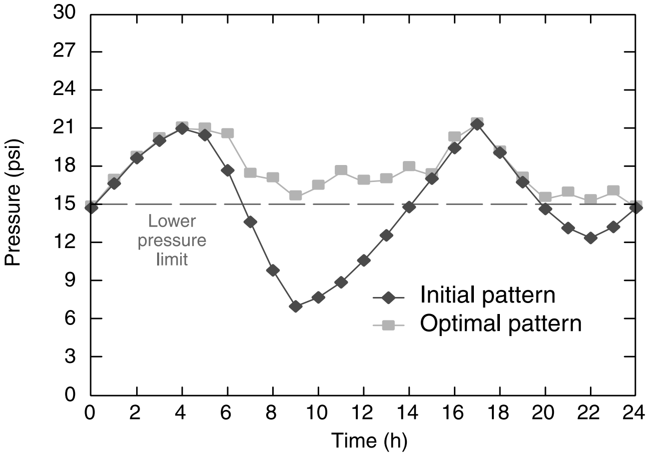

Figure 7 | Water pressure at junction #4,627 using two different patterns, May 1962 (see Figure 1, Part A for location) (from Maslia *et al*. 2001).

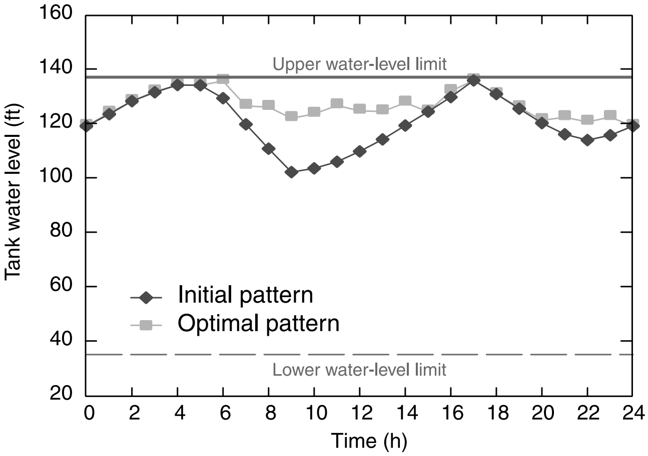

Figure 8 | Water level at Horton Street tank and standpipe using two different patterns, May 1962 (see Figure 1, Part A for locations) (from Maslia *et al*. 2001).

optimization calculation of four iterations, they are reduced to 2,910.2, 0.98 and 2,911.2, respectively.

Figure 10a, b and c shows the convergence of the objective function, junction violation cost and tank violation cost as a function of time step and iterations. Although the tank violation cost has some fluctuation during the iteration, the objective function is monotonically reduced. In this case, the water pressure violations in many junctions are very large when the initial patterns are used. The water pressure violations are significantly improved although the violation still exists in some junctions using the optimal patterns, as shown in Figures 11 and 12. These violation junctions are at the suction sides of the pumps and not at demand nodes, and are expected to occur. The water pressure at junction #55,024 is much higher than the upper pressure limit (110 psi, 75.8 N cm^{-2}) from 0800 to 1000 hours and at a maximum reaches 240 psi (165.5 N cm^{-2}) using the initial patterns. This situation should be prohibited in real operation of the water system distribution. In contrast, the water pressure at the junction stays within the lower pressure limit and upper pressure limits at all times using the optimal patterns (Figure 11). A similar result can be seen from Figure 12 for junction #55,022.

The water level in the tank is within its lower level and upper level limits at all times for both the initial and the optimal patterns, which is enforced by the EPANET2 water distribution system model as shown in Figures 13 and 14. However, the initial water level in a tank at the beginning of the day may not be equal to the water level at the end of the day. The water levels in the Indian Hill tank at the beginning and the end of the day are 196.00 ft (59.74 m) and 195.92 ft (59.72 m), respectively, resulting in a difference of 0.08 ft (0.02 m) using the initial patterns, reducing to 0.01 ft (0.003 m) using the optimal pattern (Figure 14). For the North Dover tank, the difference reduces from 0.48 ft (0.15 m) in the initial patterns to 0.01 ft (0.003 m) in the optimal patterns (Figure 13). In the reconstruction simulation, the optimal patterns selected will be continuously operated for 1,200 h. Even though the difference in water levels in a tank between the beginning and end of the day is small, the long-term simulations may result in a large difference in the water level in the tank between the beginning and the end of the simulation period. This may affect the flexibility of the patterns, and create negative pressures at junctions and cause water delivery problems. This is the reason why the penalty coefficient for this term is selected to be large compared with other terms in the objective function.

Selection of alternative patterns

As described in Part A of this paper (this issue), the reconstructed water supply strategy to operate the water

