Optimal Sampling Design Methodologies for Water Distribution Model Calibration

Zoran S. Kapelan[1]; Dragan A. Savic[2]; and Godfrey A. Walters[3]

Abstract: Sampling design (SD) for water distribution systems (WDS) is an important issue, previously addressed by various researchers and practitioners. Generally, SD has one of several purposes. The aim of the methodologies developed and presented here is to find the optimal set of network locations for pressure loggers, which will be used to collect data for the calibration of a WDS model. First, existing SD approaches for WDS are reviewed. Then SD is formulated as a multiobjective optimization problem. Two SD models are developed to solve this problem, both using genetic algorithms (GA) as search engines. The first model is based on a single-objective GA (SOGA) approach in which two objectives are combined into one using appropriate weights. The second model uses a multiobjective GA (MOGA) approach based on Pareto ranking. Both SD models are applied to two case studies (literature and real-life problems). The results show several advantages and one disadvantage of the MOGA model when compared to SOGA. A comparison of the MOGA SD model solution to the results of several published SD models shows that the Pareto optimal front obtained using MOGA acts as an envelope to the Pareto fronts obtained using previously published SD models.

DOI: 10.1061/(ASCE)0733-9429(2005)131:3(190)

CE Database subject headings: Sampling; Water distribution; Calibration; Pipe networks; Algorithms.

Introduction

Hydraulic simulation models are widely used by planners, water utility personnel, consultants and others involved in analysis, design, operation, and maintenance of water distribution systems. To make the models useful they must be calibrated (Walski 1983). This is achieved by determining various parameters that, when input into a hydraulic simulation model, yield a reasonable match between measured and predicted pressures and flows in the network (Shamir and Howard 1968).

In general, an optimal sampling design (SD) procedure for water distribution systems (WDS) model calibration should aim to determine: (1) which WDS model predicted variables (pressures, flows, etc.) to observe; (2) where in the WDS to observe them; (3) when to observe them (duration and frequency); and (4) under what conditions to observe them (e.g., demand condition, pumps on/off). The objective is to collect data that, when used for calibration of the model, will yield the best results. In the SD model developed and presented here, it is assumed that (1), (3), and (4) are known in advance. Hence efforts are concentrated on determining the optimal set of WDS measurement locations.

First, background information regarding existing SD approaches for WDS is given. After that, SD objectives and constraints are defined. Two new SD models are then developed and presented. This is followed by application of both SD models to two case studies. Finally, a summary is provided and relevant conclusions are drawn.

Background

Walski (1983) was among the first to suggest where in a WDS to collect data on pressures and flows for model calibration. He suggested observing pressures near points of high demand, preferably on the perimeter of the skeletonized network, away from water sources. Yu and Powell (1994) formulated the so-called meter placement problem for state estimation of WDS, their main objectives being maximization of estimation accuracy, and minimization of metering cost. A method employing dynamic analysis of the covariance matrix of state variables and decision-tree techniques was developed. Ferreri et al. (1994) suggested the selection of measurement points by analyzing the relative sensitivities of nodal heads with respect to roughness calibration parameters. Instead of formulating an optimization problem, they ranked the WDS nodes according to their overall relative sensitivity.

Bush and Uber (1998) developed three new, relatively simple yet efficient SD methods: max-sum, weighted sum, and max-min methods. All are based on analysis of the Jacobian matrix and were derived from so-called *D*-optimality criteria, but do not directly solve the *D*-optimal SD problem. A SD solution is given in the form of ranked WDS nodes. Piller et al. (1999) formulated the SD problem with the objective of minimizing the influence of measurement errors in the state vector estimation subject to the constraint that the Jacobian matrix is of maximum rank. The problem was solved using a greedy algorithm which starts from

[1]Research Fellow, Dept. of Engineering, School of Engineering and Computer Science and Mathematics, Univ. of Exeter, North Park Rd., Exeter, EX4 4QF UK. E-mail: z.kapelan@ex.ac.uk

[2]Professor, Dept. of Engineering, School of Engineering and Computer Science and Mathematics, Univ. of Exeter, North Park Rd., Exeter, EX4 4QF UK. E-mail: d.savic@ex.ac.uk

[3]Professor, Dept. of Engineering, School of Engineering and Computer Science and Mathematics, Univ. of Exeter, North Park Road, Exeter, EX4 4QF UK. E-mail: g.a.walters@ex.ac.uk

Note. Discussion open until August 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on June 4, 2003; approved on March 26, 2004. This paper is part of the ***Journal of Hydraulic Engineering***, Vol. 131, No. 3, March 1, 2005. ©ASCE, ISSN 0733-9429/2005/3-190–200/$25.00.

an empty solution set and sequentially selects and adds measurement locations, one at each algorithm step.

Meier and Barkdoll (2000) were the first to use genetic algorithms (GA) to solve the optimal SD problem. Their objective was to find a (fixed size) set of calibration test locations that produces non-negligible flow rates in as large a portion of the WDS as possible. Three new SD approaches were suggested by de Schaetzen et al. (2000). The first two approaches are based on the shortest path algorithm and rank potential measurement locations. The third approach solves the optimization problem based on maximization of Shannon's entropy. The SD cost is also taken into account, either as a model constraint or as a second objective (in which case the two objectives are normalized and combined using weight coefficients). Most recently, Lansey et al. (2001) developed a sensitivity-based heuristic SD procedure, which uses a first order second moment (FOSM) model to propagate uncertainties in measurements to calibrated WDS model predictions.

Uncertainty Modeling in Calibration of Water Distribution System Models

Calibration Problem

The problem of optimal sampling design is closely related to that of calibration. Calibration of a hydraulic model is formulated here as an optimization problem with an objective function of the weighted least squares type

$$\text{Minimize } E = \mathbf{r}^T\mathbf{W}\mathbf{r} \tag{1}$$

where E = scalar objective function value to be minimized; $\mathbf{W}$ = weight matrix (N_o rows and columns); $\mathbf{r} = \mathbf{y}^* - \mathbf{y(a)}$ = vector of N_o residuals (errors), i.e., differences between observed $\mathbf{y}^*$ and model predicted variables $\mathbf{y(a)}$; $\mathbf{a}$ = vector of N_a unknown calibration parameters; T = vector/matrix transpose operator; N_o = number of measurement data in both spatial and temporal domains; and N_a = number of calibration parameters. There are two sets of constraints: (1) implicit type constraints, consisting of equations representing the analyzed WDS hydraulic model and (2) explicit type constraints used to impose maximum and minimum bounds on the calibration parameter values.

Uncertainty Quantification

During the calibration process, errors (i.e., uncertainties) in pressure/flow measurements are propagated to calibration parameters, resulting in uncertain model predictions. Quantification of the parameter and prediction uncertainties presented here is based on linear regression theory, a method known in the literature as the FOSM model (Bard 1974). Using this model, a first-order approximation of the parameter variance–covariance matrix $\mathbf{Cov}_{\boldsymbol{a}}$ (henceforth referred to as the parameter covariance matrix) is defined as

$$\mathbf{Cov}_{\mathbf{a}} = s^2 \cdot (\mathbf{J}^T\mathbf{W}\mathbf{J})^{-1} \tag{2}$$

where s^2 = calculated error variance; $\mathbf{J}$ = Jacobian matrix of derivatives $\partial y_i/\partial a_k$ ($i = 1, \ldots, N_o$; $k = 1, \ldots, N_a$), i.e., of derivatives calculated for model predictions $\mathbf{y(a)}$ that spatially and temporally correspond to measurements $\mathbf{y}^*$. When all measurements have the same error indicated by standard deviation σ_y, Eq. (2) can be approximated as (Lansey et al. 2001)

$$\mathbf{Cov}_{\mathbf{a}} = \sigma_{\mathbf{y}}^2 \cdot (\mathbf{J}^{\mathbf{T}}\mathbf{J})^{-1} \tag{3}$$

Uncertainty in the calibration parameter values determined is indicated by parameter variances. The variance of the ith calibration parameter is estimated as the value of the ith diagonal element of matrix $\mathbf{Cov}_{\boldsymbol{a}}$.

