Optimal Sampling Design for Model Calibration Using Genetic Algorithm: a Case Study

Kourosh Behzadian[i]; Abdollah Ardeshir[ii] and Farhad Sabour[iii]

ABSTRACT

Before calibrating a water distribution model, selection of the best points to collect data is undoubtedly an important task for relevant experts. This paper presents a systematic process of sampling design (SD) in a real water distribution network (WDN). The purpose is to find a specific number of optimal monitoring locations in which measurement devices (pressure loggers) will be installed. At first, skeletonization is applied to Mahalat WDN for selecting only the parts of WDN that have a significant impact on the behavior of the system. SD is then formulated and solved as an optimization problem by using a single objective genetic algorithm (SOGA) model. Model prediction accuracy is defined as the objective function. The solutions obtained by SOGA are compared to the ones obtained by expert choice (EC). The results show that SOGA can find measurement locations with significantly better prediction accuracy rather than EC. Furthermore, the model is tested for different number of sampling locations in addition to the ones in EC. It shows that SOGA can achieve the same model uncertainty of EC with only less than half of the number of EC's monitoring locations.

KEYWORDS

Monitoring location; Calibration; Water distribution network; Genetic algorithm; Expert choice.

1. INTRODUCTION

The data for calibration of a water distribution network (WDN) model is usually collected from a series of field tests at strategic locations within the network, in which pressure heads are recorded [1]. The accuracy of calibration depends on the quality and quantity of the collected data. Therefore, selection of appropriate locations, called sampling design (SD), has been a challenge among researchers and practitioners, especially in recent years [2], [3]. Note that, after monitoring data are collected from measurement locations, they will be used later on in the calibration of the analyzed WDN model to adjust calibration parameters.

The practitioners and modellers of WDN often use some simplified approaches for sampling design since they seek methods which are straightforward to apply and without any analytically complex computation. These approaches often employ some special criteria such as clustering based on physical and hydraulic characteristics of WDN [14]. Walski (1983) suggested monitoring locations near the high-demand locations and on the perimeter of the skeletonized network [4].

However, this is prone to errors as it may lead to the selection whose accuracy is less than the required one for SD. Therefore, the need for finding optimal locations can lead to a better selection regarding the time and the money being spent on this way.

Lee and Deininger (1992) and Yu and Powell (1994) addressed optimal SD in a WDN model for the first times [5], [6]. Finding optimal solutions on a real world WDN model are challenging because the procedure usually needs coupling an optimization model with hydraulic simulation models (e.g., EPANET [7]) to evaluate potential solutions. These coupled models can make a complex optimization model. This model may be difficult to be solved by traditional optimization model because of being non-convex, and/or discontinuous. In recently developed applications of water resources optimization, genetic algorithms (GAs) have been noticed by many researchers and practitioners [1], [8], [9], [10]. Meier and Barkdoll (2000) solved a SD problem in real world WDN model using GAs [10]. They formulate objective function to find fire hydrants in the network in such a way that will cause water to flow at nonnegligible velocities in more pipes.

Bush and Uber (1998) proposed three relatively simple SD methods, including max-sum, weighted sum and max-

[i] Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran (e-mail: Behzadian@aut.ac.ir).
[ii] Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran (e-mail: Ardeshir@aut.ac.ir).
[iii] President of Sadra Negar Consulting Engineers, Tehran, Iran (e-mail: sufarhad@yahoo.com).

min, for WDN model calibration based on the analysis of relevant Jacobian matrix. The methods were derived from minimization of parameter covariance matrix [11]. Lansey et al. (2001) then addressed prediction covariance matrix as a better index of measuring errors for finding optimal sampling locations in WDN model [12].

Kapelan (2002) proposed that optimal sampling locations are better determined by evaluating the trade-off between calibrated model accuracy and the cost of SD [3]. Model accuracy is usually evaluated using some norms of parameter or prediction covariance matrix which, in turn, is calculated from the relevant Jacobian matrix [11], [12]. The cost of SD is typically surrogated by the number of sampling devices used (e.g., pressure loggers) [3], [11].

A newly developed model by Kapelan et al. (2003) presented a multi-objective genetic algorithm (MOGA) for SD with the aim of calibration of WDN models. In this approach, elements of the Jacobian matrix are calculated prior to optimization model run. They applied and verified their model in a small artificial case study [13].

The aim of this paper is to assess the performance of suggested sampling design optimization approach on a real case study. More specifically, the objectives consist of: (1) applying a genetic algorithm optimization for sampling design on a real world WDN problem; (2) comparing the optimal SD solution obtained by the developed model to the ones suggested by expert. In the next section, sampling design problem is briefly illustrated and formulated. Then, a real case study and the process of its skeletonization are demonstrated, respectively. Results of applying the model to the case study are then presented. Finally, a summary is given and conclusions are made.