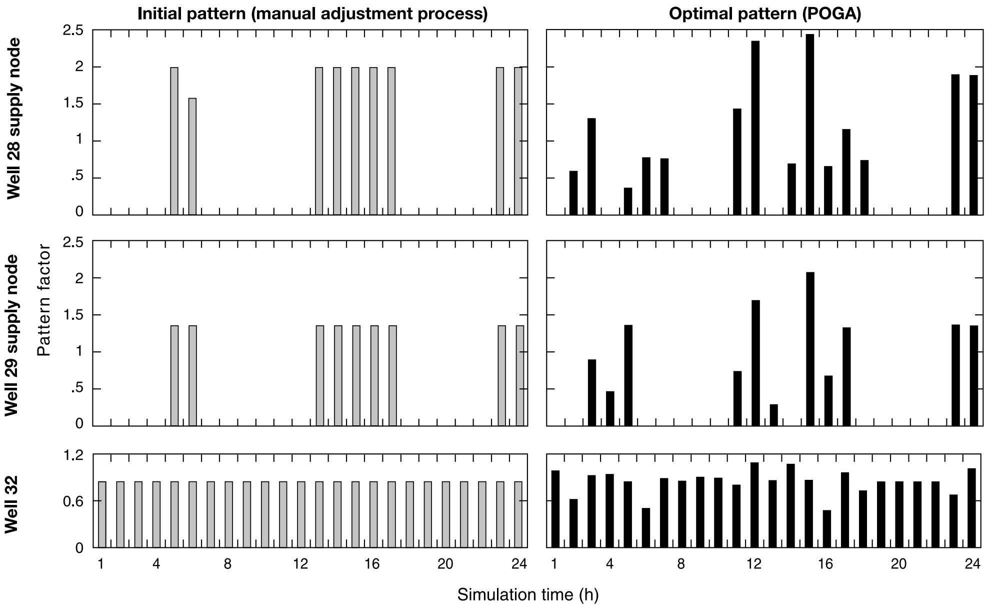

Figure 9 | Initial and optimal operation patterns of water supply wells, October 1996 (see Figure 2, Part A for locations) (from Maslia *et al*. 2001).

distribution system may not be unique. The dataset for October 1996 is chosen as an example to address the effect of non-uniqueness of solutions on the supply of water from the distribution system. The model in Equation (19) (Part A of this paper, this issue) and POGA algorithm are employed with the optimal patterns obtained from Equation (15) (Part A of this paper, this issue) as the initial solution to select new patterns. The new patterns should have an objective function value calculated from Equation (14) (Part A of this paper, this issue) that would be near the objective function value of the first solution (optimal patterns) and it must produce a different pattern schedule when compared with the original optimal pattern. Obviously, the second solution will not be optimal when compared with the first solution, but the pattern schedule will be as different as possible while satisfying the constraints. In this optimization, the maximum iteration is selected as eight considering the characteristics of the model in Equation (19). The resulting patterns for this model are shown in Figure 15 for three water supply wells. For some of the simulation hours, the new patterns can have significant differences when compared with the original optimal patterns. Other computational characteristics such as the objective value and differences in pressure and storage tank water level constraints between the original optimal patterns and the alternative patterns are discussed in detail in Aral *et al*. (2001) and Maslia *et al*. (2001, Appendix E). The effect of the new patterns on the supply of water (i.e. the sensitivity of the water supply strategy to the change in patterns) is discussed below in the section on sensitivity analysis.

Computational efficiency

In POGA, the computation cost consists of two parts. First is the computational cost of calculation of the objective function and second is the calculation of GA operations, such as crossover, mutation and comparison of fitness values. The calculation of objective function accounts for more than 95% of total computational cost so that the effect of the second calculation cost can be neglected. The computational cost for calculation of the objective function depends on the scale of the water distribution system. For example, the computational cost for each calculation of the objective function needs about 2 seconds for the system in 1962, and about 10 seconds for the system in 1996 on an IBM/PC Pentium system. If we define t as the

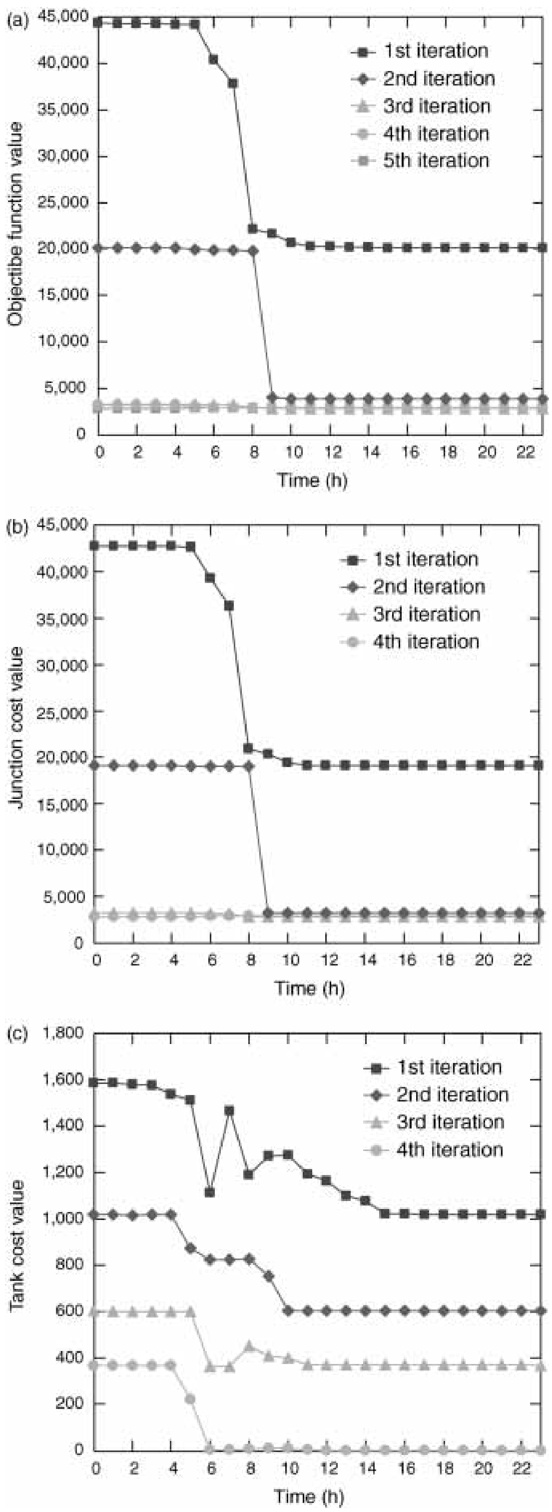

Figure 10 | Convergence of (a) objective function as a function of iteration number and time step, (b) junction cost as a function of iteration number and time step, (c) tank cost as function of iteration number and time step, October 1996 (from Maslia *et al*. 2001).

