OPTIMAL MEASUREMENT SITE LOCATIONS FOR INVERSE TRANSIENT ANALYSIS IN PIPE NETWORKS

John P. Vítkovský[1], James A. Liggett[2], Angus R. Simpson, M. ASCE[3],

and Martin F. Lambert[4]

ABSTRACT

The quality of leak detection and quantification, and calibration for friction coefficients, in pipelines and networks by the inverse transient method are dependent on the quantity and location of data measurement sites. This paper presents an approach for determining the configuration of measurement sites that produces optimal results. Three performance indicators, two that are based on A- and D-optimality criteria and one that is based on the sensitivities of the heads with respect to the parameters, show which configurations are superior. These are illustrated by two case studies, the first of which is a small pipe network in which all configurations are considered directly (fully enumerable) and the second is a larger pipe network in which statistics are drawn from a sampling of configurations. For the large network, a genetic algorithm—with a new crossover operator—performs a search of

[1] Research Associate, School of Civil and Environmental Engineering, University of Adelaide, Adelaide SA 5005, Australia. (Corresponding Author)
Email: jvitkovs@civeng.adelaide.edu.au; Tel: +61 8 8303 4324; Fax: +61 8 8303 4324

[2] Professor Emeritus, School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853-3501, USA. Email: jal8@cornell.edu

[3] Associate Professor, School of Civil and Environmental Engineering, University of Adelaide, Adelaide SA 5005, Australia. Email: asimpson@civeng.adelaide.edu.au

[4] Senior Lecturer, School of Civil and Environmental Engineering, University of Adelaide, Adelaide SA 5005, Australia. Email: mlambert@civeng.adelaide.edu.au

possible measurement site configurations to determine the optimal measurement locations. The number of sites as well as time length of data at each site are considered also.

INTRODUCTION

Inverse analysis has been applied in a variety of fields to determine parameters of problems, boundary conditions and even the basic equations governing a process. By definition an inverse problem is one where measurements of one or more events are known but the parameters defining the physical condition, the boundary or initial conditions, and/or the governing equations are unknown. Inverse analysis has been applied under transient conditions to leak detection and friction factor calibration in pipelines and pipe networks by Liggett and Chen (1994) and Vítkovský (2001). It typically requires a large quantity of data for accurate calculation, and an unsteady event provides much more data than a steady event.

The basic objective of inverse analysis in a piping system is to find leaks, but Liggett and Chen (1994) noted that unless the frictional properties are well known—which is seldom the case—leak detection and quantification could not be carried out with sufficient precision. Thus, the analysis requires a simultaneous calibration for friction factors and leak areas. In addition, wave speed in a pipe is seldom known accurately and that factor is often included in the sought-for parameters. Although the primary objective has been leak detection and quantification, the calibration aspect forms a major side benefit as the frictional properties are required for the analysis of a network, design of additional infrastructure and maintenance of networks.

A transient event in a pipeline system can be generated using a change in valve or pump conditions. The measured data are the pressures observed periodically and simultaneously at various locations in a pipe or pipe network during the transient event. Since flow rates are more difficult and expensive to measure they are not generally used.

There are two methods to obtain an inverse solution, a direct and an indirect method. Each has advantages and disadvantages (Neuman 1973). The direct method treats the model parameters as the dependent variables in a formal inverse boundary value problem from which a direct solution of the parameters is made. The direct method requires that the data (and derivatives of the data) are exact and complete. Errors in the data may cause the problem to become improperly posed and solutions might only exist for certain restricted conditions. The indirect method minimizes the difference between measured and calculated data. Essentially, the indirect method uses a "guided search" by a minimization algorithm. An advantage of the indirect method is that data (and derivatives of the data) need not be known at all points and times in the network. In pipe networks, the data are measured at particular locations and certainly not over the entire domain of the dependent variables (required by the direct method). Hence, the indirect method is used in this research.

Accuracy of the inverse method is very dependent on the quality and quantity of measurements. However, all measurements are not equally effective. The objective of this paper is to explore data collection methods that make transient inverse analysis effective and economical. Questions include:

1. How many measurement sites are necessary in a network for adequate results?
2. Where should these measurement sites be placed to produce the best results?
3. What (time) length of data is needed to produce accurate results?

4. What degree of confidence is associated with the results?

The optimum location of measurement sites is a combinatorial problem, i.e., for a given number of measurement sites there are many combinations of site configurations. Two case studies are considered. One is a completely enumerable network (meaning that all possible measurement sites and measurement configurations are considered); the other is a large network where full enumeration is not practicable.

The "optimal sampling design" consists of a plan of measurement sites that optimize the inverse solution. Historically, the field of groundwater monitoring has made considerable inroads into optimal sampling design. Carrera and Neuman (1986) suggested the reduction of parameter variances (A-optimality criterion) be used to determine the optimal locations to make measurements. Knopman and Voss (1989) optimized the accuracy to which the parameters are determined, cost of sampling and even the type of model used. Their optimization contained multiple objectives and produced an optimal front of solutions. Loaiciga *et al*. (1992) give a review of groundwater sampling design.

In the context of this paper, sampling design is applied to water distribution systems. Walski (1983) suggested some rules-of-thumb for the steady state calibration of water distribution systems based on practical experience. Yu and Powell (1994) determined optimal sampling designs using a decision-tree technique for optimization based on the A-optimality criterion, sampling cost and distance from the sampling locations to a control center. Bush and Uber (1998) used a ranking of three different criteria based on sensitivities to generate near optimal sampling designs for calibration. Their results compared well with the D-optimality criterion. Meier and Barkdoll (2000) considered the calibration problem using a number of different

flow tests generated by the opening of fire hydrants. Genetic algorithms determined where and what number of hydrants were to be opened to ensure satisfactory network coverage. De Schaetzen et al. (2000) defined a single objective function from a weighted combination of Shannon's entropy (related to sensitivity) ands the cost of measurement. Genetic algorithms were used to optimize the sampling design problem.

All approaches for optimal sampling design (including this paper) depend on (i) the decision of what constitutes a good sampling design, (ii) definition of criteria that indicate a good sampling design, and (iii) an optimization technique to determine the optimal design or optimal front of sampling designs. This paper approaches the selection of measurement sites for the inverse transient problem in a similar fashion. Indicators are defined that assess the performance of particular sampling designs for the inverse transient problem. Additionally, the calculation of these indicators should be efficient. Optimal sampling designs are determined in the following sections for a fully enumerable network and a larger network, where full enumeration is not practical, in which the search for the optimal design is performed using a genetic algorithm.