The WDS model prediction variance-covariance matrix $\mathbf{Cov}_{\mathbf{z}}$ (henceforth, the prediction covariance matrix) can be estimated as (Bard 1974)

$$\mathbf{Cov}_{\mathbf{z}} = \mathbf{J}_{\mathbf{z}} \cdot \mathbf{Cov}_{\mathbf{a}} \cdot \mathbf{J}_z^T \tag{4}$$

where $\mathbf{J}_{\mathbf{z}}$ = Jacobian matrix of derivatives $\partial z_i/\partial a_k$ ($i = 1, \ldots, N_z$; $k = 1, \ldots, N_a$); and $\mathbf{z}$ = vector of N_z model predictions of interest, i.e., vector consisting of chosen model predictions $\mathbf{y}$ whose uncertainties are being evaluated. The variance of the ith model prediction is estimated as the value of the ith diagonal element of matrix $\mathbf{Cov}_{\mathbf{z}}$.

Sampling Design for Water Distribution Systems Model Calibration

Assumptions

In the SD approach presented, the following assumptions and simplifications are made: (1) the only source of errors (uncertainties) is imprecise measurement, i.e., the WDS hydraulic model and all its input parameters (except pipe roughnesses) are assumed error free; (2) the only measurement devices that are subject to optimal placement are pressure loggers, which can only be placed at a number (N_{ml}) of previously identified network nodes; (3) all pressure loggers have identical measurement accuracies defined by standard deviation σ_h; (4) measurement data are collected to calibrate the steady-state hydraulic model under multiple loading conditions for unknown pipe roughness coefficients only; (5) calibrated model accuracy is evaluated by calculating nodal pressure uncertainties only; and (6) the only unknowns in the optimal SD analysis are pressure logger locations: all other data necessary to define the SD problem are known.

Even though some of these assumptions may seem quite restrictive, they are actually not. For example, all pressure loggers do not need to have the same accuracy for the methodologies to work. Where pressure loggers have different accuracies, Eq. (2) should be used instead of Eq. (3) to evaluate matrix $\mathbf{Cov}_{\mathbf{a}}$. Also, the methodologies can easily be extended to determine optimal flow measurement locations in addition to pressure measurement locations, and for any calibration parameter (not just pipe roughness coefficients).

Objectives and Constraints

There are two distinct objectives: (1) maximization of calibrated model accuracy by minimization of calibrated model prediction uncertainty and (2) minimization of total costs associated with SD. A tradeoff between the two objectives exists.

Here, the calibration accuracy objective is formulated as minimization of the average absolute model prediction uncertainty

$$\text{Minimize } F_1 = \frac{1}{N_z}\sum_{i=1}^{N_z} \mathbf{Cov}_{\mathbf{z},ii}^{1/2} \tag{5}$$

where $\mathbf{Cov}_{\mathbf{z},ii}$ = ith diagonal element of the prediction covariance matrix $\mathbf{Cov}_{\mathbf{z}}$ defined by Eq. (4) and N_z = number of model predic-

tions (nodal pressures here) in both spatial and temporal domains for which uncertainties are being evaluated.

From Eqs. (3) and (4) it is obvious that evaluation of F_1 requires calculation of the Jacobian matrices $\mathbf{J}$ and $\mathbf{J_z}$. Note that matrix $\mathbf{J_z}$ needs to be calculated only once, prior to the optimization process. Unlike $\mathbf{J_z}$, matrix $\mathbf{J}$ depends on the selection of measurement locations, and therefore needs constant updating during the search process. However, this does not mean that derivatives in matrix $\mathbf{J}$ have to be calculated for update. Matrix $\mathbf{J}$ can be simply constructed from the full Jacobian matrix $\mathbf{J_{ml}}$ by copying the rows corresponding to the currently analyzed set of measurement locations. $\mathbf{J_{ml}}$ is the Jacobian matrix with rows corresponding to all N_{ml} analyzed sampling design locations and needs to be calculated only once, prior to the optimization process. Elements of the Jacobian matrix $\mathbf{J_{ml}}$ (and $\mathbf{J_z}$, if different from $\mathbf{J_{ml}}$) are calculated using the sensitivity equation method for the best calibration parameter estimates available. Details of this method can be found in Bush and Uber (1998) and Kapelan (2002).

The second objective addresses the problem of SD costs, consisting of capital and operational components. Capital costs may include: investment in new measurement equipment or on-site work associated with installation of measurement devices. Operational costs may relate to labor involved in conducting field tests, equipment maintenance, equipment insurance, electricity, etc. In a particular case it may be possible to estimate costs relatively accurately. However, it is very difficult to generalize them, and derive a general expression to estimate total cost. Here, the following surrogate measure is used as the total SD cost (Yu and Powell 1994; de Schaetzen et al. 2000):

$$\text{Minimize } F_2 = N \tag{6}$$

In addition to the two aforementioned objectives, the following constraint is used:

$$N_{\min} \leq N \leq N_{\max} \tag{7}$$

where N = actual number of measurement devices; $N_{\min}$ = minimum required number of devices ($N_{\min} > 0$); $N_{\max}$ = maximum allowed number of devices; $N_{\max} \leq N_{\text{ml}}$; and N_{ml} = total number of measurement locations to be analyzed when solving a particular SD problem. Note that if all N_{ml} locations are used to evaluate calibration accuracy then $\mathbf{J_z} = \mathbf{J_{ml}}$.

An optimal SD solution should lead to the collection of data that can be used to formulate and solve well-posed calibration problems. The well-posed solution is one that has a Jacobian matrix of full rank (Carrera and Neuman 1986). However, checking this is computationally very demanding, and therefore is not done here during the search process. Instead, a constraint requiring a minimum number of measurement devices $N_{\min}$ is introduced to ensure that the solution obtained will lead to, at least, an overdetermined calibration problem. Since overdeterminedness is no guarantee that the SD solution obtained will produce a well-posed calibration problem, once the search process is stopped, a singular value decomposition analysis (Press et al. 1990) is performed to verify *a posteriori* the optimal SD solution. Constraint (7) on the maximum number of measurement devices $N_{\max}$ is introduced as a surrogate measure for the SD budget.

Sampling Design Procedure

Assuming only pressures are measured, the following procedure should be used for solving the optimal SD problem: (1) define the optimization model input data, by specifying: the set of N_{ml} potential pressure measurement locations, the duration (e.g., 24 h) and frequency (e.g., 15 min) of pressure measurements (for extended period simulation calibration) or the number of characteristic loading conditions (for multiple steady-state simulation calibration), the number, grouping, and estimated values of potential calibration parameters (to estimate values of parameters use either any existing measurements or engineering judgment), and other data (network configuration, demands, etc.); (2) calculate elements of the full Jacobian matrix $\mathbf{J_{ml}}$; if different from $\mathbf{J_{ml}}$, calculate matrix $\mathbf{J_z}$ too; and (3) solve the optimal SD problem using the signle objective genetric algorithm (SOGA) or multiobjective genetic algorithm (MOGA) SD models.