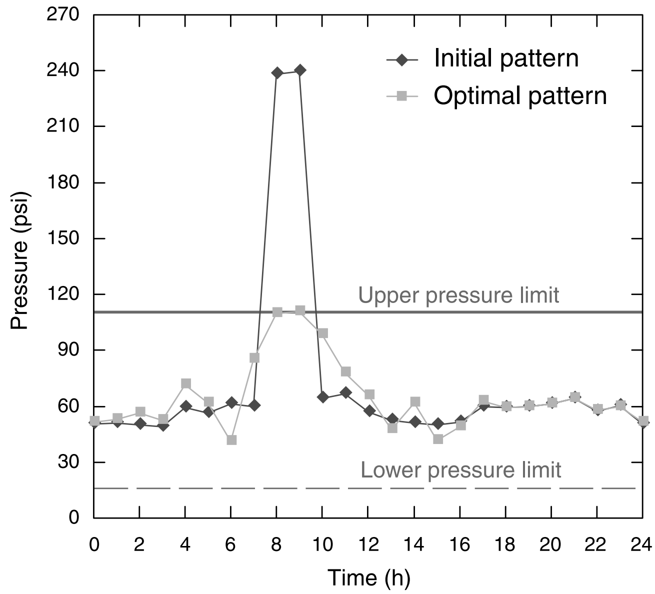

Figure 11 | Water pressure at junction #55,024 (Parkway well 24) using two different patterns, October 1996 (see Figure 2, Part A for location) (from Maslia *et al*. 2001).

computation time of one calculation of the objective function, G as the maximum number of generations, P as the population size, N as the maximum number of iterations, M as the number of time steps and T as the total time, then total time using POGA can be approximately estimated by:

$$T = G^*P^*N^*M^*t \tag{1}$$

In searching optimal patterns, $G = 2$, $P = 5$, $M = 23$, $N = 4$, $T = 920\,t$. In the case of the 1962 simulation, $t = 2$ s, then total computational time $T = 1{,}840$ s $= 30.6$ min $= 0.5$ h. In the case of the 1996 simulation, $t = 10$ s, then total computational time $T = 9{,}200$ s $= 153.3$ min $= 2.6$ h.

From Equation (1), the total computational time seems to be independent of the number of decision variables. In fact, the effect of the number of decision variables on the total computational cost is implicitly reflected in population size and generations. In general, the greater the number of decision variables, the larger the number of population size and generations should be set. For example, in the case of the 1996 simulation, there are 20 water supply wells in the system. The resulting number of decision variables is ($20 \times 24 = 480$). If the GA is directly

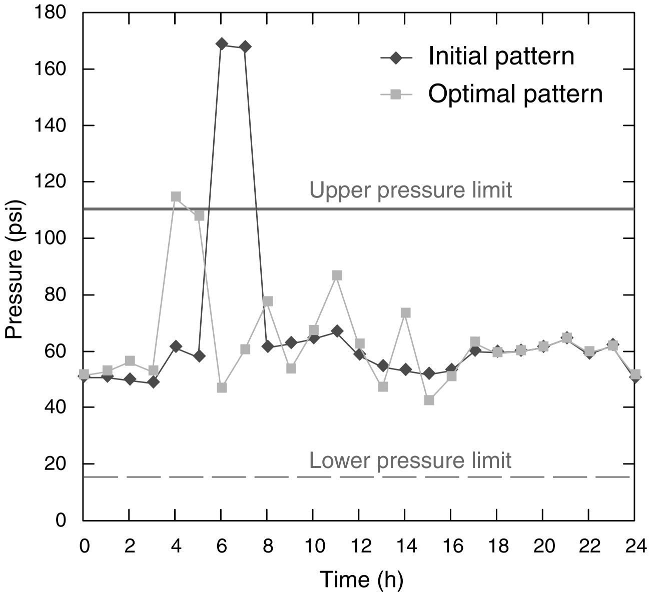

Figure 12 | Water pressure at junction #55,022 (Parkway well 22) using two different patterns, October 1996 (see Figure 2, Part A for location) (from Maslia *et al*. 2001).

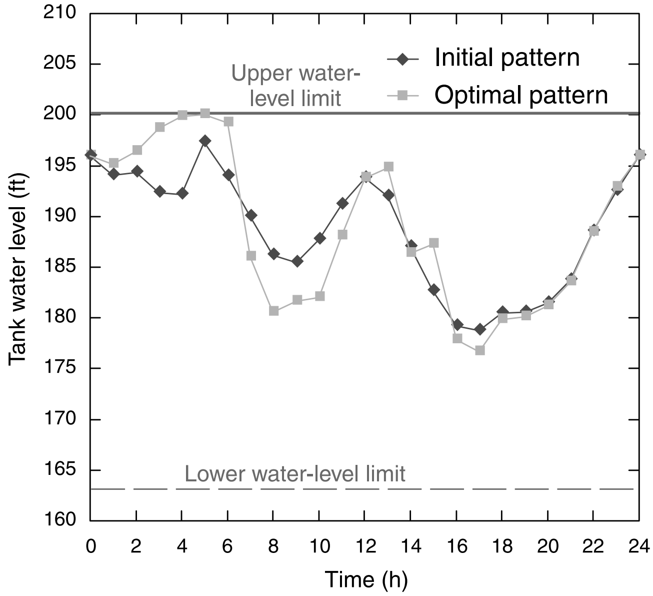

Figure 14 | Water pressure at Indian Hill elevated storage tank using two different patterns, October 1996 (see Figure 2, Part A for location) (from Maslia *et al*. 2001).