PERFORMANCE INDICATORS

Liggett and Chen (1994) suggested that sensitivities with respect to the parameters be used to decide where measurement sites should be placed, i.e., that points of measurement be located where the desired parameters are sensitive to the measurements. Secondly, they suggested that the degree of confidence that one has in an inverse-transient result largely depends on the sensitivity. If in the forward problem, the variables at measurement points are insensitive to

the parameters, then a large change in the parameters has little effect on the variables, or conversely, a small error in measurement has an unreasonably large effect on the parameters and thus those parameters will be poorly determined.

Liggett and Chen (1994) used two measures of sensitivity for the inverse transient problem. The first uses an objective function, *E*, which is frequently expressed as the sum of the of the squares of the differences of one or more observed variables (usually head, *H*) and the calculated variables based on a set of values of the parameters (e.g., $E=\Sigma(H^m-H_{calc})^2$ where H^m is measured head and H_{calc} is calculated head). The sensitivity with respect to each of the parameters can be measured by the partial derivatives of the objective function, *E*, with respect to each of the parameters, a_j

$$M_j^{(1)} = \left\{ \frac{\partial E}{\partial a_j} \right\} \tag{1}$$

which is the gradient vector (length N_P, where N_P is the number of parameters) of the objective function. These derivatives represent a measure of the convergence rate towards the inverse transient solution and also indicate the slope of the objective function surface. For small values of the gradients, a large range of parameter values will satisfy the minimizing criterion, and thus the parameters will not be well determined under experimental error. However, at the point where the inverse transient method has found a minimum, $M_j^{(1)} = 0$; therefore, it is not a particularly useful quantity for determining probable inverse transient performance.

A second sensitivity measure is the partial derivative of the head, *H*, in each of the sampling points with respect to the parameters,

$$M_{ij}^{(2)} = \left[\frac{\partial H_i}{\partial a_j} \right] \tag{2}$$

in which $M_{ij}^{(2)}$ is a Jacobian (size $N_P \times N_M$, where N_M is the total number of measured data points) and H_i is the head. $M_{ij}^{(2)}$ indicates directly the influence of a parameter on the head in both the temporal and spatial domains. For example, if the pressure at a leak is low in a region of the pipe network, then the sensitivity of the heads in that region with respect to a leak is lower than a similarly sized leak in a higher pressure region due to the fact that a leak in low pressure will discharge less fluid than a leak under high pressure. Thus, a leak in a low-pressure region will be found with less precision than a similarly sized leak in a high-pressure region. As a general rule, measurement points should be placed in locations of high values of $M_{ij}^{(2)}$.

In addition to the indicators defined by Eqs. 1 and 2, an alternative indicator of inverse transient analysis performance is given by the probable error in the solution parameters. The size of the probable error depends on a number of factors such as the parameters to be determined, length of measured data and type of transient test, but arguably the most important factors are the number and location of measurement sites. By choosing measurement site configurations that minimize the probable error in the parameters, a superior inverse transient result is ensured. For steady state systems, the analysis of error using a first-order error approximation has been performed previously for pressure and flow uncertainty (Bargiela and Hainsworth 1989), for reliability (Xu and Goulter 1998) and for calibration (Lansey *et al*. 2001). The inverse transient problem can be defined as $\{a\} = f(\{H^m\})$ where the elements of the parameter vector, $\{a\}$, are a function of the measured heads, $\{H^m\}$. Using maximum likelihood estimators (Press *et al*. 1992), the chi-square statistic for a least-squares fit is

$$E = \sum_{i=1}^{N_M} \frac{\left(H_i^m - H_i\right)^2}{\sigma^2_{H_i^m}} \tag{3}$$

where $\sigma^2_{H^m}$ = variance of the measured head data. Using a first order Taylor Series expansion about the correct parameter values, the error propagation equation for a parameter a_j is

$$\sigma^2_{a_j} = \sigma^2_{H_1^m}\left(\frac{\partial a_j}{\partial H_1^m}\right)^2 + \sigma^2_{H_2^m}\left(\frac{\partial a_j}{\partial H_2^m}\right)^2 + \ldots + 2\sigma_{H_1^m H_2^m}\left(\frac{\partial a_j}{\partial H_1^m}\right)\left(\frac{\partial a_j}{\partial H_2^m}\right) + \ldots \tag{4}$$

If the errors are uncorrelated with each other then the error propagation equation reduces to

$$\sigma^2_{a_j} = \sum_{i=1}^{N_M} \sigma^2_{H_i^m}\left(\frac{\partial a_j}{\partial H_i^m}\right)^2 \tag{5}$$

Typically, the partial derivatives in Eq. 5 cannot be calculated easily. If, however, the minimization is accomplished by a method that uses the Hessian of *E*, an efficient method of calculation is available. The curvature matrix of the objective function, [κ], is defined as half the Hessian matrix (Press *et al*. 1992)

$$[\kappa] = \left[\frac{1}{2}\frac{\partial^2 E}{\partial a_j \partial a_k}\right] \tag{6}$$

The covariance matrix of the parameters, [*C*], is defined such that it is equal to the inverse of the curvature matrix of the objective function,

$$[C] = [\kappa]^{-1} \tag{7}$$

The variance of the error in parameter a_j is the j^{th} diagonal element of [*C*],

$$\sigma^2_{a_j} = C_{jj} \tag{8}$$

Similarly, the covariance of the errors between two parameters a_j and a_k is the element

$$\sigma_{a_j a_k} = C_{jk} \tag{9}$$

[*C*] is commonly called the covariance matrix. An element of the Hessian matrix, used in the calculation of [*C*], can be estimated efficiently from Jacobian elements using

$$\frac{\partial^2 E}{\partial a_j \partial a_k} \approx 2\sum_{i=1}^{N_M} \frac{1}{\sigma^2_{H_i^m}} \frac{\partial H_i}{\partial a_j} \frac{\partial H_i}{\partial a_k} \tag{10}$$