2. Sampling Design

The current SD is carried out under the following assumptions: (1) the type of predicted variables, which include nodal pressure, pipe flows or both, is assumed to be only nodal pressure head; (2) pipe roughness coefficients are considered as calibration parameters; (3) the steady-state WDN hydraulic model is calibrated under average demand condition.

A. Single-objective Optimization

The SD problem is formulated and solved here as a single-objective genetic algorithm (SOGA) optimisation problem. The objective is to maximise the model prediction accuracy with a fixed number of sampling devices. Constraints are mass and energy equations of hydraulic equality, which are handled through a well-known simulation model (EPANET) [7].

1) Problem Formulation

To quantify the calibrated model prediction accuracy, a first-order second-moment (FOSM) model and linear regression theory are used to approximate both parameter covariance matrix and prediction covariance matrix [15]. Based on this quantification, if a set of measurement locations are assumed, the uncertainty of WDN model parameters associated with those measurement locations is estimated as the diagonal elements of covariance matrix defined as follows [2], [11]:

$$\mathbf{Cov_a} = s^2.(\mathbf{J}^{\mathrm{T}}\mathbf{J})^{-1} \tag{1}$$

where s=standard deviation of measurement devices; and $\mathbf{J}$=Jacobian matrix of derivatives $\partial y_i / \partial a_k$ ($i = 1,\dots,N_p; k = 1,\dots,N_a$), y=vector of pressure predicted variables in locations of interest, a=vector of calibration parameters (pipe roughness coefficients), N_p=number of measurement locations of interest, N_a=number of calibration parameters (pipe roughness coefficients). Therefore, the value of *i*th diagonal element in matrix Cov_a corresponds to the uncertainty of *i*th model parameter (pipe roughness coefficient).

For a better representation of the model uncertainty, the model prediction (pressure head) uncertainty needs to be evaluated. This is performed through the propagation of parameter uncertainty to model predictions as follows [2], [12]:

$$\mathbf{Cov}_z = \mathbf{J}_z.\mathbf{Cov_a}.\mathbf{J}_z^T \tag{2}$$

$\mathbf{J}_z$=Jacobian matrix of derivatives $\partial z_i / \partial a_k$ ($i = 1,\dots,N_z; k = 1,\dots,N_a$); z=vector of model pressure predictions in all nodes, and N_z=number of model pressure predictions of interest corresponding to all nodes of WDN in which pressure prediction is important (here referred to all potential locations of pressure logger installation). The value of the *i*th diagonal element in matrix $\mathbf{Cov}_z$ indicates the uncertainty of *i*th model pressure prediction, and somehow it indicates the variance of *i*th model pressure prediction.

To totally represent the value of model prediction uncertainty, the average of square root of all diagonal elements in matrix $\mathbf{Cov}_z$ (i.e., all model prediction uncertainty) is assumed as model uncertainty representation which is defined as follows:

$$F = \frac{1}{N_z}\sum_{i=1}^{N_z}\mathbf{Cov}_{z,ii}^{1/2} \tag{3}$$

The above objective function is in absolute terms. An alternative to this form is to address in relative and dimensionless form. Therefore, the objective value is defined as the normalised (relative) model prediction accuracy as follows:

$$Max\ f = \frac{F_{ml}}{F} \tag{4}$$

where F_{ml}=the value of model pressure uncertainty for ideal state where all potential measurement locations are monitored. Here, it is assumed that the total budget of sampling is constant. As a surrogate, the number of pressure loggers is introduced as an indicator of sampling cost, which is assumed to be fixed.

II) Genetic Algorithm

A standard genetic algorithm (GA) is used in this study. In the SD model, each chromosome represents a set of sampling locations within the WDN model. Integer-value encoding is used as the location of measurement device in the WDN. The length of each chromosome is the number of pressure loggers. In fact, number of genes for each chromosome is equal to the number of sampling locations, which is shown in Fig. 1. Fitness function is assigned for each chromosome and is defined as the aforementioned objective value (f) in (4).

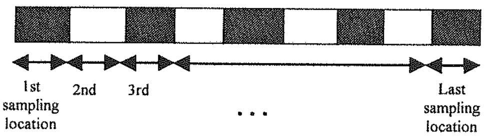

Fig. 1: Schematic representation of each chromosome.

The GA operators are selection, crossover, and mutation. Elitism operator is also used in this study to keep the best solution in the next generations. The GA generation is repeated until some finishing criteria are met. In this study, after a pre-specified number of generations in which all GA runs converge, the GA is stopped.