Single Objective Genetic Algorithm Sampling Design Model

Problem Formulation

In the SOGA approach, the SD objective functions are first normalized. To apply single-objective optimization, all normalized functions must be either maximized or minimized. Since the GA library used here maximizes objectives by default, all objectives were normalized following that logic. First objective F_1 defined by Eq. (5) is normalized

$$\text{Maximize } f_1 = \frac{\sum_{i=1}^{N_z} \mathbf{Cov}_{\mathbf{z},\text{ml},ii}^{1/2}}{\sum_{i=1}^{N_z} \mathbf{Cov}_{\mathbf{z},ii}^{1/2}} \tag{8}$$

where $\mathbf{Cov}_{\mathbf{z},\mathbf{ml},ii} = i$th diagonal element of the model prediction covariance matrix $\mathbf{Cov_z}$ calculated for $\mathbf{J} = \mathbf{J_{ml}}$; i.e., for the case where all analyzed WDS locations are being observed. Normalization of the second objective F_2 defined by Eq. (6) is done as follows:

$$\text{Maximize } f_2 = 1 - \frac{N}{N_{\text{ml}}} \tag{9}$$

Since the standard GA cannot handle constraints directly, a penalty function is introduced to handle constraint (7). The penalty function Pe is introduced as

$$\text{Pe} = \begin{cases} \text{pc}_1(N_{\min} - N), & N < N_{\min} \\ \text{pc}_2(N - N_{\max}), & N > N_{\max} \\ 0, & N_{\min} \leq N \leq N_{\max} \end{cases} \tag{10}$$

where pc_1 and pc_2 = arbitrarily chosen, positive penalty constants. The penalty function Pe always has non-negative values. Finally, the SOGA model objective O becomes

$$\text{Maximize } O = \left(\sum_{i=1}^{2} w_i f_i^p \right)^{1/p} - \text{Pe} \tag{11}$$

where p = norm order (typically 1, 2, or ∞); and w_1 and w_2 = weights used to express preferences, chosen to satisfy the condition $\Sigma_{i=1}^{2} w_i = 1$.

Single Objective Genetic Algorithm Solution

A standard GA is used to solve the optimal single-objective SD problem. The GA software library used here was developed at the Univ. of Exeter (Morley et al. 2001). The library has a large number of built-in features: modeling of generational or steady-state GAs; support for various coding schemes (binary, real, integer); various selection schemes (uniform random, rank biased, roulette wheel, tournament, etc.); various replacement schemes

(random, by rank, weakest, etc.); various crossover operators (simple one point, uniform random, arithmetic, etc.); various mutation operators (random by gene, random nonuniform, etc.); optional use of elitism; etc.

When solving the optimal SD problem, two types of GA coding are used here: (1) binary and (2) integer. For binary coding, each of the N_{ml} potential SD locations (network nodes) is represented as a single gene, with gene values of one or zero used to indicate whether a measurement device is/is not placed at a particular location (e.g., network node). When using integer coding, each of the N_{max} measurement devices is represented as a single gene. The gene value indexes the network location at which that particular device is to be placed, with a zero value indicating that the device is not to be placed anywhere. A drawback is that cases may occur in which two or more genes have the same integer value, suggesting that two or more measurement devices should be placed at the same location. This is corrected as follows: (1) when calculating the calibration accuracy objective, it is assumed that only one device is used to collect data at the corresponding location, i.e., all extra devices are ignored; and (2) when calculating the cost objective, all measurement devices are taken into account. This way, solutions with multiple measurement devices linked to the same location are made less desirable since they cost more while having no additional benefit on calibration accuracy.

In the work presented here, binary coding is used when $N_{max} = N_{ml}$, this being typical of theoretical SD analysis on relatively small, artificial networks. Conversely, integer coding is preferred to binary when $N_{max} \ll N_{ml}$, this being.typical of SD analysis on large, real-life networks where only a small subset of locations may be available for use. In this case, the most critical, second part of constraint (7), i.e., $N \leqslant N_{max}$, is handled automatically leading to solutions with zero penalty (not the case for binary coding). As a consequence, the GA search becomes more effective.

Multiobjective Genetic Algorithm Sampling Design Model

Problem Formulation

Instead of transforming the two-objective problem into a single-objective format, a true, two-objective approach can be taken. In the MOGA approach, the two objectives are treated separately. These are: (1) the calibration accuracy objective defined by Eq. (8), and (2) SD cost defined by the surrogate measure in Eq. (6). The calibration accuracy objective is kept in a normalized form, as suggested by Bush and Uber (1998).

Once the MOGA problem is solved, a whole set of optimal solutions defining the tradeoff surface or Pareto-optimal front is available. Constraint (7) is then used to identify the feasible region of the Pareto-optimal front.

Multiobjective Genetic Algorithm Solution

Several MOGA approaches exist today (Veldhuizen and Lamont 1998). The MOGA model presented here is based on Fonseca and Fleming (1993). Generally speaking, a MOGA optimization run is similar to a standard GA run, the most fundamental difference being in the way chromosome fitness values are calculated. Due to limited space available, only a brief overview of the Fonseca and Fleming's MOGA model is presented here (more details can be found in the reference).

Chromosome Coding. The two chromosome coding schemes used in the SOGA SD model, are also used in the MOGA model.

Fitness Evaluation. In the MOGA model, the value of each objective function is calculated for each chromosome in the population. Once these values are obtained, chromosomes are ranked according to Pareto domination rules (Goldberg 1989). Briefly, Solution A is Pareto inferior to Solution B if B is partially better than A, i.e., if B is better than or equal to A in all function dimensions and better in at least one function dimension. Obviously, definition of better depends on whether the function is to be minimized or maximized. Further, Solution A is said to be superior to B if and only if B is inferior to A. Finally, Solutions A and B are noninferior to one another if B is neither inferior nor superior to A. Application of the Pareto domination rules effectively divides the GA population into a number of subpopulations, each containing a number of noninferior chromosomes.

In Fonseca and Fleming's model, each chromosome is assigned a rank value equal to one plus the number of chromosomes in the same population that are dominating that chromosome. Therefore, all solutions in a Pareto optimal front (i.e., subpopulation) are assigned a rank of one. The chromosome fitness value is then calculated as the reciprocal of its rank (indicating that fitter solutions have smaller rank values). However, assigning each chromosome a fitness value based on Pareto ranking alone, does not guarantee that the Pareto optimal set will be uniformly sampled, i.e., that chromosomes will be uniformly distributed along the Pareto optimal front. To maintain population diversity, the fitness value is reduced according to the number of other chromosomes in the proximity, i.e., according to the number of chromosomes occupying the same "niche." This number is known as the chromosome niche count, and is evaluated for each chromosome by calculating its distance (in the multiobjective space) to all other chromosomes in the population and comparing these distances to a prespecified value (called niching radius). If the distance between two chromosomes is smaller than the niching radius, the two chromosomes are considered to share the same niche.

Mating Restriction. To avoid excessive competition between distant members of the population (i.e., production of lethals), Fonseca and Fleming (1993) suggested mating restriction. The concept is based on the idea that two individuals located at the extremes of a search space are unlikely to form a highly fit individual. The implementation of mating restriction follows the logic of niching. An individual (chromosome) will be able to mate with another, if and only if the distance between those two individuals is less than the mating restriction radius (another MOGA model parameter).

Selection Operator. In Fonseca and Fleming's (1993) MOGA approach, a selection operator called stochastic universal sampling (SUS) is used. In a manner similar to proportionate selection, SUS uses a biased roulette wheel. However, in SUS, the wheel is only spun once. The remaining individuals to be selected are then chosen by moving sequentially around the wheel by a predesignated amount. The use of this incremental step means that the mating process is positionally biased between individuals that are situated alongside each other in the old population. According to Fonseca and Fleming, the SUS operator is optimal in terms of both bias and spread. They also claim that SUS is preferred to standard roulette wheel selection because it has lower stochastic errors and, thus, lower genetic drift.