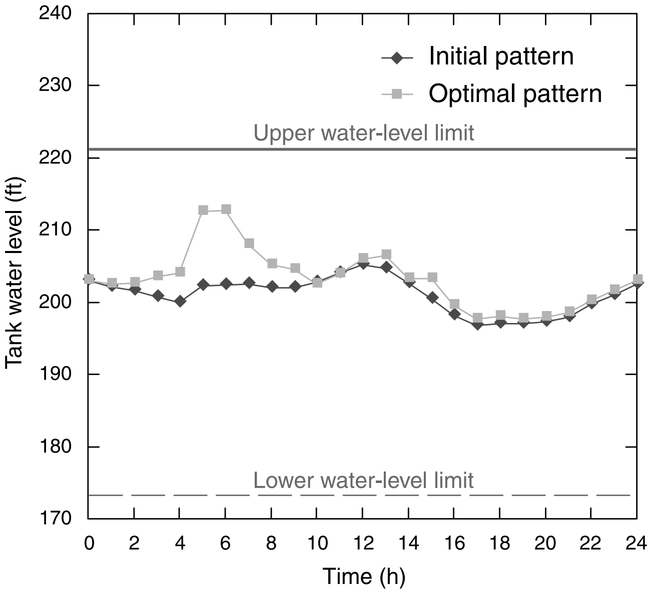

Figure 13 | Water pressure at North Dover elevated storage tank using two different patterns, October 1996 (see Figure 2, Part A for location) (from Maslia *et al*. 2001).

used to solve this problem, the population size should be set as 50 and the number of generations should be set to at least 20. The total computational time T would be at least ($50 \times 20 \times t = 10{,}000$ s $= 2.8$ h). In POGA, the optimization problem is divided into 23 sub-problems during each iteration. The number of variables in the sub-problem reduces to the number of water supply wells. In the case of the 1996 simulation, it is 20. Therefore, we can choose a small population size and generations to reduce the computational cost. Furthermore, the equality constraints cannot be easily handled as they are treated in POGA, if GA is directly applied to this problem.

SENSITIVITY ANALYSIS

Model parameter uncertainty and variability may occur because of spatial and temporal variability of data, incomplete or missing data, or measurement errors. Sensitivity analyses are typically conducted as part of a model calibration process to assess changes in simulation results when adjustments or modifications are made to certain model parameters (Walski *et al*. 2001). Sensitivity analyses conducted as part of the historical reconstruction of water distribution system operations were designed to assess

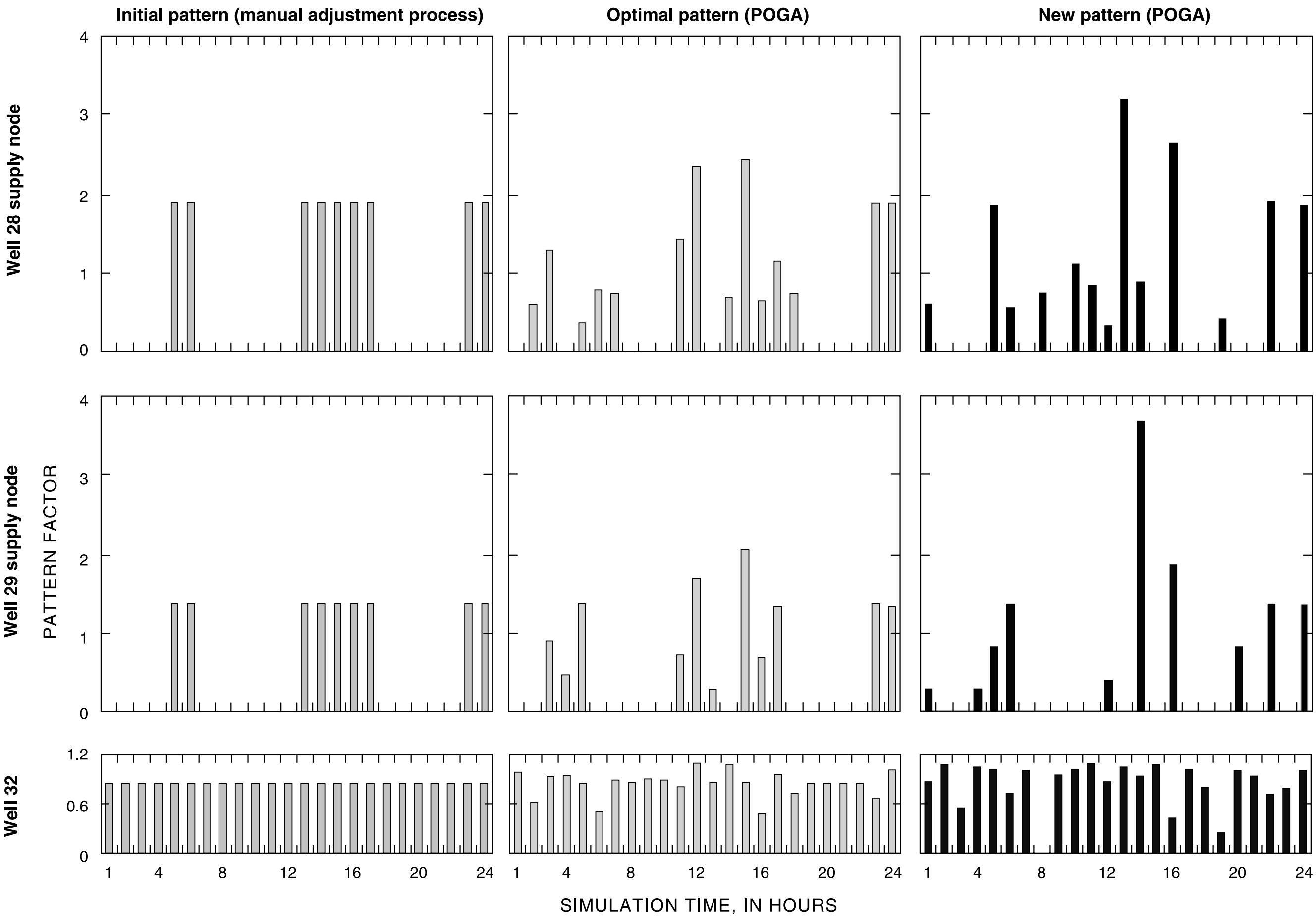

Figure 15 | Initial, optimal and new operation patterns of water supply wells, October 1996 (see Figure 2, Part A for well locations) (from Maslia *et al*. 2001).