B. Expert choice

Without applying an optimization model, expert choice (EC) is used by professionals to find critical points of the system for pressure logger installation. The following EC method is an initiative proposed by practitioners to identify pressure logger installation based on pressure zones approach [14]. To select a set of monitoring stations, the following steps are sequentially performed:

In the first step, an analysis on the WDN topography and morphology is done to recognize different zones with clustering WDN nodes' elevation. In the second step, an analysis on flow distribution statement is performed to recognize different zones which are supplied by the same source(s). Consequently, a better behavior of pressure head loss variations is identified based on these two steps. Then, the WDN is divided into a number of pressure zones based on the classification performed in steps 1 and 2. Finally, pressure logger stations are determined for each zone based on the following criteria.

At first, weighted average of elevation over the customers located in each zone is calculated as a representative elevation of that zone as follows:

$$El_{ave}^{j} = \frac{\sum_{i=1}^{n_j}(m_i^j \times El_i^j)}{\sum_{i=1}^{n_j} m_i^j} \tag{4}$$

where El_i^j =*i*th elevation line in zone j; m_i^j= number of customers at *i*th elevation line in zone j; n_j =number of elevation line in zone j. Measurement station is then determined at a junction with the elevation equal to the representative elevation of that zone. The selected monitoring station must satisfy other required conditions (e.g., imposed by executive limitations). For instance, pressure logger must be installed in a main pipeline in the relevant zone.

3. CASE STUDY

Proposed SD problem is applied to Mahalat WDN model as a real world case study. The city of Mahalat is located in the central part of Iran. The general layout is presented in Fig. 2. A brief summary of the case study is given here; for more details, refer to [14].

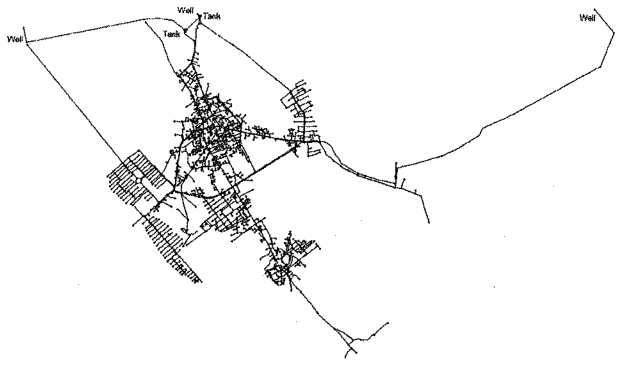

Fig. 2: Layout of Mahalat WDN model.

The WDN covers approximately 46 km^2, with a population of around 160,000. Model demands are predominantly domestic with some commercial users. To reduce the high pressure head induced by steep slope of the city, six pressure reduced vales (PRVs) are used to decrease pressure heads to a fixed pre-specified values. The main characteristics of the pipes are shown in Table1. The majority of main pipes material is ductile iron and the majority of small-size pipes are made of PVC; and asbestos cement pipes cover the larger part of middle-size pipes material in the network. An EPANET hydraulic model was constructed including 1814 pipes with the total length of approximately 101 Kilometers, 1771 junctions, 2 tanks, and six PRVs based on the available data.

TABLE 1: SUMMARY OF PIPES MATERIALS AND DIAMETER.

No.	Original Material	Number of Pipes	Range of Diameter (mm)
1	Asbestos Cement	406	80-250
2	Ductile Iron	470	100-500
3	Galvanized Iron	113	25-125
4	PVC	657	25-110
5	Steel	166	20-65

The WDN is supplied by gravity from three wells and two service tanks (reservoirs) around the city, whose position are shown in Fig. 2. The average water demand is 158.9 L/S. The water is pumped into the system with a constant rate. The reservoirs store and balance the fluctuations of water daily consumption. Estimation of pipe roughness coefficients (Hazen-Williams C-factor) were performed based on the main characteristics of the pipes and water quality [16], [17].

Based on the circumstances in the EC section, the minimum and maximum elevation of the WDN was identified equal to 1554.4 m and 1924.4 m above sea level, respectively. Then, The WDN was divided into 13 pressure zones with respect to elevation variations and water supply sources. 20 stations were initially determined for installing pressure logger and collecting pressure measurements. Two pressure loggers were excluded since they had some obvious anomalies after gathering the data. Fig. 3 depicts locations of the remaining 18 pressure loggers as EC (expert choice).

4. SKELETONIZATION

Skeletonization is the process of selecting only the parts of the hydraulic network that have a significant impact on the behavior of the system for inclusion in a water distribution model [18]. Although the portions of the network that play a minor role in the system are removed, they are not ignored. Rather, the effects of these elements are accounted for within the parts of the system that are included in the model.

With respect to the huge number of pipes and junctions in the case study, two approaches of skeletonization are sequentially used here as below.