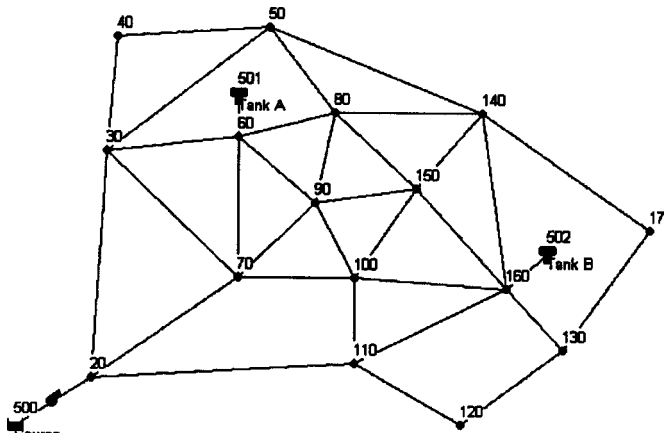

Fig. 1. Case 1: Anytown network layout

Crossover and Mutation Operators. The same crossover and mutation operators used in the SOGA SD model are used here.

Case Studies

Case Study 1: Literature Sampling Design Problem

Problem Description

This case study tests the performance of the two suggested SD approaches on an artificial network known in the literature as the "Anytown" model (Walski et al. 1987). The model has previously been used to test different calibration models (Ormsbee 1989; Lansey and Basnet 1991; Ferreri et al. 1994). The specific objectives of this case study are: (1) to apply and verify both SOGA and MOGA SD models; (2) to compare SOGA and MOGA SD models and identify their advantages and disadvantages; (3) to compare results obtained by the MOGA SD model to results obtained using other published SD models (Ferreri et al. 1994; Bush and Uber 1998; de Schaetzen et al. 2000).

The objective of the SD analysis presented here is to determine the best network locations for collection of pressure data for calibrating, as accurately as possible, the Anytown hydraulic model (see Fig. 1). The steady-state hydraulic model is calibrated for five loading conditions, with model and calibration data taken from Ormsbee (1989).

A total of $N_{ml}=16$ nodes (Nodes 20–170) are considered as possible locations for pressure observations. Fixed head nodes (reservoir and two tanks) are excluded from the analysis. The SD optimization problem is solved for the calibration of the hydraulic model for a total of $N_a=5$ grouped Hazen–Williams (HW) pipe roughness coefficients [the first four parameters are as defined in Ormsbee (1989) while the fifth parameter represents the grouped fifth and sixth parameters from the same paper]. True parameter values are used to obtain the full Jacobian matrix $\mathbf{J_{ml}}$. The matrix has a total of $N_o=80$ rows (16 nodes times 5 loading conditions) and $N_a=5$ columns. Pressure uncertainties are evaluated at all N_{ml} analyzed SD locations ($\mathbf{J_z}=\mathbf{J_{ml}}$).

Case 1.1: Application of Single Objective Genetic Algorithm Model. Analyses were carried out to verify the SOGA SD model. The optimal solution is compared to the best solution found by an enumeration method for two equally weighted SD objectives. Sensitivity of the SOGA model with respect to weights and the order of the Euclidian norm used, are also investigated.

A typical SOGA run and optimal solution are illustrated using the following parameters: $w_1=0.5$, $w_2=0.5$, $N_{min}=1$, $N_{max}=16$, $p=1$, and $pc_1=pc_2=10$. The following GA configuration was used: steady state with population size 50, simple one-point crossover with probability 0.90, and a mutation operator with probability 0.05.

The optimal solution found is to place $N=5$ pressure loggers at the following nodes: 40, 90, 110, 120, and 160. The optimal value of the first normalized function is $f_1=0.699$, the value of the second normalized function is $f_2=0.688$, and the value of the overall objective is $O=0.693$.

Enumeration was then used to solve the same problem. A total of $65{,}535=2^{16}-1$ possible SD solutions were evaluated. The best solution identified is identical to that obtained using the SOGA model.

To analyze the sensitivity of the SOGA optimal solution with respect to weight w_1 (w_2 is always $1-w_1$) and the norm order p, a series of SOGA runs were performed. The results are presented in Table 1. The following are noted: (1) as the weight w_1 increases (and w_2 decreases), parameter accuracy becomes more important than SD cost, causing the optimal number of SD locations to increase; (2) the optimal solution is sensitive to w_1 and p; (3) for $p=2$ and $p=20$ and a relatively large weight increment of 0.1, the behavior of the Euclidian norm emphasizes extreme values of f_1 and f_2, leading to the identification of two extreme SD solutions ($N=N_{min}=1$ and $N=N_{max}=16$, see bottom part of Table 1).

Case 1.2: Application of Multiobjective Genetic Algorithm Model. Analyses were carried out to verify the MOGA SD model for the same SD problem. The following MOGA parameters/settings were used: binary coding, population size 50, uniform crossover with probability 0.7, mutation probability 0.9, niche radius 0.02, and mating restriction radius 0.10.

The optimal Pareto front found by MOGA is shown in Fig. 2, and contains 17 SD solutions. Details about each solution are given in Table 2. Since the solutions for $N=0$ and $N=16$ are trivial, the optimal solution should, based on one's preferences, be chosen from the other 15 available solutions. Note that each point on the Pareto-optimal front can be identified by solving the relevant single objective SD problem. For example, the point on the Pareto front corresponding to $N=5$ can be identified by running the SOGA SD model with the following set of parameters: $w_1=1.0$, $w_2=0.0$, $N_{min}=5$, $N_{max}=5$, and $p=1$. The results from the MOGA SD model were thus verified by formulating and solving 16 independent SOGA SD problems, the results confirming that all points on the MOGA Pareto optimal front are correctly determined.

Finally, from Fig. 2 and Table 2 the following can be noted: (1) a comparatively small number of network nodes need to be observed to achieve relatively high calibration accuracy, i.e., to significantly reduce calibration uncertainty. (2) The most frequently chosen, i.e., best pressure measurement locations are those near locations of high demand and far from fixed-head nodes, as previously suggested by Walski (1983). Indeed, according to the largest number of "ones" in Table 2 (see also last row in Table 3), the best pressure measurement location is Node 90. This node has the highest sum of demands for all five loading conditions and is relatively far from all three fixed-head nodes. The same is true for Node 120. On the other hand, even though it has the second highest total demand, Node 60 is not a desirable measurement location since it is very close to Tank A. Furthermore, Node 170 is desirable even with a relatively low total demand,

Table 1. Case 1.1: Optimal Single Objective Genetic Algorithm Solutions for Different Values of p and w_1

						Network node ID															
p	w_1	O	f_1	f_2	N	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170
1	0.0	0.938	—	0.938	1	Any node															
	0.1	0.848	0.043	0.938	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0.2	0.759	0.043	0.938	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0.3	0.697	0.574	0.750	4	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0
	0.4	0.692	0.699	0.688	5	0	0	1	0	0	0	0	1	0	1	1	0	0	0	1	0
	0.5	0.693	0.699	0.688	5	0	0	1	0	0	0	0	1	0	1	1	0	0	0	1	0
	0.6	0.708	0.806	0.563	7	0	0	1	0	0	0	0	1	1	1	1	0	0	0	1	1
	0.7	0.748	0.908	0.375	10	0	0	1	1	0	0	0	1	1	1	1	1	1	0	1	1
	0.8	0.811	0.967	0.188	13	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
	0.9	0.900	0.993	0.063	15	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1.0	1.000	1.000	0.000	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0.0	0.938	—	0.938	1	Any node															
	0.1–0.4	0.89–0.73	0.043	0.938	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0.5–1.0	0.71–1.00	1.000	0.000	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	0.0	0.938	—	0.938	1	Any node															
	0.1–0.2	0.93–0.92	0.043	0.938	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0.3–1.0	0.94–1.00	1.000	0.000	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Note: "1" means that node should be monitored, "0" opposite.