changes in the percentage of water contributed by a well or well field to pipeline locations (proportionate contribution) rather than to assess changes in the simulated hydraulics of the distribution system. Output from the source trace analyses – the simulated proportionate contribution of water – was considered as one of the risk factors in the epidemiological study. If large but reasonable variations in model parameter values result in correspondingly large variations in the percentage of water contributed by a well or well field to pipeline locations, the estimates of exposure to the different water sources may result in exposure misclassification. On the other hand, if changes in the simulated proportionate contributions are small regardless of the magnitude of change in model parameters, then simulation variability will not greatly detract from the confidence assigned to exposure classifications. The bases of comparison for all sensitivity analysis results were the corresponding results obtained through the manual adjustment of supply well and supply node patterns, previously described in the section 'Historical reconstruction analysis using manually adjusted pattern factors (base conditions)'. A complete and detailed description of the sensitivity analyses is provided in Maslia *et al*. (2001).

Three types of operational and hydraulic constraints involving the GA simulations were varied during the sensitivity analyses in order to determine the effects of constraint changes on the simulated proportionate

Table 3 | Description of operational and hydraulic constraints varied for sensitivity analyses[1]

Type of variation	Sensitivity simulation identification	Description of parameter variation and operational and hydraulic constraints
Well and supply node pattern factors	SENS0	Initial well and supply node pattern factors derived by use of POGA. Minimum allowable pressure, 15 psi (10 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 0.0 ft (0.0 m)
	SENS1	Alternative well and supply node pattern factors that are not as optimal as simulation SENS0, but still provide a system operation that satisfies constraints. Minimum allowable pressure, 15 psi (10 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 0.0 ft (0.0 m)
	SENS2	Minimum allowable pattern factor, 0.25; minimum allowable pressure, 15 psi (10 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 0.0 ft (0.0 m)
	SENS3	Minimum allowable pattern factor, 0.75; maximum allowable pattern factor, 1.25; minimum allowable pressure, 15 psi (10 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 0.0 ft (0.0 m).
Minimum pressure criteria	SENS4	Minimum allowable pattern factor, 0.15; minimum allowable pressure, 20 psi (13.8 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 0.0 ft (0.0 m)
	SENS5	Minimum allowable pattern factor, 0.15; minimum allowable pressure, 30 psi (20.7 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 0.0 ft (0.0 m).
Storage tank water level difference criteria	SENS6	Minimum allowable pattern factor, 0.15; minimum allowable pressure, 20 psi (13.8 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 3.0 ft (1.0 m)
	SENS7	Minimum allowable pattern factor, 0.15; minimum allowable pressure, 30 psi (20.7 N cm^{-2}); maximum allowable pressure, 110 psi (75.8 N cm^{-2}); difference between storage tank starting and ending water level in a 24-hour period, 3.0 ft (1.0 m)

[1]The bases of comparison for all sensitivity analyses (SENS0–SENS7) are proportionate contributions derived using the manual adjustment of pattern factors described in the 'Historical reconstruction analysis using manually adjusted pattern factors (base conditions)' section.

contribution results. The constraints subjected to variations were (Table 3):

- pattern factors assigned to supply nodes (operational variation in value and time of day);
- minimum pressure requirements at model nodes; and
- allowable storage tank water level differences between starting and ending time of a simulation (hour 0 and 24, respectively).

It is important to note that, in Table 3, sensitivity analysis SENS0 represents the differences in the supply of water simulated using the manually adjusted pattern factors and the pattern factors derived using the original POGA formulation. Sensitivity analysis SENS3 restricts the value of pattern factors to a very narrow range of 0.75–1.25 (otherwise a value of zero was assigned). This analysis was conducted to effectively simulate the operations of wells and supply nodes as being either 'on' (pattern factor value of 0.75–1.25) or 'off' (pattern factor value of 0.0).

A comparison of simulation results obtained from sensitivity analyses SENS0 (POGA) with corresponding results obtained using the manual adjustment of pattern factors for each month of the historical period, indicated that the simulated proportionate contributions of water were highly similar regardless of the simulation approach. Using 1988 as an example, a comparison of the simulated proportionate contribution of water from wells and well fields to five selected pipeline locations (A–E) derived from the manual adjustment of pattern factors and sensitivity analyses SENS0 indicate that results are nearly identical (Figure 16). The graphs in Figure 16 further demonstrate that, at specific historical pipeline locations in the Dover Township area, the difference between results obtained using the two simulation approaches is insignificant.