Liggett and Chen (1994), Nash and Karney (1999) and Vítkovský (2001) give methods of efficiently calculating the Jacobian, although Nash and Karney's method was only applied to a two-pipe series pipeline. If similar measurement devices are used at different measurement sites, the variance in the errors of the measured signals will be similar. Based on this assumption, both the curvature and covariance matrices may be normalized using the common variance in the measured data. The resulting normalized curvature matrix approximation $[\kappa^*]$ is

$$\left[\kappa^*\right] = \sigma^2_{H^m}\left[\kappa\right] \approx \left[\sum_{i=1}^{N_M} \frac{\partial H_i}{\partial a_j} \frac{\partial H_i}{\partial a_k}\right] \tag{11}$$

and the normalized covariance matrix approximation $[C^*]$ is

$$\left[C^*\right] = \frac{\left[C\right]}{\sigma^2_{H^m}} \approx \left[\sum_{i=1}^{N_M} \frac{\partial H_i}{\partial a_j} \frac{\partial H_i}{\partial a_k}\right]^{-1} \tag{12}$$

The normalized covariance matrix element C^*_{jj} represents the measurement error transmission to the parameter error in parameter a_j. Both $[\kappa^*]$ and $[C^*]$ are independent of both the measured head data, H^m, and the error in the measured head data, $\sigma^2_{H^m}$.

In the following section, the quantities in Eqs. 2, 11 and 12 are used to define a performance indicator for a particular measurement site configuration.

OPTIMAL PERFORMANCE

The performance of the inverse transient method for a particular configuration of measurement sites can be estimated using a performance indicator, η_J, based on the Jacobian of the heads in the objective function,

$$\eta_J = \sum_{k=1}^{N_P} \sum_{j=1}^{N_S} \sum_{i=1}^{N_L} \left| \frac{\partial H_{ij}}{\partial a_k} \right| \tag{13}$$

where N_S = number of measurement sites and N_L = number of measurements per site. Note that the definition of the head (H_{ij}) in Eq. 13 differs from the definition of the head (H_i) in the preceding equations such that H_i represents all measurements (i = 1, …, N_M) and H_{ij} represents all measurements on a site-by-site basis (i = 1, …, N_L and j = 1, …, N_S where N_M = $N_S \times N_L$). For the indicator η_J, the optimal set of measurement sites is determined when η_J is maximized. A second performance indicator, η_A, based on the variance of the parameter errors is

$$\eta_A = \text{trace}\left[C^*\right] = \sum_{j=1}^{N_P} C^*_{jj} \tag{14}$$

For η_A, the optimal set of measurement sites is achieved when the indicator is minimized. Although η_A only includes the diagonal elements of $[C^*]$, corresponding to the variance of the error in each parameter, other definitions might also include the covariances. An alternative indicator definition is $\max(C^*_{jj})$ where the maximum diagonal element of $[C^*]$ is minimized when searching for optimal measurement site configurations. Possibly $[C^*]$ might not exist for parameters that have no influence on the heads at the locations of the measurements. In that case, the partial derivatives $\partial H/\partial a$ are zero for a particular parameter, creating both a column and a row of zeros in the Hessian so that it is singular and cannot be inverted. However, the use of singular-value decomposition (SVD, see Press *et al*. 1992) can be used to

solve for those parameters that do have an influence on the head where measurements are made. Conditions where $[C^*]$ does not exist are those when the inverse problem is under-determined or mixed-determined, for example when using short-duration transients in a large pipe network so that a signal from a leak does not have time to reach a measurement point during the period of measurement.

The η_A indicator is an A-optimality criterion where the objective is to minimize the sum of the estimated variances of the model parameters (or minimize η_A). An alternative to an A-optimality criterion is a D-optimality criterion, which maximizes the determinant of $[\kappa^*]$ (Bush and Uber 1998),

$$\eta_D = \det\left[\kappa^*\right] \tag{15}$$

An advantage of the D-optimality criterion is that $\det[\kappa^*]$ is more efficiently calculated than $[\kappa^*]^{-1}$. Maximization of η_D also tends to reduce the correlation between parameters and is not susceptible to the dominance of large parameter variances, as compared with η_A. For these reasons η_D is usually preferred over η_A. If the inverse problem is indeterminate for some parameters but not all parameters (as is the case for an inverse transient problem with short lengths of measurement data), then $\det[\kappa^*]$ will equal zero resulting in an impossible minimization problem. Again, SVD is used to calculate $\text{rank}[\kappa^*]$, which determines the number of solvable parameters. If $\text{rank}[\kappa^*] < N_P$ then some parameters are unsolvable. After finding which of the parameters can be determined, η_D is calculated using only the solvable parameters. Note that inverse transient parameters can be indeterminable for three reasons, the first being that the sensitivity of head at the measurement sites with respect to a parameter is zero (short lengths of data or a network configuration in which a signal is not transmitted to some parts). The second is that there is a perfect correlation between two parameters resulting in an infinite number of solutions for those two parameters. The third is that the

sensitivity of a parameter is very low (but not zero), such that either under measurement error the parameter is undeterminable or round-off errors due the machine accuracy fully contaminate the determination of that parameter. In these situations, where some parameters are not determinable, the optimal sampling design problem using the indicators previously defined is not well-posed and should only be attempted subject to all parameters being solvable. Additionally, if some parameters are indeterminable then any solution must be sub-optimal.

Only heads with corresponding measurements are used in optimization in this paper, although it would be possible to do it for flows also if they could economically be measured instantaneously. Determination of both leak areas and friction factors form a minor complication in that the sensitivities with respect to the leaks are typically three orders of magnitude greater than those with respect to the friction factors. This mismatch in sensitivities means that optimization will be carried out for the leak parameters rather than for both parameter types. Simply scaling the sensitivities is equivalent to using a weighting function that emphasizes any selected parameter. Examples of scaling factors are the quantities $|\partial H/\partial a|$ or $[C^*]$. The use of prior estimates of parameters also can improve the performance of inverse transient solutions (Kapelan *et al*. 2001) by better conditioning the inverse transient problem.

The following section determines optimal measurement sites configurations for a small, "fully enumerable" pipe network.