mainly due to its location on the network edge, relatively far from all fixed head nodes. (3) The optimal set for N measurement locations is not always a superset of that for N-1 locations (e.g., optimal locations for $N=4$ are 40, 90, 110, and 120, while optimal locations for $N=3$ are 90, 120, and 140). Therefore, SD models that use ranking (Ferreri et al. 1994; Bush and Uber 1998) or any other methodology in which the optimal set of N measurement locations is derived from the optimal set of N-1 locations (Yu and Powell 1994; Piller et al. 1999) may fail to identify the best solution.

Case 1.3: Comparison of Different Sampling Design Models. Analyses are carried out to compare solutions obtained by the MOGA model to equivalent solutions obtained by: (1) three Bush and Uber (1998) models; (2) the de Schaetzen et al. (2000) entropy model; and (3) the Ferreri et al. (1994) model.

The three Bush and Uber (1998) SD models (max-sum, max-min, and weighted-sum) were developed with the D-optimality criteria in mind. However, rather than maximizing a determinant of the curvature matrix directly, all models try to achieve that indirectly using a simplified, very fast, numerical procedure. To be fully compatible with the Bush and Uber approach, the full Jacobian matrix is normalized here as originally suggested by the authors.

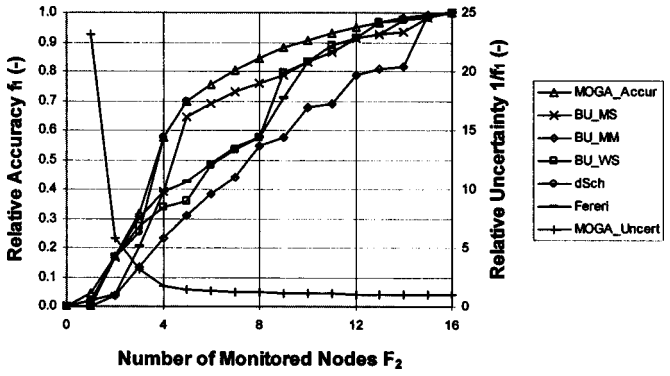

Fig. 2. Cases 1.2 and 1.3: Comparison of different sampling design model solutions

The de Schaetzen et al. (2000) entropy model solves the optimal SD problem using standard GAs. Conceptually, it is very similar to the SOGA SD model. The main difference is in the formulation of the first objective, with calibration accuracy addressed indirectly by maximizing an entropy measure. The entropy model was originally developed and tested for calibration of a steady-state hydraulic model using a single loading condition. However, the Anytown model analyzed here has five loading conditions. To overcome this, each element of the full Jacobian matrix (corresponding to a potential pressure measurement location and calibration parameter) used for the entropy calculations is created by summation of corresponding absolute pressure sensitivities for all loading conditions.

The Ferreri et al. (1994) SD model is also based on the direct analysis of the full Jacobian matrix and is similar to the max-sum model of Bush and Uber (1998). As with the de Schaetzen et al. (2000) model, it was originally developed for a steady-state hydraulic model with a single loading condition. Consequently, elements of the full Jacobian matrix were calculated here by summing up corresponding absolute pressure sensitivities for all loading conditions.

Comparisons of the best solutions obtained using different SD models is done by: (1) comparison of the ranking of analyzed measurement locations and (2) comparison of relevant Pareto fronts for different SD solutions.

While all the Bush and Uber (1998) models and the Ferreri et al. (1994) model directly rank measurement locations, other models do not. To aid comparisons, the optimal measurement locations obtained by the MOGA model are ranked according to their frequency of occurrence in the Pareto-optimal front. The same approach was used for the de Schaetzen et al. (2000) entropy model once the Pareto-optimal front was identified by solving multiple GA optimization problems. Results of the comparisons

Table 2. Case 1.2: Multiobjective Genetic Algorithm Pareto-Optimal Front Points

		Network nodes															
f_1	F_2	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170
0.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.043	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0.173	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
0.319	3	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0
0.574	4	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0
0.699	5	0	0	1	0	0	0	0	1	0	1	1	0	0	0	1	0
0.757	6	0	0	1	0	0	0	0	1	0	1	1	0	0	0	1	1
0.806	7	0	0	1	0	0	0	0	1	1	1	1	0	0	0	1	1
0.846	8	0	0	1	1	0	0	0	1	1	1	1	0	0	0	1	1
0.880	9	0	0	1	1	0	0	0	1	1	1	1	1	0	0	1	1
0.908	10	0	0	1	1	0	0	0	1	1	1	1	1	1	0	1	1
0.932	11	0	0	1	1	0	1	0	1	1	1	1	1	1	0	1	1
0.950	12	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1
0.967	13	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1
0.982	14	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.993	15	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.000	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Note: "1" means that node should be monitored, "0" opposite.

are presented in Table 3. The following can be noted: (1) Optimal ranking of possible measurement locations differs with the SD model used. However, most models identified Nodes 90 and 120 as the two best measurement locations. (2) The maximum number of measurement locations that can be ranked by the de Schaetzen et al. (2000) entropy model is equal to the number of calibration parameters, which is why only the best five locations are ranked here, i.e., all other nodes have the same rank.

Further comparison of the best SD solutions from different models is made using Pareto fronts. If a particular SD model produces ranked measurement locations, it is assumed that the optimal solution for N measurement locations contains the N highest ranked locations. Results of the model comparisons are presented in Fig. 2. The following can be noted: (1) The Pareto front obtained by the MOGA model is the optimal one, i.e., it represents an envelope of the best solutions obtained using all other models. (2) The Pareto front of the Bush and Uber (1998) max-sum model is, in this particular case, in the closest agreement with the MOGA Pareto-optimal front. (3) The Pareto front of the Bush and Uber (1998) max-min model is, generally, the worst. (4) The performance of the Ferreri et al. (1994) model is fairly average and is similar to that of the Bush and Uber weighted-sum (1998) model. This is no surprise, since the two models are similarly formulated. (5) As noted above, the de Schaetzen et al. (2000) entropy model solution ranks only five nontrivial Pareto-front points. However, the points identified are in excellent agreement with the corresponding points from the MOGA solution. (6) As the number of measurement locations increases, the Pareto-optimal fronts of all models are, generally, less different from each other.

Case Study 2: Real-Life Sampling Design Problem

Problem Description

This case study aims to test the performance of the suggested SD approaches on a real WDS. Specifically, the objectives are: (1) to apply and verify use of the proposed SOGA and MOGA models on a real WDS network problem and (2) to compare optimal SD solutions obtained by the SOGA model to solutions suggested by an expert practitioner.