(1) Branch trimming is first applied in which short dead-end branches and their corresponding junctions are removed. After removing the link, any nodal demand of the removed junction is reallocated to the upper junction at the beginning of the branch. For removing pipes within performance of this process, other special criteria can be defined. The criterion defined here is to remove small diameters pipes (less than 100 mm) and the pipes out of the reach of existing measurement points (EC). Note that pipe roughness coefficients for the pipes on dead-end branches and out of the reach of existing measurement points are inefficient in the calibration process [1]. Therefore, removal of these pipes has no effect on declining the generality of the SD model developed here.

(2) Series pipe removal is the process of removing the intermediate node and merging the series pipes. Further, specific criteria can be defined as which conditions the two series pipes must have in order to complete the process. Here, it is assumed that only equal diameter series pipes are merged. Moreover, a distance-weighted technique is used to distribute any available demand of the removed node to the two end nodes of the newly merged pipe. This technique divides the demands between the two nodes based on their proximity to the node being removed. Note that, when the condition for removal is satisfied between two series pipes, the merging process is carried out in such a way that hydraulic capacity keeps constant after merging.

As one of the purposes of this paper is the comparison of the proposed method with the EC, the junctions which have been used as observation junctions in EC should be kept during skeletonization process. Therefore, these junctions are retained to be considered afterwards as the potential nodes for monitoring locations in the optimization SD problem.

With respect to the mentioned steps in skeletonization, the WDN model was skeletonized by WATERGEMS software [18] ten times for removing dead-end branches until there were no meaningful trimming performed; i.e., dead-end branches have been removed up to ten sequential times if they have satisfied the criteria for branch trimming in each time. At the second step, series pipes removal was continued five times until they were no significant removal occurred. Finally, the skeletonized WDN model was made of 237 pipes and 195 junctions including the junctions of EC (Fig. 3).

5. RESULTS AND DISCUSSION

As our purpose is the comparison of the sampling design by using SOGA with the ones obtained by the EC, The optimization problem is solved to find the best 18 monitoring locations (equal to number of the EC's measurement locations) as the positions for pressure logger installation. After skeletonizing the WDN model, all nodes of the system were considered as potential pressure measurement locations (N_{ml}=195). Therefore, number of sets of possible solutions is equal to $\binom{195}{18} \cong 1.16\times10^{25}$, which shows a large space of feasible solutions and justifies the use of genetic algorithm as an optimization model to solve the problem.

Further, it is assumed that the WDN model are calibrated for N_a=7 groups. Although there are a large number of pipes (237), number of parameter calibration groups is assumed to be small number because (1) model prediction error will increase if number of calibration parameters increase [12], [19] (2) it was shown that the computational time for running the model will exponentially be enlarged. Grouping was done by dividing the range of HW pipe roughness coefficients into a 7 distinctive ranges. After estimating HW pipe roughness coefficients, their variations were between 78 and 155. Therefore, they have been classified as the ranges of (78, 90], (90,100], ..., (130,140] and (140,155]. Then, the average of the HW pipe roughness coefficients in each group was considered as the representative roughness coefficient of all pipe of the group. Note that the standard deviation of all pressure loggers is assumed to be equal to s=1.0 m.

When evaluating fitness of each chromosome, Jacobian matrix ($\mathbf{J}$) is required to be calculated each time. For calculating $\mathbf{J}$, hydraulic simulation model (EPANET) needs to be run by N_a+1 times. Instead of this huge computational effort in each generation, an new method is suggested here. Before starting optimization model, Full Jacobian matrix ($\mathbf{J}_{ml}$) is calculated. Matrix $\mathbf{J}$ can be constructed from full Jacobian matrix ($\mathbf{J}_{ml}$) by copying the rows corresponding to the currently analysed set of measurement locations. Note that full Jacobian matrix $\mathbf{J}_{ml}$ is obtained using all potential measurement locations. The matrix has N_{ml}=195 rows and N_a=7 columns.