FULLY ENUMERABLE EXAMPLE

The network in Figure 1 has been previously investigated in other studies and is based on one in Pudar and Liggett (1992) and Liggett and Chen (1994). The network is partly gravity fed with a constant inflow of 20 L/s at node 7 and consists of eleven pipes and seven nodes. All of the pipes share a common diameter of 254 mm, a common length of 762 m and a common wave speed of 1,316 m/s. Leaks are assumed to occur at the nodes except at the reservoir. A valve at node 4 was partially closed at a time of 2.0 s, reducing the initial outflow from 58 to 28 L/s in a linear fashion in 10 s then opened restoring the flow to 58 L/s in another 10 s. All pipes have a Courant number of unity so no interpolation is necessary, thus minimizing the numerical error. The "measured" pressure heads are obtained using a transient solver and are specified at nodes 2, 3, 4, 5, 6 and 7 (node 1 being a constant head reservoir). The variation in the head during the transient at the possible measurement sites is shown in Figure 2.

The analysis presented in the following sections establishes how the inverse transient method should be best applied for the small example network. Since the true leak sizes and locations, if any exist, are not known before the inverse transient method is implemented, the analysis assumes equally sized small leaks at all nodes except node 1.

The Objective

The search for the optimal location of measurement sites requires an objective function, F, to distinguish between different configurations. In the case of the indicators η_J, η_A and η_D, F is optimized so that

$$F = \max(\eta_J) \quad \text{or} \quad \min(\eta_A) \quad \text{or} \quad \max(\eta_D) \tag{16}$$

All of these objectives can be used to find preferred measurement site configurations that maximize confidence in the resulting parameters. The number of possible measurement site combinations is

$$N_C = \binom{N_{TS}}{N_S} = \frac{N_{TS}!}{N_S!(N_{TS} - N_S)!} \tag{17}$$

where N_{TS} = total number of possible measurement sites and N_S = number of measurement sites selected. For a small network the number of measurement site combinations is small enough to be enumerated, but for large networks the calculation time for complete enumeration becomes impracticable. If all numbers of measurement sites are considered then the total number of possible configurations is $2^{N_{TS}} - 1$. In the current example the total number of measurement site combinations is 63, all shown in Table 1.

Optimal Location of Measurement Sites

Using the indicators η_J, η_A and η_D for particular measurement site configurations, the example network was completely enumerated for all configurations of measurement sites. A total measured data length of 40.0 s is used in the analysis. Table 2 shows the optimal measurement site configurations for each number of measurement sites based on both indicators. Figure 3 shows the optimal front that allows the selection of optimal measurement site configurations given a certain number of measurement sites for the data in Table 2.

Figure 3 indicates that more measurement sites leads to a better solution. However, the incremental improvement in the solution falls rapidly as the number of sites increases. In any practical problem, the user must decide whether the improvement in the solution is worth the effort of extra measurements. The analysis leading to Figure 3 will aid in such a decision.

The consideration of each parameter's component in η_J and η_A indicates the confidence with which that parameter will be determined from the inverse transient method. For example, each parameter's component in both η_J and η_A for the example network using all 40.0 s of measurement data is shown in Figure 4. Figure 4 shows that for both the η_J and η_A indicators the parameter corresponding to a leak at node 4 would be more accurately determined than the other parameters. Additionally, if the variance in the measurement error is known then the variance in the solution parameters can be calculated in a manner similar to the calculation of η_A.

Consideration of the Number of Measurement Sites

Figure 3 is useful for deciding how many measurement sites are necessary. The indicators η_A and η_D show a large improvement in the inverse transient solution when a small number of measurement sites are increased. However, there is little improvement in the inverse transient solution when large numbers of measurements are increased. The indicator η_J does not depict this behavior, although it does show improvement of the inverse transient solution with increasing numbers of measurement sites.

Figure 3 shows an apparent correlation between both the optimal measurement site configuration and number of measurement sites that should be considered in the optimization. However, when considering a certain number of measurement sites, say n, the indicator for the optimal measurement site configuration for n sites will always be worse than the indicator for the optimal measurement site configuration for $n + 1$ sites, because the optimal

measurement site configuration for n sites is a sub-set of the search space for the optimal measurement configuration for $n + 1$ sites. The extra site contains extra measured data and must improve the optimal indicator value.

Figure 5 considers the optimal measurement location fronts when different lengths of transient data are used. In general, the optimal measurement site configurations change little when using different data lengths, but they do change. The families of curves show that for each indicator, the curves do not intersect one another. The optimal indicator value for a certain number of measurement sites always improves as the data length increases. Other considerations about the length of data are presented in the following section.

Consideration of the Length of Measurement Data

It is logical to suggest that the longer the length of data used in the inverse transient method, the more confidence one would have in the results. The performance indicators can be used to observe the parameter confidence for differing lengths of measurement data. The curve plotting the indicator value versus length of measurement data is a useful tool for making the decision of what length of data is sufficient. Such a curve is determined for the small example network for two measurement sites at nodes 4 and 6 in which the inverse problem is to determine the existence of a leak using leak candidates at all nodes except at the reservoir. Figure 6 shows the performance indicators η_J, η_A and η_D and number of solvable parameters as a function of the length in time of the measurement data.

Initially at time $t = 0.0$ s, the inverse problem is indeterminate, i.e. there is one measurement at each of nodes 4 and 6 and six unknown leak parameters. At this time the calculation of the

indicator η_A is impossible ($[\kappa^*]^{-1}$ is undefined). At later times with greater measurement lengths per node singular-value decomposition (SVD) is used to invert $[\kappa^*]$ since it may still be singular. The singular values that result from the SVD show which parameters can be determined. In some cases, parameters in portions of the pipe network are determinable, while in other cases the whole inverse problem is unsolvable until a sufficient time has elapsed for the information from each leak location to be propagated to all of the measurement sites. For the current example, the SVD analysis shows that all six leak parameters are determinable at times greater than 3.5 s, which coincides to the time taken for the initial transient (generated at node 4 at a time t = 2.0 s) to reach node 7. The value of the indicator η_A decreases and the values of η_J and η_D increase as the length of measurement data increases after t = 3.5 s, which means that the more data, the better the inverse transient results. From the point at which all parameters are determinable, the magnitude of the η_A indicator decreases by approximately four orders of magnitude. As the measurement data length increases in Figure 6, the trade-off curve begins to flatten out. The asymptotic behavior of the predicted inverse transient performance corresponds to the dying out of the transient (see Figure 2), after which steady state conditions return to the system and additional measurement data have little effect on the performance of the inverse transient method.