Table 3. Case 1.3: Ranking of Measurement Locations for Different Sampling Design Model Solutions

Single objective genetic algorithm model/ Node ID	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170
Bush and Uber (1998) max-sum	14	13	5	6	16	11	8	1	12	3	2	10	7	9	15	4
Bush and Uber (1998) max-min	16	14	12	10	11	13	9	1	8	15	5	3	7	6	2	4
Bush and Uber (1998) Weighted-Sum	15	14	9	10	16	12	8	1	11	6	2	5	7	4	13	3
Ferreri et al. (1994)	15	13	10	9	16	11	7	4	12	2	1	5	6	8	14	3
de Schaetzen et al. (2000)	6	6	3	6	6	6	6	2	6	4	1	6	6	6	5	6
Multiobjective genetic algorithm	16	15	3	8	13	11	14	1	7	3	2	9	9	12	6	3

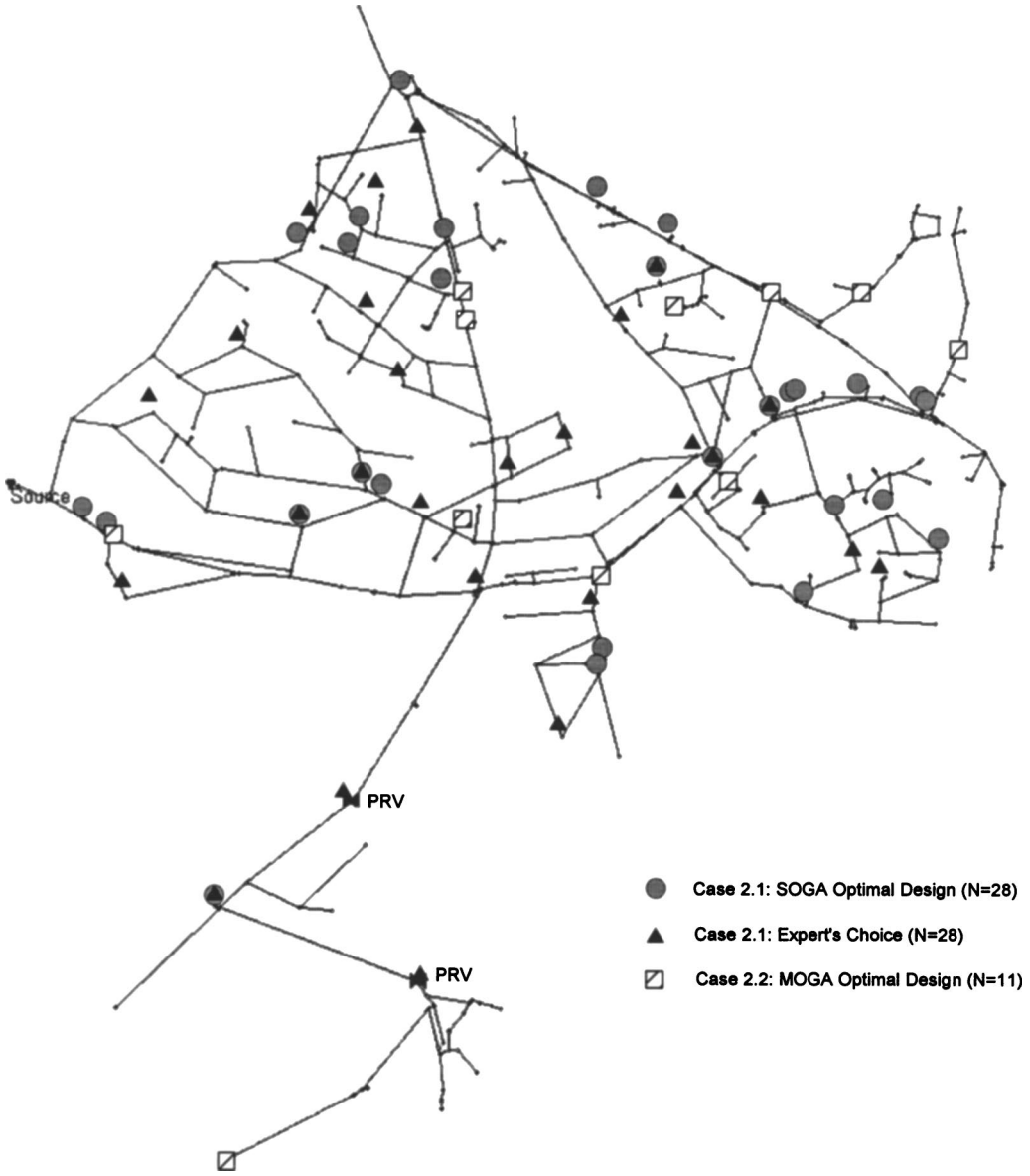

Fig. 3. Case 2: Optimal pressure measurement locations

The SD problem investigated is the calibration of the hydraulic model of a real United Kingdom WDS, studied previously by de Schaetzen (2000). The general layout is presented in Fig. 3. The WDS covers approximately 6 km^2, with a population of around 4,500. Ground levels vary from 54 to 200 m above ordinance datum. Model demands are predominantly domestic with some commercial users to the east. The system is supplied by gravity from a service reservoir (see "Source" in Fig. 3) and includes two pressure reducing valves (PRVs) in the south. The majority of pipes are cast iron or ductile iron. A normal water use field test was carried out on June 29 1994, with an estimated average demand of 14.4 L/S. Based on all available information, an *EPANET* (Rossman 2000) hydraulic model was constructed containing one tank, 451 nodes, 497 pipes, and two PRVs.

A total of 32 loggers were initially used to observe pressures in the system. Evidence of obvious anomalies (e.g., nodal heads higher than the corresponding reservoir level) led to the exclusion of data collected by four loggers. Locations of the remaining 28 pressure loggers are depicted in Fig. 3 (marked with solid triangles), and are hereafter referred to as the expert's choice.

The network is calibrated as a multiple loading condition steady-state hydraulic model. The three characteristic loading conditions analyzed are: (1) maximum demand hour (8 a.m.); (2) average demand hour (3 p.m.), and (3) minimum demand hour (3 a.m.). Sampling design analyses are carried out to collect nodal pressure data for calibrating the model for unknown HW pipe roughness coefficients.

Two cases are analyzed here. First, the single-objective problem is solved for a fixed number of $N=28$ pressure measurement locations. This is done for comparison of the best SOGA model solution to that suggested by the expert. In the second case, the MOGA model is used to solve the multiobjective SD problem. In both cases, pressure uncertainties are evaluated at all analyzed SD locations ($\mathbf{J_z}=\mathbf{J_{ml}}$).

Case 2.1: Comparison of Single Objective Genetic Algorithm Solution to Expert's Choice. Sampling design analyses are carried out to compare the expert's choice for $N=28$ pressure measurement locations (see Fig. 3) with the corresponding set of 28

Table 4. Case 2: Calibration Parameter Values

Parameter ID	Original material	Lining	Diameter (mm)	Case 2.1 Hazen–Williams (HW) roughness coefficient	Case 2.2 HW roughness coefficient
1	Cast iron	None	76	50	25.7
2	Cast iron	None	102	50	46.9
3	Cast iron	None	152	55	40.6
4	Cast iron	None	254	60	63.2
5	Ductile iron	Cement	100	100	101.7
6	Ductile iron	Cement	150	100	98.3
7	Ductile iron	Cement	250	110	111.0
8	Cast iron	Epoxy	76	90	94.7
9	Cast iron	Epoxy	102	—[a]	100[b]
10	Medium density polyethylene	None	73	—[a]	120[b]
11	Medium density polyethylene	None	101	—[a]	125[b]
12	Medium density polyethylene	None	145	130	132.7
13	Polyvinyl chloride	None	102	—[a]	125[b]
14	Polyvinyl chloride	None	152	130	130.2

[a]Parameter does not exist in this case.

[b]Estimated value.

optimal locations obtained using the SOGA model.