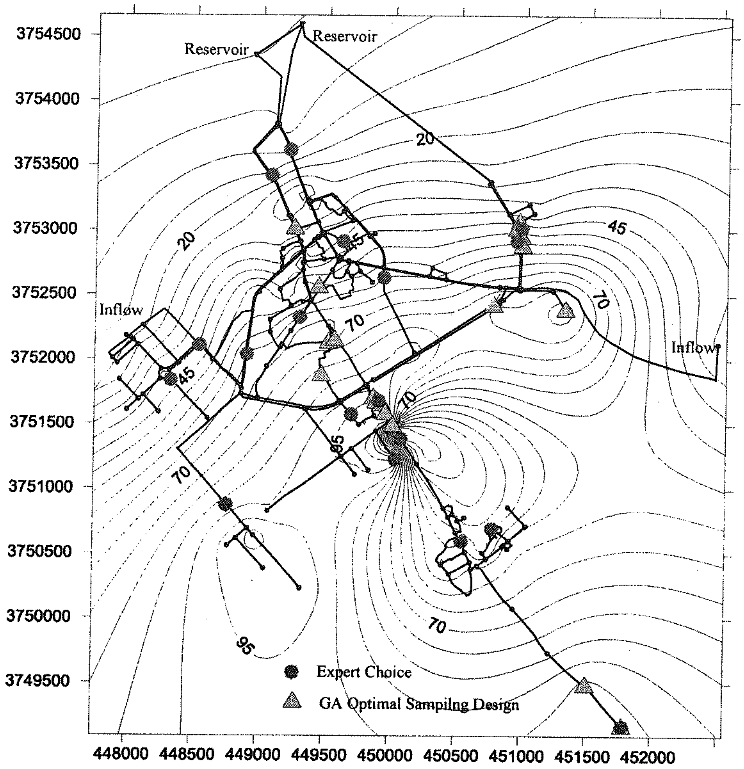

Fig. 3: Pressure monitoring locations for the Expert Choice (EC) and the Genetic Algorithm (GA) sampling design in the skeletonized WDN model of Mahalat; both approaches for 18 measurement points.The contours denote pressure head lines in meters.

A sensitivity analysis based approach was used here for setting the SOGA model parameters. Therefore, they were determined after a limited number of trial runs using different randomly created initial populations. The parameters used in the problem are as follows: population size of 50 chromosomes, roulette wheel selection operator, mutation with the probability of 0.05 and one point crossover with the probability of 0.8. These values were rigorously checked for the parameters in such a way that the fastest convergence of finding optimal solution is obtained. Note that there is no systematic approach to find the best composition of GA parameters setting.

Due to the large space of feasible solutions, the optimal solutions obtained by GA should be treated as suboptimal solutions. It means there is no guarantee to find global optimum. Therefore, GA was run by twenty different times, each time with different randomly generated initial populations. Then, the most frequently selected measurement locations were determined as a suboptimal solution of optimization model. All SOGA runs converged after 5,000 generations. Fig. 4 shows a typical SOGA run convergence of fitness function for the best solution.

The set of 18 optimal measurement locations obtained using SOGA and measurement locations proposed by EC are shown in Fig. 3. The network nodes associated with these solutions are also presented in Table 2. Since the SOGA solution was run twenty times, the optimal measurement locations are accompanied by a percentage which shows the relative frequency of selection of optimal measurement locations among different runs. This percentage denotes the robustness of GA optimal solutions. For example, the selected locations with the percentage of 100 percent are robust as GA optimal solutions. However, some selected locations such as N1735 and N4648 have the percentage of 75%, which implies that there are some other competitive locations within WDN. Nevertheless, the selected points are more robust than other competitive points of WDN.

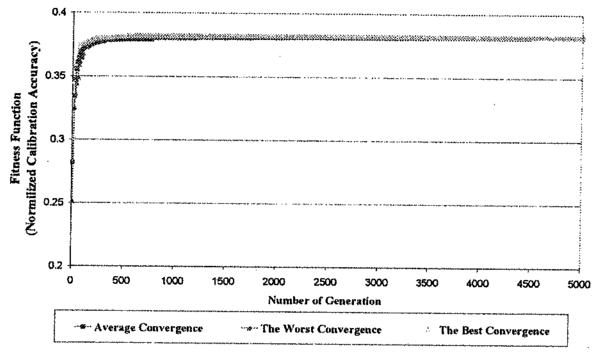

Fig. 4: The convergence of SOGA model runs.

As can be seen in Table 2, four measurement locations (22% of measurement locations) are identical and some other points are close between the two approaches. The occurrence of such clustering can be mainly a consequence of the following possibilities: (1) sensitive nodes, which are likely candidates for SOGA selection, have spread in each zone of the network and have been located in main pipes in each zone (i.e., the locations that had been selected most by EC); (2) since grouping pipe roughness coefficients is somewhat related to the special pipes in each zone, the high uncertainty of some grouped pipe roughness coefficients may have dictated some critical (sensitive) nodes to be selected in that zone. However, there are some points in the network that the solutions are quite different. For instance, on the left of the city, EC have selected three measurement locations while SOGA have not selected any points since there are no pints in this area to be sensitive when calibrating the whole system. In other words, the variations of HW pipe roughness coefficients for the pipes in this area have no or little effect on the resulted pressure heads. Therefore, Jacobian matrix, if it includes any point in this area, is less sensitive rather than the ones which include the points of the other regions.