Steady state measurement can gather an equal amount of data as transient measurement only if a (large) number of different steady states are used, each of which takes much longer to set up and to be sure that the flow is really steady. But even with data collected from several steady states, the inverse analysis may not determine the parameters as accurately as transient analysis. Consider the example and assume that measurements are taken at all nodes in the small pipe network to determine six unknown leak areas (one at each node). Now consider two different scenarios where (i) no transient event is generated but measurements are made

at several steady state conditions, and (ii) a transient event is generated in a manner similar to that previously described.

Figure 7(a) shows a plot of the partial derivative (the change) of the head at node 3 with respect to the leak area at node 2. By definition, the steady state value does not vary with time. The transient derivative, on the other hand, does vary with time and presents a large change in values, many of which are larger than the steady value. At many points the transient the head at node 3 is more sensitive to a leak at node 2 than the steady head. Even if the user were wise enough to choose a steady state where the leak at node 2 is near the maximum sensitivity, it is likely that would not be the case for other nodes. Thus, by passing through a range of sensitivities, the transient analysis can better determine parameters than a series of steady states. This idea is illustrated in the dramatic difference between the steady η_J, η_A and η_D and the transient η_J, η_A and η_D values (indicators of the error magnitude in the solution parameters), as shown in Figures 7(b), 7(c) and 7(d). In the case of η_A, if the partial derivative of the head with respect to the parameters is constant, as it is in the steady case, then the indicator η_A using N_L data points per site is equal to $1/N_L \times (\eta_A)_1$, where $(\eta_A)_1$ corresponds to the indicator value using 1 data point per site. When using all 40 seconds of transient data, the transient η_A is three orders of magnitude smaller than the steady η_A, showing that inverse transient analysis will by far outperform inverse steady analysis for similar measurement data lengths.

This section highlights the influence of measurement data length on inverse transient results with respect to measurement error and solvability of the parameters. However, there are other data length issues not considered here. There are arguably six data length issues to consider that influence inverse transient results:

1. The solvability of parameters (as discussed previously) that promotes a minimum data length below which the information from the transient event has not yet propagated to all measurement sites.
2. The effect of measurement error (also discussed previously) that indicates a long data length should be used.
3. The effect of model error (e.g., caused by inaccurate model properties or exclusion of unsteady friction) that promotes shorter data lengths where the transient model prediction is the most accurate.
4. The dynamic nature of distribution systems whereby demands and tank levels are not static, suggesting that short data lengths should be used to reduce the variation of such system properties.
5. The expense of computation times, which increases with long data lengths.
6. The possibility of contamination of measured data by uncontrolled events, which is less likely for short sampling periods.

These six considerations have competing objectives suggesting that the optimal data length is most probably an optimal front of some kind. For example, there is a correlation between the number of measurement sites and the length of measurement data used (i.e., longer data lengths and less measurement sites may produce equally good results as shorter data lengths and many sites). In addition, careful planning can improve inverse transient results, such as conducting tests at 3am when there is little variation in demands. A full consideration of these effects suggests further study is required. Indeed, some considerations are site specific and thus general rules are difficult to make.

Discussion of the Small Example Network Results

The optimal measurement sites for the example network were found in the previous sections through the full enumeration of measurement site combinations. The use of each performance indicator (η_J, η_A and η_D) produces different sets of optimal measurement sites. A reason for this difference is better explained in the next section based on a larger network. Using results from the small example network, a strategy for the selection of optimal measurement site configurations can be formed as follows:

1. Define the inverse transient problem including model data.
2. For a given pipe network identify the desired parameters to be determined.
3. Decide on the strength, sharpness and location of the initial transient.
4. Determine possible measurement site locations.
5. Set the measured data length to an estimate of the predictive range of the transient model.
6. Using performance indicators (perhaps η_D), determine the optimal measurement site front for both location and number of measurement sites.
7. Using engineering judgement (perhaps based on cost of measuring data at a number of sites) to choose the number of measurement sites and associated optimal measurement site configuration.
8. Perform inverse transient analysis.
9. Calculate the expected variance (using η_A) and solvability of parameters to assess the confidence of the inverse transient solution.

This method to find the optimal site configuration could also incorporate pipe roughness calibration as well as the leak areas as illustrated in the previous sections. The same

principles are applicable to find the optimal number and locations to estimate pipe roughnesses. A difference in sensitivity magnitudes (as, e.g., the three order-of-magnitude difference in leak and friction sensitivities) would be manifested in the performance indicators since they are all dependent on some form of partial derivative (e.g., $\partial H/\partial a$). A suitable scaling of each performance indicator alleviates this problem.

In a real pipe network, with a large number of pipes, performing a complete enumeration for every possible combination of measurement sites is impracticable. Alternatively, a sample optimal curve can be constructed by randomly generating a set of site configurations and then computing their respective expected inverse transient performance (using the indicators) and creating a curve that is an approximation of the true optimal front curve.

In general, the use of more measurement sites will provide better inverse transient results than a less number of sites. The approach for deciding the optimal number of measurement sites is the based on the trade-off between the cost of setting up such measurement sites and the predicted inverse transient performance (as indicated by the indicators). Ultimately, the inverse transient practitioner would decide how many sites to use.

As shown in the analysis, the longer the measured data length, the better the inverse transient results. This suggests that long data lengths will dictate the choice of the best data length. However, the best data length is more likely to be related directly to the predictive range of the transient model. Therefore, the predictive range of the model should be used to set the measured data length. In addition, a short sampling period may be useful in minimizing the possibility of contamination by uncontrolled factors in the pipe network (such as induced transients from an uncontrolled source).

APPROACH FOR LARGE NETWORKS

Finding the optimal number of measurement sites and their locations is a more difficult problem (and it is more important to find a solution) in a large pipe network as compared to a small pipe network. Different solutions to the problem are presented in this section.

The Optimization Method

Finding the optimum number of measurement sites and their locations has two objectives. One is to use the least number for the desired accuracy. The second is to find what combination gives the optimal inverse transient performance. The genetic algorithm (GA) is well suited as an optimization technique to meet these objectives. It is evolutionary-based whereby a population of solution strings are maintained and subjected to evolutionary pressures. Each solution string is comprised of bits, which for the optimal measurement site case is a string of possible sites. The fitness of each string is evaluated using a performance indicator. The population evolves through the application of genetic operators such as selection, crossover and mutation. After a number of generations the stings converge to the solution that maximizes the fitness of the strings. GAs perform an efficient optimization by only searching a small proportion of the search space. GAs have been used as the minimization algorithm in the inverse transient method by Vítkovský *et al*. (2000).