A total of $N_{\text{ml}}=210$ (out of 451) network nodes were analyzed as potential pressure measurement locations leading to a large number of possible solutions equal to $\binom{210}{28}\approx 5\times 10^{34}$. These nodes were restricted to those pipes for which calibration was possible using the expert's choice of measurement locations, i.e., on pipes upstream of logger locations. Without this restriction, meaningful comparison of expert and SOGA sampling designs would not be possible. The drawback associated with this approach is that the restricted selection of nodes artificially improves the expert's choice of measurement locations by putting a relatively large number of devices on locations near the network boundaries. However, it will be shown that even in such conditions it is possible to identify a SOGA solution that is better than that suggested by the expert.

It is assumed that the hydraulic model will, once relevant data are collected, be calibrated for $N_a=10$ calibration parameters (see Table 4). Initially, the idea was to define a larger number of calibration parameters (up to one per pipe). However, preliminary analysis showed that under the demand conditions analyzed, pipe roughness coefficients had to be grouped to gain enough information (from measurements). Therefore, HW pipe roughness coefficients were grouped using the "American" criterion (de Schaetzen 2000), according to their material (i.e., lining) and diameter. In this particular case, this was correct since the ages of pipes within a group were approximately the same. When estimating parameter values, it was assumed that existing pressure measurements are not available, since they were not available when the expert's choice of measurement locations was made. Therefore, calibration parameter values, shown in Table 4, were estimated using engineering judgment based on other information available.

Once estimated, parameter values were used to define a full Jacobian matrix $\mathbf{J_{ml}}$ which is a fundamental matrix for any SD analysis. Here, the matrix has a total of $N_o=630$ rows ($N_x=210$ nodes times $N_t=3$ loading conditions) and $N_a=10$ columns. All potential pressure loggers are assumed to be of similar accuracy with a standard deviation of $\sigma_h=1.0$ m.

The following parameters were used in all SD runs: $N_{\min}=N_{\max}=28$, $w_1=1$, $w_2=0$, and $p=1$. The GA settings used were: steady-state GA, integer coding, population size 200, allowed to run for $N_{\text{gen}}=20{,}000$ generations, simple one-point crossover with probability 0.90, and mutation probability 0.05. Multiple GA runs were carried out using different randomly created initial populations. At the end, the best solution was selected. Note that due to the large size of the optimization problem search space, solutions found should be treated as suboptimal, i.e., there is no guarantee that the global optimum is identified.

The set of 28 optimal pressure measurement locations is presented, together with the expert's equivalent choice, in Fig. 3. As can be seen, some clustering of measurement locations occurs. This is mainly a consequence of: (1) the chosen parameter grouping; (2) the locations of sensitive (e.g., high demand) nodes; and (3) possible correlation among sensitivities.

The comparison between the expert's choice and the optimal sampling design is documented numerically in Table 5. The optimal SOGA solution is clearly better than the expert's choice of measurement locations, as indicated by several metrics. On the basis of calibrated model pressure uncertainty, the SOGA design is 22% better than the expert's.

Case 2.2: Application of Multiobjective Genetic Algorithm Model. Analyses were carried out to apply and verify the MOGA SD model using a real problem. The aim was to determine optimal locations for the collection of pressure data to be used later in calibrating a multiple loading condition steady-state hydraulic model.

It is assumed that the network model will, once pressure data are collected, be calibrated for unknown HW pipe roughness coefficients. The HW coefficients were grouped using the same grouping criterion as in Case 2.1, leading to $N_a=14$ unknown calibration parameters (see Table 4). The number of parameter groups is larger here than in the previous case since all network junctions ($N_{\text{ml}}=451$) are considered as potential locations for pressure loggers.

Unlike Case 2.1, when determining values of the calibration parameters, it was assumed that pressure data were available for the 28 measurement locations forming the expert's choice. Therefore, values for 10 of the 14 parameters were identified by cali-

Table 5. Case 2.1: Comparison of Sampling Designs

Metric	Expert design	Best single objective genetic algorithm sdesign	Relative improvement (%)
Relative calibration accuracy f_1 [see Eq. (8)]	0.338	0.423	25
Average pressure prediction uncertainty F_1 (m) [see Eq. (5)]	0.27	0.21	22
Maximum pressure prediction uncertainty (i.e., square root of the largest diagonal element of $\mathbf{Cov_z}$) (m)	1.50	0.97	35
Trace (i.e., sum of the diagonal elements) of $\mathbf{Cov_z}$ (m^2)	88.8	53.8	39

brating the hydraulic model for the three characteristic loading conditions using this existing pressure data (Kapelan 2002). The other four parameters (9, 10, 11, and 13) were estimated using engineering judgment and other available information. Once determined, calibration parameter values (see Table 4) were used in the calculation of the full Jacobian matrix $\mathbf{J_{ml}}$.

To solve the optimal SD problem, multiple MOGA runs were carried out using randomly created initial populations. Note that due to the extremely large search space (order of 10^{46}), solutions should be treated as suboptimal, i.e., there is no guarantee that the Pareto-optimal front is identified. The following model settings were used in all optimization runs: integer coding, population size 100, uniform crossover with probability 0.7, mutation probability 0.9, niche radius 0.05, and mating restriction radius 0.25. All runs were stopped after 10,000 generations. Finally, note that in this case, typical MOGA runs took approximately 2 h (on a 3 GHz PC computer) to converge.

The best Pareto front identified for N between $N_{\min}=10$ and $N_{\max}=30$ is presented in Fig. 4. For each point on the Pareto-optimal front, a corresponding set of optimal locations for pressure loggers exists, which are not shown here due to limited space. From the Pareto front, an optimal SD solution can be selected by either fixing the number of measurement devices N according to the budget limit, or by specifying the maximum allowed pressure uncertainty (F_1).

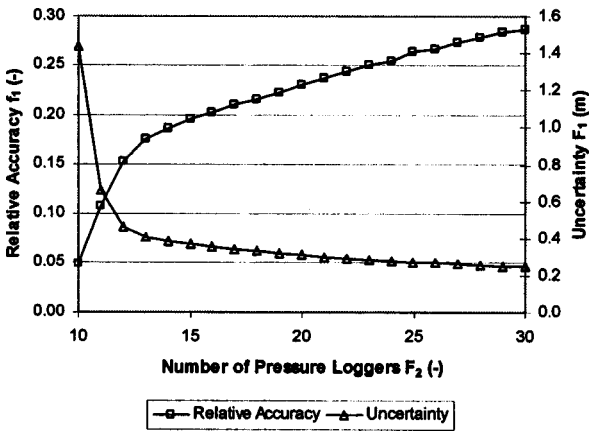

Fig. 4. Case 2.2: Pareto-optimal multiobjective genetic algorithms solution

Fig. 4 shows that for the number and grouping of calibration parameters defined here, only 11 pressure loggers (a total of 33 pressure measurements) are necessary to achieve average pressure prediction uncertainty below 1.0 m. The 11 optimal locations are presented in Fig. 3, which shows that the logger locations are reasonably well distributed throughout the network.

Summary and Conclusions

The objectives and constraints used to solve the problem of optimal SD for calibration of WDS models are first identified. The two conflicting objectives are: (1) maximize calibration accuracy by minimizing some overall model prediction or parameter uncertainty measure and (2) minimize the total SD cost. Two new SD methods are then presented. In the first, the two objectives are normalized and the optimal SD transformed into a single objective problem, which is then solved using a standard, SOGA. In the second, the SD problem is solved as a true multiobjective problem using MOGA based on Pareto ranking. Both SD methodologies were verified on case studies including both literature and real-life problems.