As can be seen in Fig. 3 for optimal measurement locations, some of them are mainly prone to be selected on the main pipes and where the pressure contour lines are close to each other (i.e., where the hydraulic gradients are too much). Despite this fact of selection for some points, it is difficult for us to be able to generalize to all measurement locations. Of course, one of the main factors in selecting optimal locations is the sensitivity of nodal head over the calibration parameter variations. As pipe roughness coefficients (calibration parameters) are grouped, the type of grouping and the number of pipes in each group can affect the selected locations and therefore should be rigorously checked and determined.

For further comparison, the objective functions and other sensitivity parameters of the two approaches are shown in Table 3. As expected, SOGA is obviously better than EC in all cases. Main objective function of this study (*f*) has a relative improvement of 51%. Interestingly, the maximum pressure prediction uncertainty for EC is 60.10 m in a node. It means that there is negligible monitoring effect on this node when the EC measurement locations are used. However, this value for SOGA is only 1.0 m. Rather, it means that the prediction uncertainty (standard deviation) for the worst node in the network is 1.0 m if the model is calibrated with the measurement locations obtained by SOGA. The next row in Table 3 has the same interpretation.

Further, the uncertainty of the calibration parameters associated with each SD method is shown in Table 3. Standard deviation of parameters indicates the parameter uncertainty which is equal to the square root of diagonal elements in Cov_a. These uncertainties will be acquired for calibration parameters if the model is calibrated with the observing data collected from measurement locations of the relevant method. As can be seen, improvement in parameter estimation (i.e., estimation with less uncertainty) will be obtained for all parameters in SOGA rather than EC.

SOGA was analyzed more to find optimal solutions with different number of monitoring locations. Thus, optimal set of locations were found for number of monitoring locations of 7 to 18. Monitoring locations should be more than 7 since there are 7 calibration parameters, otherwise they will not lead to a well-posed solution [13].

The relative accuracy and associated pressure prediction uncertainty for each specified number of monitoring locations are shown in Fig. 5. As can be seen, relative prediction accuracy decreases when number of monitoring locations decreases. Note that for each point on the trade-off, there is a set of optimal locations for installing measurement devices corresponding to the associated number of pressure logger.

When comparing the SOGA optimal solutions with the EC's solution in Fig. 5, the following can be noted: optimal solution of 8 monitoring locations has the relative accuracy of around 25% (or model uncertainty of 0.65 m), which are approximately equal to the relative accuracy (or model uncertainty) of 18 monitoring locations in EC. It implies that the same model uncertainty of EC can be achieved by SOGA, despite that only less than half of the number of EC's monitoring locations are required. In other words, the accuracy of the EC's can be achieved by SOGA with a few optimal sampling locations. The position of these 8 optimal locations compared to the EC's is depicted in Fig. 6. The SOGA optimal solution shows a uniform spread of locations within the WDN, which can produce approximately the same model uncertainty as EC's produces.

TABLE 2: MEASUREMENT LOCATIONS (NETWORK NODES) OF EC AND SOGA SAMPLING DESIGN.

No.	1	2	3	4	5	6	7	8	9
Expert Choice	N1009	N1205	N1412	N1536	N1632	N1655	N3330	N4124	N5550
SOGA optimal solution	3	N1009	N1015	N1622	N1628	N1632	N1735	N1742	N4648
percentage of selection for SOGA optimal solutions	100	100	100	100	100	100	75	100	75
No.	10	11	12	13	14	15	16	17	18
Expert Choice	N5790	N6042	N6148	N6654	N7273	N8226	N8640	N8910	N8912
SOGA optimal solution	N4687	N7058	N7145	N7192	N7346	N8904	N8910	N8912	N8922
percentage of selection for SOGA optimal solutions	100	100	100	100	100	100	75	100	75

TABLE 3: COMPARISON OF SAMPLING DESIGN RESULTS BETWEEN EXPERT CHOICE AND GENETIC ALGORITHM SAMPLING DESIGN.

No.	Expert design	Best single objective genetic algorithm SD	Relative improvement (%)
Relative pressure prediction accuracy (f)	0.25	0.38	51
Average pressure prediction uncertainty (F) (m)	0.63	0.42	34
Maximum pressure prediction uncertainty (i.e., square root of the largest diagonal element of $\mathbf{Cov}_z$ (m)	60.10	1.00	98
Trace (i.e., sum of the diagonal elements) of $\mathbf{Cov}_z$ (m^2)	164.29	38.09	77
Parameter 1: Value and SD[a]	85.2 and 13.6	85.2 and 7.4	46
Parameter 2: Value and SD	94.4 and 4.1	94.4 and 3.2	22
Parameter 3: Value and SD	101.6 and 8.6	101.6 and 6.9	20
Parameter 4: Value and SD	117 and 469.4	117 and 444.9	5
Parameter 5: Value and SD	123 and 57.5	123 and 33.3	42
Parameter 6: Value and SD	134.6 and 1047.	134.6 and 125.4	88
Parameter 7: Value and SD	147.4 and 20.8	147.4 and 14.4	31
Average of selected locations' demands (L/S)	1.21	0.53	-
Average of selected locations' elevation (m)	1728.14	1720.38	-
Average of selected locations' pressure head (m)	55.29	85.61	-