The GA optimization is only required for the indicators η_A and η_D. The optimal measurement site configuration using the indicator η_J is determined by simply calculating η_J value for each

separate measurement site, ranking these values and selecting the sites with the highest rankings to form the optimal measurement site configuration. In this respect, the set of optimal measurement locations for n sites is a superset of the set of optimal measurement locations for $n - 1$ sites. This ranking process may be performed because the calculation of η_J is a summation of independent components for each measurement site (see Eq. 13).

New Genetic Algorithm Operators

Within a genetic algorithm formulation the operators of selection and mutation are applied in the standard manner (Goldberg 1989) for the combinatorial search problem. Binary tournament selection and random mutation (excluding the creation of illegal chromosomes) have been used here. However, standard *N*-point crossover and partially mapped crossover (PMX), described in Goldberg (1989), do not work for this problem for a variety of reasons. In the case of *N*-point crossover, there is no facility by which the genetic algorithm operator can deal with the non-repeatability of bits in a chromosome. PMX can ensure that bits are not repeated in the chromosome, given that the number of bit values is equal to the chromosome length. However, when applying it to a chromosome where ordering is not important and there are a greater number of possible bit values than the chromosome length (such as in the problem at hand), PMX can form illegal chromosomes. The non-ordered, extended-set, combinatorial crossover (NOESCX) provides a flexible crossover operator that does not create illegal chromosomes. Figure 8 shows the process for NOESCX. Essentially each bit that is located in the crossover zone of one chromosome is checked to see if there is a matching bit in the other chromosome. If there is no match, then the bit is copied to the corresponding position in the other chromosome. The NOESCX operator produces no

repeated bit values in the chromosomes (if both parent chromosomes are legal) and order is not important.

Mutation is performed by randomly choosing a string from the population given some probability of mutation. Then one bit in the string is randomly chosen and replaced with a randomly generated value. If the new string is illegal, such as containing two identical measurement sites, the process is repeated until a legal string is achieved.

These genetic operators (together with selection) can be used to create a new population. Before the strings' fitnesses are evaluated, the entire population is checked for multiple copies of the same measurement site configuration, thus preventing evaluations of identical strings and reducing unneeded evaluations. The production of successive generations is repeated until convergence, or until the number of generations exceeds some maximum number. Anderson and Simpson (1996) developed the base source code used for this GA. For low numbers of configurations, the number of possible combinations may be lower than the expected number of function evaluations used by the GA, in which case full enumeration is used. The combination of full enumeration and the genetic algorithm is referred to as EGA searching.

LARGE NETWORK EXAMPLE

The configuration of the large network is shown in Figure 9 and is based on one used by Jeppson (1976). This network is large compared to the fully enumerable network previously used, but not as large as those commonly found in practice. In practice, isolating portions of

the network using valves can reduce their size. The system is gravity fed by three reservoirs and consists of 51 pipes and 35 nodes. The transient event was generated by instantaneously closing a valve that discharged to the atmosphere at node 7 at time $t = 1.0$ s. There were constant demands at all non-reservoir nodes with the exception of nodes 2, 16, 18, 20 and 31. The pipe properties for the network ranged between 400 and 900 m for pipe lengths, 0.3 and 1.5 m for diameters and 1,100 and 1,400 m/s for wave speeds. Some pipes of the original network were divided into two or more pipes so the minimum Courant number is 0.8 for computational accuracy, thus increasing the size to 75 pipes and 59 nodes. A cubic timeline scheme was used to minimize interpolation errors (Yang and Hsu 1990). A simulation time of 40.0 s was used, which was a more than adequate time for the transient to reach every part of the network. All parameters can be theoretically determined, although some parameters are better determined than others.

Possible measurement sites are located at all nodes, excluding those added from the pipe division and the reservoir nodes, making a total of 32. The number of measurement sites was limited to ten because it is likely that, in reality, there will be considerably fewer measurement sites than nodes in a pipe network. Table 3 shows the number of possible site configurations as a function of the number of sites (up to ten). Full enumeration was used for configurations with the number of sites equal to or less than four, whereas GA optimization was used for higher numbers. The GA parameters used for the optimization of the measurement site location are a population size of 1,000, tournament selection, a probability of crossover 0.7, and a probability of mutation 0.01. The maximum number of generations allowed is 100 (although, typically, the GA converged in less than 50 generations). The maximum number of evaluations of η_J, η_A or η_D used by the GA is 100,000. Therefore, if the number of possible site configurations for a particular number of sites is less than 100,000, full

enumeration is used, if not then a GA search is used. GAs are stochastic algorithms and are not guaranteed find to true global minimum. In this paper three GA runs with different seeds for the random number generator were used for each optimization. If the same optimal solution was not determined in two of the three differently seeded GA runs, additional GA runs were performed.

Sampled Results

Due to the large number of measurement site combinations, the calculation of the performance indicator for every possible configuration is too computationally intensive. Only a few sample configurations are used to generate an approximation of the optimal site front and the optimal curve. The sample results are representative of the population results given that the sample size is adequately large. Figure 10 shows approximate curves for the average indicator values from 30 randomly chosen configurations. These curves can be used to determine quickly an adequate number of sites for the inverse transient method. The best measurement site configuration for each number of sites selected is shown in Tables 4, 5 and 6 together with its corresponding indicator value.

In addition to being useful for the consideration of the number of measurement sites for the inverse transient method, the result also serves as a check for the EGA optimization to find the best locations of sites (presented in the following section). The following section applies the EGA technique to produce better optimal configurations than by taking random samples.

Optimization Results

The search space sizes for the different numbers of measurement site combinations dictates what type of strategy should be used in the EGA optimization. Full enumeration was used for numbers of sites equal to 1, 2, 3, and 4, and the genetic algorithm was used for higher numbers. Figure 10 shows the optimal front formed by either EGA optimization for η_A and η_D ranking for η_J for the placement of measurement sites where each point on the optimal curve front represents a configuration that provides the best indicator values (largest η_J or η_D, smallest η_A) for the large network. The optimal front bounds the sampled results (from 30 randomly generated configurations) indicating that the optimization is working. Tables 4, 5 and 6 shows the optimal sites and the indicator values for each number of sites considered. Some measurement sites such as sites 11, 12, 23 and 34 are consistently found in the optimal set (see Tables 4, 5 and 6). These will produce good convergence of the minimization algorithm due to their high sensitivities.