Pattern factors derived using the manual adjustment and corresponding pattern factors from all sensitivity analyses (SENS0–SENS7) derived using the POGA approach are shown in Figure 17 for the supply node representing Parkway well 22. The pattern factors, derived for October 1996 demand conditions, show the effect of conducting the different sensitivity analyses with their respective constraints. For example, sensitivity analyses SENS3 were conducted to simulate the wells and pumps in either the 'on' or the 'off' position (Table 3). Therefore, the values of pattern factors were constrained to a range between 0.75 and 1.25 for the 'on' position or 0.0 for the 'off' position. This constraint was in addition to the pressure and storage tank water level constraints. As shown in Figure 17, the resulting pattern factors range between 0.75 and 1.25 for hours that the supply node is operational, and are 0.0 for simulation hours 0400 to 0500 and 1700 to 1900 when the supply node is not operational. The pattern factors resulting from all of the sensitivity analyses (Figure 17) also show some significant differences in terms of values and hours of operation when compared with the pattern factors derived using the manual adjustment process. However, regardless of the value or origin of the pattern factors derived using the sensitivity analyses (the POGA approach), the simulated proportionate contributions of water when compared with corresponding results obtained using the manual adjustment process were highly similar.

Statistical analyses were conducted on the differences in the simulated proportionate contribution of water using the results of the manual adjustment process as the bases of comparison with all sensitivity analyses (SENS0–SENS7). The statistical analyses assumed that the differences could be characterized by a normal distribution. The statistical analysis can be interpreted as providing a quantitative evaluation for the differences in the proportionate contribution of water for any plausible operational mode for the historical water distribution system. For all sensitivity analyses, the mean, mode and median of the differences are 0% and the standard deviation of the differences of proportionate contribution of water is 3.9%. For the entire historical period, sensitivity analyses indicate that the differences in the proportionate contribution of water – simulated by the range of operating conditions and hydraulic constraints previously described (Table 3) – are insensitive to the manner in which the water distribution system was operated over a 24-hour period. Thus, the minor differences in the simulated proportionate contribution of water between the manual adjustment process and the sensitivity analyses (POGA) indicate that there was a narrow range of conditions within which the historical water distribution system could have successfully operated to maintain a balanced flow condition and satisfy the water supply demand.

CONCLUSIONS

In this study, we formulated the reconstruction of water supply strategies in the water distribution system as an optimization problem. In the model developed, we selected the pumping rates of water supply wells as decision variables, and penalty functions were assigned to the

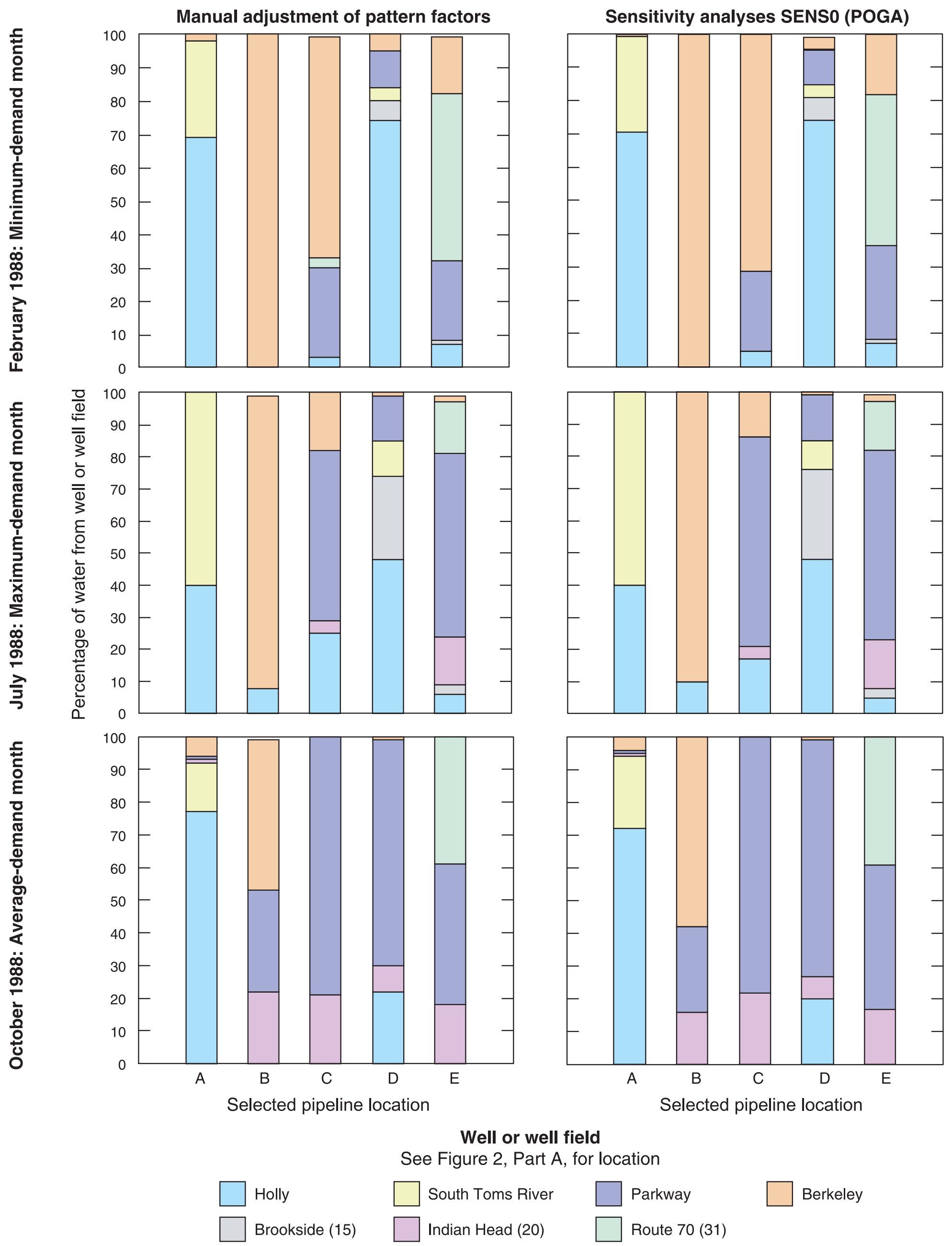

Figure 16 | Simulated proportionate contribution of water derived from the manual adjustment of pattern factors and sensitivity analyses SENS0 for selected pipeline locations, Dover Township area, New Jersey, February, July and October 1988 (number in parenthesis is individual well number) (from Maslia *et al*. 2001).