[a] Standard deviation (indicator of uncertainty) which is equal to the square root of the relevant diagonal element in $\mathbf{Cov}_a$

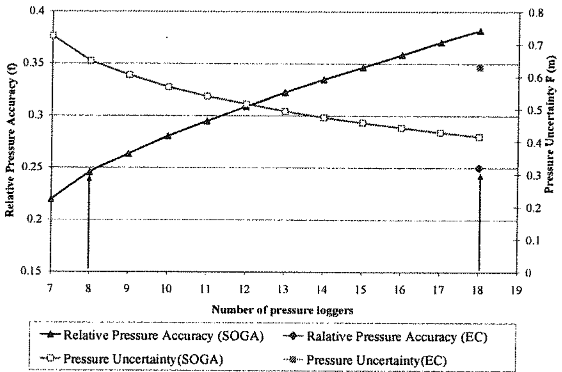

Fig. 5: Trade-off between number of measurement locations and associated relative pressure accuracy (pressure uncertainty).

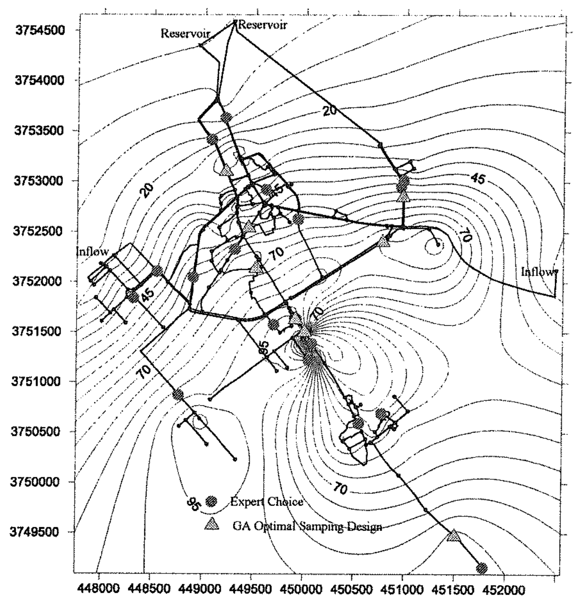

Fig. 6: Pressure logger locations for the Expert Choice (18 measurement points) and SOGA sampling design (8 measurement points). The contours denote pressure head lines in meters.

6. SUMMARY AND CONCLUSION

A single objective genetic algorithm was used to solve a sampling design problem for calibration of WDN model. The objective function was defined as relative pressure prediction accuracy. SOGA was seeking a specified number of the best locations for installing pressure logger. The model applied to a real world case study in Mahalat city, which is large enough to use an optimization model. Before solving SD problem, skeletonization was used in the WDN model to remove unnecessary pipes which have little effect on the behavior of the system. The optimal solution was then compared to that obtained by the EC.

From comparing the two methods (i.e., SOGA and EC), it is verified that SOGA can find better measurement locations, which have more effect on the whole nodes of the WDN model. Consequently, the calibration resulted from these monitoring locations will lead to the least prediction uncertainty.

In addition, SOGA was used to identify the optimal sampling locations with different number of points (Fig. 5). It was shown that the same level of accuracy of EC can be achieved by fewer numbers of optimal measurement locations. For instance, number of 8 optimal measurement locations can have the same prediction of accuracy resulted by 18 monitoring locations of EC (Fig. 5 and Fig. 6). This implies that SOGA can achieve the same model uncertainty of EC despite that only less than half of the number of EC's monitoring locations is required. This proves the cost-effective superiority and significant improvement of optimal SD over EC. Further, Table 3 shows that significant improvement will be obtained for parameter and prediction uncertainty if SOGA measurement locations are used for data monitoring.

Further, if the purpose of SD is model calibration as well as some other objectives (e.g., monitoring the whole WDN), the GA optimization problem can be still applicable. In these cases, GA optimization problem can be possible with assuming some measurement locations to be fixed in SD process. To do so, a penalty function must be defined to deteriorate the solutions' fitness without those previously fixed points. Obviously, the composition of new measurement locations will be influenced by these fixed points.

In addition, the spread of measurement locations in WDN can be dependent on the type of pipe roughness grouping. The sensitivity of such grouping must be investigated more to find the robust solutions.