The optimal configurations as indicated by each performance indicator are different. The optimal sites using η_J typically cluster near the point at which the transient was generated (node 7), which corresponds to the highest $\partial H/\partial a$ sensitivity values. The optimal sites using η_A and η_D are well distributed throughout the network and correspond to the locations that minimize the effect of measurement error on the calculation of parameters. Essentially, η_J favors those parameters that are located near the transient generation point, resulting in poor inverse transient performance for parameters located further away from the transient generation point. Is this sense, the optimal configuration using η_J does not adequately solve the original problem where the inverse transient performance for all parameters, be they close or far from the transient generation point, must be optimized. Ultimately, the best indicator of inverse transient performance is η_D and it should be used when determining the optimal

location of measurement sites. In contrast, η_A is more computationally intensive, favors correlated parameters and may be dominated by high variance parameters.

INVERSE TRANSIENT STATISTICS

The expected variance of the parameters for a particular solution from the inverse transient method is a useful quantity. It can be used to predict how much confidence to associate with each solved parameter and, ultimately, how successful the inverse transient method has been. In previous sections of this paper, the indicator η_A has been shown to predict the overall performance from the inverse transient method given a particular configuration of measurement sites. The variances in the parameters and covariances between parameters (Eqs. 8 and 9) are determined during the calculation of η_A. The marginal probability distribution function (Yevjevich 1972) is

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-(x-\mu_x)^2 / 2\sigma_x^2} \tag{18}$$

where $f(x)$ = probability density function of a parameter x, μ_x = mean of parameter x and σ_x = standard deviation of parameter x. A further step can be taken by considering the expected covariance (also found when calculating the expected variance, see Eqs. 4 to 9). The covariance can be used to check if leak areas are correlated. Similarly to Eq. 18, a bivariate probability density function, $f(x,y)$, written for two normally distributed parameters x and y, is

$$f(x,y) = \frac{e^{-Q\left(1-\rho^2\right)/2}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \tag{19}$$

where the variable Q is equal to

$$Q = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(x-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(x-\mu_y)}{\sigma_x\sigma_y} \tag{20}$$

The correlation coefficient, ρ, calculated between any two parameters (Yevjevich, 1972) is defined as

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \tag{21}$$

where σ_x = standard deviation in parameter x, σ_y = standard deviation in parameter y and σ_{xy} = covariance between parameters x and y. A value of $\rho = \pm 1.0$ represents exact correlation and $\rho = 0$ is no correlation. In the small pipe network case study presented earlier Figure 11 shows the bivariate probability density function calculated between leak areas $(C_dA_L)_4$ and $(C_dA_L)_6$ and between $(C_dA_L)_5$ and $(C_dA_L)_7$ representing both low and high correlation pairs of parameters, respectively. Figure 11 uses standardized leak areas, calculated by subtracting its mean then dividing by its standard deviation.

Little correlation exists between the two leak areas at nodes 4 and 6 (as indicated by the near circular concentric rings on the bivariate probability density contour plot) whose correlation coefficient is equal to 0.13. A different bivariate distribution is observed between the lumped leak areas at node 5 and 7 where the correlation coefficient is equal to –0.80. The correlation between the leak areas at nodes 5 and 7 is indicated by the elliptic shape of the contours with the major axis of the ellipse along the diagonal $\sigma_y = -\sigma_x$. Although the correlation between the leak areas at nodes 5 and 7 is high, it is not perfect. A unique solution for each leak area exists; however, if the measurements contain error then those pairs of leak areas with high correlation and high variances and covariance (compared to other parameters) may not be well determined. The correlation between the lumped leak areas is visualized using the correlation matrix. Figure 12 shows the correlation matrix for the inverse transient problem

for the determination of six unknown leaks. If the parameters were friction factors, the correlation matrix would be useful for indicating which friction factors might be lumped together to improve their overall sensitivity (Vítkovský 2001).

For the small example network, the first-order parameter error estimates as calculated in the preceding paragraphs were verified using a Monte Carlo method (Vítkovský 2001). Monte Carlo methods are stochastic techniques used for investigation of probabilistic problems. For each Monte Carlo simulation the measurement data were contaminated with an artificial, normally-distributed error of zero mean, then the leak areas determined using the inverse transient method. After a number of such simulations, when the error in the leak areas becomes ergodic, the leak area statistics are calculated. The leak area error statistics demonstrate the effect of measurement error. One hundred thousand Monte Carlo simulations were used to calculate parameter values. A standard Levenberg-Marquardt (Press *et al*. 1992) minimization algorithm was used in the inverse transient method. The two methods of calculation produced nearly identical results for all marginal and bivariate probability density functions.

If the measurement error becomes too large, first-order error analysis is invalid because the linearized Taylor series expansion does not adequately represent the nonlinear behavior. However, failure to match the first-order error analysis might not be due solely to nonlinear effects. A major cause of failure of the inverse transient method for large measurement errors is the minimization algorithm. The measurement data contain a greater proportion of outliers when the error becomes large. These outliers cause problems such as local minima for the least-squares fitting algorithm, slow convergence and non-convergence. The shuffled complex evolution algorithm, used with the inverse transient method, avoids some of the

pitfalls of other minimization methods and has been successful where the Levenberg-Marquardt algorithm failed, thus suggesting that the non-convergence was caused by the failure of the minimization algorithm rather than nonlinear effects. In inverse transient tests applied to experimentally collected laboratory data, a global-minimization algorithm was required (Vítkovský 2001).

CONCLUSIONS

Currently engineering judgment is used to locate measurement sites for the inverse transient method. Better site locations are determined using formal optimization combined with engineering judgment. Given the large number of combinations of site configurations in a large pipe network, optimization algorithms outperform human intuition due to the size of the problem. However, the optimal configuration gained through an optimization algorithm may not be practical; therefore, engineering judgement is still required, but at a higher level. The use of a genetic algorithm produces a number of near-optimal solutions; thus, if the optimal solution is impractical then one of the near-optimal solutions might be used. The formation of an optimal curve for η_D (or η_A) and N_S shows that there is never one solution, rather an optimal curve of solutions. Again, engineering judgment must be used to decide what level of error is acceptable in the parameters versus the cost of additional sites.