violation of the management objectives, which are incorporated into the objective function. The equality and the bounded limits of decision variables for each water supply well reflect the historical operation of the system. Considering the non-uniqueness of the solution of this model, we developed a new optimal model, which was based on the optimal solution of the first model. The objective function in the new model reflects the expectation that the new

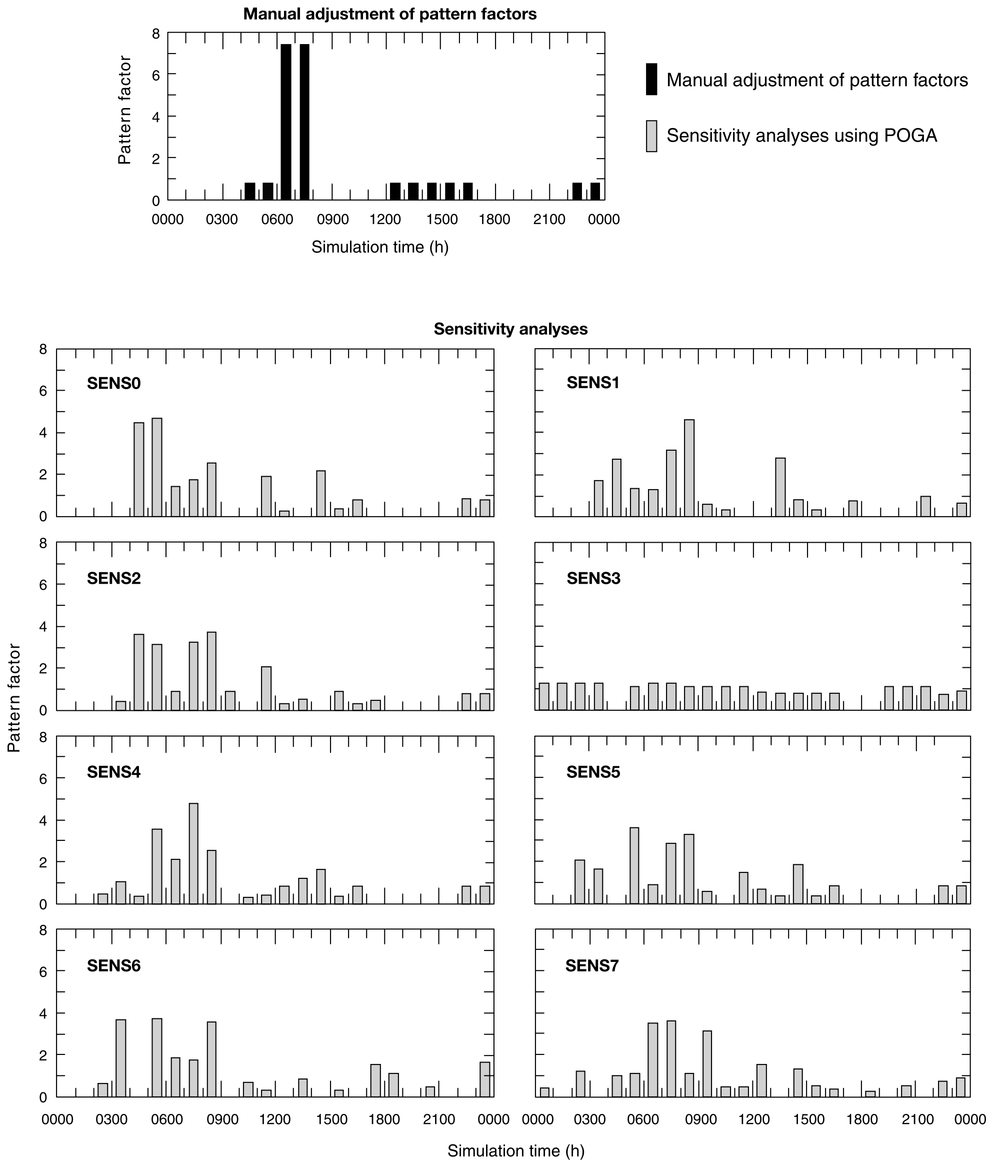

Figure 17 | Pattern factors derived using manual adjustment and sensitivity analyses (POGA), supply node represents Parkway well 22, October 1996 conditions (see Figure 2, Part A for well location) (from Maslia *et al*. 2001).

patterns obtained from the second model have an objective function value close to the optimal solution obtained from the first model. However, there should also be a large difference between the 'optimal patterns'. Both models have the same constraints so that the results obtained from these models can be used to analyse the effect of different patterns on hydraulic and proportionate water distribution characteristics of the system. Both models are multiple-stage nonlinear dynamic problems with equality constraints. For the solution of both models, a new algorithm, identified as the progressive optimality genetic algorithm (POGA), was developed and coded. POGA is dynamically linked with the EPANET2 model to construct a highly flexible system, which provides a visual tool for reconstructing the water supply strategies in a water distribution system (Part A of this paper, this issue).