7. NOMENCLATURE

a =vector of calibration parameters

$\mathbf{Cov}_a$ =parameter variance-covariance matrix

$\mathbf{Cov}_z$ =model prediction variance-covariance matrix

El_i^j =ith elevation line in zone j

F =objective function of pressure prediction uncertainty

f =objective function of relative pressure accuracy

F_{ml} =Value of F assuming that all analyzed locations are monitored

$\mathbf{J}$ =Jacobian matrix of derivatives

$\mathbf{J}_{ml}$ =full Jacobian matrix (all locations monitored)

$\mathbf{J}_z$ =prediction Jacobian matrix

m_i^j = number of customers at ith elevation line in zone j

N_a =number of calibration parameters

N_z =number of model predictions for whom uncertainty are evaluated

N_p =number of measurement devices

N_{ml} =number of total potential locations of SD

N_F =number of objectives

n_j =number of elevation line in zone j

Superscript ^{T}=vector/matrix transpose operator

s=standard deviation of measurement devices

y =vector of WDN model predicted variables

z=vector of model predictions of interest

σ_y =standard deviation of measurement devices

8. ACKNOWLEDGMENT

The authors wish to acknowledge the support and assistance the first author has received from the British Council (Iran), especially Ms. F. Ahmadi and Shell for the help provided throughout the research. The authors are also grateful to the anonymous reviewers for their effective comments.

9. REFERENCES

[1] W. de Schaetzen, "Optimal calibration and sampling design for hydraulic network models." PhD thesis, School of Engineering and computer science, Univ of Exeter, Exeter, U.K., 2000.

[2] Z. Kapelan, D. A. Savic, and G. A. Walters, "Optimal Sampling Design Methodologies for Water Distribution Model Calibration", *Journal of Hydraulic Engineering*, 131(3), pp. 190-200, 2005.

[3] Z. Kapelan, "Calibration of water distribution system hydraulic models." PhD thesis, School of Engineering and computer science, Univ of Exeter, Exeter, U.K., 2002.

[4] T. M. Walski, "Technique for calibrating network models." *Journal of Water Resources Planning and Management*, 109(4), pp. 360–372, 1983.

[5] B. H. Lee, and R. A. Deininger, "Optimal locations of monitoring stations in water distribution system." *Journal Environmental Engineering (Reston, Va.)*, 118(1), pp. 4–16 1992.

[6] G. Yu, and R. S. Powell, "Optimal design of meter placement in water distribution systems." *International Journal of Systematic Sciences*, 25(12), pp. 2155–2166 1994.

[7] L. A. Rossman, *Epanet2 user manual*, US EPA, Washington, D.C., 2000.

[8] S. Yan, and B. Minsker, "Optimal groundwater remediation design using an Adaptive Neural Network Genetic Algorithm." *Water Resources Research*, 42(5), W05407, 2006.

[9] Z. Kapelan, D. A. Savic, and G. A. Walters, "Multiobjective design of water distribution systems under uncertainty", *Water Resources Research*, 41(11), W11407, 2005.

[10] R. W. Meier, and B. D. Barkdoll, "Sampling design for network model calibration using genetic algorithms." *Journal of Water Resources Planning and Management*, 126(4), pp. 245–250, 2000.

[11] C. A. Bush, and J. G. Uber, "Sampling Design Methods for Water Distribution Model Calibration." *Journal of Water Resources Planning and Management,* 124(6), pp. 334-344, 1998.

[12] K. E. Lansey, W. El-Shorbagy, I. Ahmed, J. Araujo, and C. T. Haan, ''Calibration assessment and data collection for water distribution networks.'' *Journal of Hydraulic Engineering,* 127(4), pp. 270–279, 2001.

[13] Z. Kapelan, D. A. Savic, and G. A. Walters "Multi-objective Sampling Design for Water Distribution Model Calibration", *Journal of Water Resources Planning and Management,* 129(6), pp. 466-479, 2003.

[14] Sadra Negar Consulting Engineers, "Optimization Study of Mahalat Water Distribution System", Water and Wastewater Company of Markazi Province (in Farsi), 2005.

[15] Y. Bard, Nonlinear parameter estimation, Wiley, New York 1974, p. 174.

[16] A. Taebi, M. R. chamani, *Water Distribution Systems,* 1st ed., Isfahan, Isfahan University of Technology (in Farsi), 2001, pp. 41-51.

[17] T. M. Walski, W. W. Sharp, and F. D. Shields, "Predicting Internal Roughness in Water Mains" Miscellaneous Paper EL-88-2, US Army Engineer Waterways Experiment Station, Vicksburg, Miss, 1988.

[18] Haestad Methods, *WaterGEMS User's Guide,* CT 06708-1499 USA, 2003.

[19] K. N. Mallick, Ahmed, I., Tickle, K. S., and Lansey, K. E. ''Determining pipe groupings for water distribution networks.'' *Journal of Water Resources Planning and Management,* 128(2), pp. 130-139, 2002.