From the case studies, the main conclusions when comparing SOGA and MOGA SD models are: (1) SOGA can detect one optimal solution in a single GA run while MOGA can detect a whole set of (Pareto) optimal solutions, i.e., it can detect the whole tradeoff surface. Consequently, multiple SOGA runs are necessary to obtain the same level of information that can be obtained from a single MOGA run. (2) When SOGA is used, preferences toward different objectives need to be specified before a model run, while in MOGA preferences can be specified after a run, thereby allowing much greater flexibility for decision making. (3) A penalty function must be implemented in SOGA to handle relevant constraints, this being unnecessary in MOGA. Typically, penalty functions make the search more difficult. (4) The main drawback of the MOGA model when compared to SOGA is that it requires additional search parameters whose values need to be tuned for optimal performance.

The main conclusions in comparing the MOGA SD model solution against SD model solutions from the literature (Ferreri et al. 1994; Bush and Uber 1998; de Schaetzen et al. 2000) are: (1) The MOGA SD model is preferred since it addresses explicitly the calibrated model accuracy using a model prediction uncer-

tainty measure rather than implicitly using entropy (de Schaetzen et al. 2000) or some sensitivity measure (Ferreri et al. 1994; Bush and Uber 1998). Consequently, the MOGA model identifies the best tradeoff curve, this being an envelope of equivalent curves obtained from the other SD models mentioned above. (2) The MOGA model is preferred to ranking type methods (Ferreri et al. 1994; Bush and Uber 1998) because the optimal SD for N measurement points is not always a superset of the optimal SD for N-1 points. Therefore, SD models that use ranking, or other methodologies, in which the optimal set for N locations is derived from the set for N-1 locations (Yu and Powell 1994; Piller et al. 1999) may fail to identify the optimal SD solution. (3) When compared to other published SD models, the obvious drawback of the MOGA (and SOGA) model is the computation requirement (see Case 2.2). This is the price that must be paid for relatively accurate, direct evaluation of the relevant uncertainties (typically including matrix inversion or determinant calculation). This may be an obstacle when analyzing large real network models, especially with large numbers of calibration parameters. However, it is envisaged that with constant increases in computational power, this will be less of a problem in the future.

Acknowledgment

Results presented in this paper were obtained within a research project under the United Kingdom Engineering and Physical Sciences Research Council Grant No. GR/M66981/01, which is gratefully acknowledged.

Notation

The following symbols are used in this paper:

$\mathbf{a}$ = vector of calibration parameters;
$\mathbf{Cov_a}$ = parameter variance–covariance matrix;
$\mathbf{Cov_z}$ = model prediction variance–covariance matrix;
E = calibration objective value;
F_1 = first sampling design objective;
$F_{1,\mathrm{ml}}$ = value of F_1 assuming that all analyzed locations are monitored;
F_2 = second sampling design objective;
f_1 = normalized first sampling design objective;
f_2 = normalized second sampling design objective;
$\mathbf{J}$ = Jacobian matrix;
$\mathbf{J_{ml}}$ = full Jacobian matrix (all locations monitored);
$\mathbf{J_z}$ = prediction Jacobian matrix;
N = actual number of measurement devices;
N_a = number of calibration parameters;
$N_{\max}$ = maximum allowed number of measurement devices;
$N_{\min}$ = minimum number of measurement devices;
N_{ml} = number of analyzed sampling design locations;
N_o = number of observations (i.e., measurements);
N_z = number of model predictions for whom uncertainties are evaluated;
p = Euclidian norm order;
$\mathbf{r}$ = residual vector;
s = calculated error standard deviation;
$\mathbf{W}$ = calibration weight matrix;
$\mathbf{y}$ = model predictions vector;
$\mathbf{y}^*$ = measurement vector;
$\mathbf{z}$ = vector of model predictions of interest to optimal sampling design;
σ_h = standard deviation of measured pressure; and
σ_y = standard deviation measured value y^*.

References

Bard, Y. (1974). *Nonlinear parameter estimation*, Wiley, New York.
Bush, C. A., and Uber, J. G. (1998). "Sampling design methods for water distribution model calibration." *J. Water Resour. Plan. Manage.*, 124(6), 334–344.
Carrera, J., and Neuman, S. P. (1986). "Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability and solution algorithms." *Water Resour. Res.*, 22(2), 211–227.
de Schaetzen, W. (2000). "Optimal calibration and sampling design for hydraulic network models." PhD thesis, Univ. of Exeter, Exeter, U.K.
Ferreri, G. B., Napoli, E., and Tumbiolo, A. (1994). "Calibration of roughness in water distribution networks." *Proc., 2nd Int. Conf. on Water Pipeline Systems*, BHR Group, Edinburgh, U.K., 379–396.
Fonseca, C. M., and Fleming, P. J. (1993). "Genetic algorithms for multiobjective optimisation: Formulation, discussion and generalisation." *Proc., 5th Int. Conf. on Genetic Algorithms*, S. Forrest, ed., Morgan Kaufman, Mateo, Calif., 416–423.
Goldberg, D. E. (1989). *Genetic algorithms in search, optimisation and machine learning*, Addison-Wesley, Reading, Mass.
Kapelan, Z. (2002). "Calibration of WDS hydraulic models." PhD thesis, Univ. of Exeter, Exeter, U.K.
Lansey, K. E., and Basnet, C. (1991). "Parameter estimation for water distribution networks." *J. Water Resour. Plan. Manage.*, 117(1), 126–144.
Lansey, K. E., El-Shorbagy, W., Ahmed, I., Araujo, J., and Haan, C. T. (2001). "Calibration assessment and data collection for water distribution networks." *J. Hydraul. Eng.*, 127(4), 270–279.
Meier, R. W., and Barkdoll, B. D. (2000). "Sampling design for network model calibration using genetic algorithms." *J. Water Resour. Plan. Manage.*, 126(4), 245–250.
Morley, M. S., Atkinson, R. M., Savic, D. A., and Walters, G. A. (2001). "GAnet: Genetic algorithm platform for pipe network optimisation." *Adv. Eng. Software*, 32(6), 467–475.
Ormsbee, L. E. (1989). "Implicit network calibration." *J. Water Resour. Plan. Manage.*, 115(2), 243–257.
Piller, O., Bremond, B., and Morel, P. (1999). "A spatial sampling procedure for physical diagnosis in a drinking water supply network." *Proc., Water Industry Systems: Modelling and Optimisation Applications*, D. A. Savic and G. A. Walters, eds., Vol. 1, Research Studies Press Ltd., Exeter, U.K., 309–316.
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1990). *Numerical recipes: The art of scientific computing*, Cambridge Univ. Press, Cambridge, U.K.
Rossman, L. A. (2000). *Epanet2 users manual*, US EPA, Washington, D.C.
Shamir, U., and Howard, C. D. D. (1968). "Water distribution systems analysis." *J. Hydraul. Div., Am. Soc. Civ. Eng.*, 94(1), 219–234.
Veldhuizen, D. A. V., and Lamont, G. B. (1998). "Multiobjective evolutionary algorithm research: A history and analysis." *TR-98-03*, Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.
Walski, T. M. (1983). "Technique for calibrating network models." *J. Water Resour. Plan. Manage.*, 109(4), 360–372.
Walski, T. M., et al. (1987). "Battle of network models: Epilogue." *J. Water Resour. Plan. Manage.*, 113(2), 191–203.
Yu, G., and Powell, R. S. (1994). "Optimal design of meter placement in water distribution systems." *Int. J. Syst. Sci.*, 25(12), 2155–2166.
de Schaetzen, W., Walters, G. A., and Savic, D. A. (2000). "Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms." *Urban Water*, 2, 141–152.