We have applied the integrated system to the monthly datasets of the water distribution system serving the Dover Township area, New Jersey, for the epidemiological study period of December 1962–January 1996. We have also analysed the effect of several parameters on water supply strategies in the water distribution system. The computational results show that the objective function value improved significantly for all of the datasets, which implies that the optimized water supply strategies are better in satisfying the water supply objectives than the manually constructed water supply strategy. Based on the results and analysis, the following conclusions can be reached:

i. The models described in this paper reflect the historical conditions associated with the water distribution system, and can be used to reconstruct the water supply strategies. The POGA is an effective and efficient algorithm for the solution of both models. The idea developed in POGA may be applied to the solution of other multi-stage dynamic problems.
ii. Optimized water supply configuration for the water distribution system may improve the hydraulic condition of the system operation, reduce the violation of water pressure constraints, and heighten the efficiency of the system.
iii. The reconstruction problem may not have a unique solution. However, the different patterns obtained using two different optimization models satisfying optimal conditions of each model, seem to have little effect on water pressure distribution at junctions, the operation of tanks in the water distribution system, and consequently have little effect on the final proportionate water distribution pattern in the system.
iv. The constraints on minimum water pressure at junctions and water level difference in the tanks at the beginning and at the end of the simulation period directly affect the optimal patterns. The relaxed allowable water level difference in tanks may reduce the violation of water pressure at junctions, but may result in a large accumulated water level difference between the beginning and the end of the simulation period for longer simulation periods.
v. The minor differences in the simulated proportionate contribution of water between the manual adjustment process and the sensitivity analyses (POGA) indicate that there was a narrow range of conditions within which the historical water distribution system could have successfully operated to maintain a balanced flow condition and satisfy the water supply demand.

The results of the analysis presented in this paper provided health scientists with data that were needed for the epidemiological study of childhood leukaemia and nervous system cancers that occurred in the period 1979 through 1996 in Dover Township, New Jersey. This approach allowed epidemiologists to more accurately assess the association between the occurrence of childhood cancers and exposure to each of the sources of potable water entering the distribution system, including those known to have been historically contaminated. As a result, the health scientists conducting the epidemiological study (NJDHSS 2003) documented a statistically significant association and consistency in multiple measures of association between prenatal exposure to time-specific contaminated community water source (1982–1996) and leukaemia in female children of all ages (odds ratio (OR) of 5.0 and 95% confidence interval (CI) of 0.8–3.1). These results are significant because out of hundreds of cancer cluster investigations, only two – Woburn, Massachusetts and Dover Township, New Jersey – have shown an association between environmental exposures and childhood cancer (Costas *et al*. 2002; NJDHSS 2003).

ACKNOWLEDGEMENTS

The research described in this paper was supported by Cooperative Agreement award number U50/ATU499828-06 for the Research Program for Exposure-Dose Reconstruction between the Agency for Toxic Substances and Disease Registry (ATSDR), US Department of Health and Human Services and the Georgia Institute of Technology. The authors express appreciation to their colleagues at ATSDR, the New Jersey Department of Health and Senior Services and the US Environmental Protection Agency, National Risk Management Research Laboratory, for assistance with, and suggestions for, various phases of the project. Additionally, appreciation is expressed to RADM (ret.) Barry L. Johnson, PhD for his initial support of the project, and to Henry Falk, MD, Director, National Center for Environmental Health/ATSDR, for his continued support of the project.

DISCLAIMER

Use of trade names and commercial sources is for identification only and does not imply endorsement by the Agency for Toxic Substances and Disease Registry or the US Department of Health and Human Services.

REFERENCES

Aral, M. M., Guan, J., Maslia, M. L. & Sautner, J. B. 2001 *Reconstruction of Hydraulic Management of a Water-distribution System Using Optimization*, Report No. MESL-01-01. Multimedia Environmental Simulations Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology.

Aral, M. M., Guan, J., Maslia, M. L., Sautner, J. B., Gillig, R. E., Reyes, J. J. & Williams, R. C. 2004 Optimal reconstruction of historical water supply to a distribution system: A. Methodology. *J. Wat. Health* **2**(3).

Board of Public Utilities, State of New Jersey 1962 *Annual Report of Toms River Water Company, Year Ended 31 December 1962*. Newark, New Jersey: Board of Public Utilities.

Board of Public Utilities, State of New Jersey 1996 *Annual Report of United Water Toms River, Year Ended 31 December 1996*. Newark, New Jersey: Board of Public Utilities.

Costas, K., Knorr, R. S. & Condon, S. K. 2002 A case-control study of childhood leukemia in Woburn, Massachusetts: the relationship between leukemia incidence and exposure to public drinking water. *Sci. Total Environ*. **300**, 23–35.

Maslia, M. L., Sautner, J. B., Aral, M. M., Reyes, J. J., Abraham, J. E. & Williams, R. J. 2000*a* Using water-distribution system modeling to assist epidemiological investigation. *ASCE J. Wat. Resour. Plann. Managmt* **126**(4), 180–198.

Maslia, M. L., Sautner, J. B. & Aral, M. M. 2000*b* *Analysis of the 1998 Water-Distribution System Serving the Dover Township Area, New Jersey: Field-Data Collection Activities and Water-Distribution System Modeling*. Atlanta, Georgia: Agency for Toxic Substances and Disease Registry.

Maslia, M. L., Sautner, J. B., Aral, M. M., Gillig, R. E., Reyes, J. J. & Williams, R. C. 2001 *Historical Reconstruction of the Water-Distribution System Serving the Dover Township Area, New Jersey: January 1962–December 1996*. Atlanta, Georgia: Agency for Toxic Substances and Disease Registry.

NJDHSS (New Jersey Department of Health and Senior Services) 2003 *Case-control Study of Childhood Cancers in Dover Township (Ocean County), New Jersey*. Trenton: New Jersey Department of Health and Senior Services, Division of Epidemiology, Environmental and Occupational Health.

Rossman, L. A. 2000 *EPANET2 Users Manual*. National Risk Management Research Laboratory, US EPA, Cincinnati, Ohio.

Walski, T. M., Chase, D. V. & Savic, D. A. 2001 *Water Distribution Modeling*, 1st edition. Waterbury, Connecticut: Haestad Press.