The characterization of inverse transient analysis performance using performance indicators allows comparison between different measurement site configurations. Although three such indicators have been investigated in this paper, other types of indicators exist, with each indictor yielding slightly different optimal solutions. As to which indicator is the best, the

answer depends on how the indicator performance is judged, e.g., computational efficiency, projected parameter error, etc. Considering the large network example in this paper, the η_D indicator is preferred over both the η_J and η_A indicators.

In the same way that the performance indicators have been used to determine optimal measurement site locations, the method presented herein could be used to determine where a transient should be generated or even the transient type. In addition, there is a correlation between the number of measurement sites and the length of measurement data (i.e., longer data lengths and less measurement sites may produce equally good results as shorter data lengths and many sites). Expanding the inverse transient design problem to its fullest, the location of the measurement sites, transient generation location and length of measurement data used at each site could all be optimized simultaneously forming a multi-objective optimization problem. The methods discussed in this paper can be applied to inverse problems other than the inverse transient problem where measurement of data may occur at a number of discrete locations.

ACKNOWLEDGEMENTS

This work has been supported by a large grant from the Australian Research Council and this support is gratefully acknowledged. Furthermore, the authors would like to thank Dr. Zoran Kapelan and the reviewers for their useful comments and input to this research.

NOTATION

The following symbols are used in this paper:

a_j or $\{a\}$ = generic parameter representing C_dA_L or f;

$[C]$ = covariance matrix;

$[C^*]$ = normalized covariance matrix;

C_dA_L = effective leak area;

E = objective function, chi-square statistic;

f = Darcy-Weisbach friction factor, probability density distribution;

F = optimization objective;

H = head;

H_i^m or $\{H^m\}$ = measured head;

$M_j^{(1)}$ = sensitivity measure based on objective function;

$M_{ij}^{(2)}$ = sensitivity measure based on Jacobian;

N_C = number of measurement site combinations;

N_L = number of measurement data per site;

N_M = total number of measurement data ($=N_S \times N_L$);

N_P = number of parameters;

N_S = number of measurement sites;

N_{TS} = total number of measurement sites;

Q = bivariate distribution variable;

η_A = performance indicator based on the A-optimality criterion;

η_D = performance indicator based on the D-optimality criterion;

η_J = performance indicator based on the Jacobian of H;

$[\kappa]$ = curvature matrix;

$[\kappa^*]$ = normalized curvature matrix;

ρ_{xy} = correlation between x and y;

σ_x = standard deviation of x;

σ_x^2 = variance of x;

σ_{xy} = covariance between x and y.

REFERENCES

Anderson, A., and Simpson, A.R. (1996). "Genetic Algorithm Optimisation Software in Fortran." *Research Report No. R136*, March, Department of Civil & Environmental Engineering, The University of Adelaide, Australia.

Bargiela, A., and Hainsworth, G.D. (1989). "Pressure and Flow Uncertainty in Water Systems." *Journal of Water Resources Planning and Management*, ASCE, 115(2), March, 212-229.

Bergant, A., Simpson, A.R., and Vítkovský, J.P. (2001). "Developments in Unsteady Pipe Flow Friction Modelling." *Journal of Hydraulic Research*, IAHR, 39(3), 249-257.

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

Carrera, J., and Neuman, S.P. (1986). "Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information." *Water Resources Research*, 22(2), February, 199-210.

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(2), 114-152.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Co., Inc.

Kapelan, Z.S., Savic, D.A., and Walters, D.A. (2001). "Use of Prior Information on Parameters in Inverse Transient Analysis for Leak Detection and Roughness Calibration." *1st World Water & Environmental Resources Congress*, ASCE, 21-24 May, Orlando, Florida, USA. [CD-ROM]

Knopman, D.S., and Voss, C.I. (1989). "Multiobjective Sampling Design for Parameter Estimation and Model Discrimination in Groundwater Solute Transport." *Water Resources Research*, 25(10), October, 2245-2258.

Lansey, K.E., El-Shorbagy, W., Ahmed, I., Araujo, J., and Haan, C.T. (2001). "Calibration Assessment and Data Collection for Water Distribution Networks." *Journal of Hydraulic Engineering*, ASCE, 127(4), April, 270-279.

Liggett, J.A., and Chen L.C. (1994). "Inverse Transient Analysis in Pipe Networks." *Journal of Hydraulic Engineering*, ASCE, 120(8), August, 934-955.

Loaiciga, H.A., Charbeneau, R.J., Everett, L.G., Fogg, G.E., Hobbs, B.F., and Rouhani, S. (1992). "Review of Ground-Water Quality Monitoring Network Design." *Journal of Hydraulic Engineering*, ASCE, 118(1), 11-37.

Meier, R.W., and Barkdoll, B.D. (2000). "Sampling Design for Network Model Calibration Using Genetic Algorithms." *Journal of Water Resources Planning and Management*, ASCE, 126(4), July/August, 245-250.

Nash, G.A., and Karney, B.W. (1999). "Efficient Inverse Transient Analysis in Series Pipe Systems." *Journal of Hydraulic Engineering*, ASCE, 125(7), July, 761-764.

Neuman, S.P. (1973). "Calibration of Distributed Parameter Groundwater Flow Models Viewed as a Multiple-Objective Decision Process under Uncertainty." *Water Resources Research*, 9(4), August, 1006-1021.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.

Pudar, R.S., and Liggett, J.A. (1992). "Leaks in Pipe Networks." *Journal of Hydraulic Engineering*, ASCE, 118(7), July, 1031-1046.

Vítkovský, J.P. (2001). Inverse Analysis and Modelling of Unsteady pipe Flow: Theory, Applications and Experimental Verification. *PhD Thesis*, March, Department of Civil & Environmental Engineering, The University of Adelaide, Australia.

Vítkovský, J.P., Simpson, A.R., and Lambert, M.F. (2000). "Leak Detection and Calibration Using Transients and Genetic Algorithms." *Journal of Water Resources Planning and Management*, ASCE, 126(4), July/August, 262-265.

Walski, T.M. (1983). "Technique for Calibrating Network Models." *Journal of Water Resources Planning and Management*, ASCE, 109(4), October, 360-372.

Wang, X.J., Lambert, M.F., Simpson, A.R., Liggett, J.A., and Vítkovský, J.P. (2001). "Leak Detection in Pipeline Systems Using the Damping of Fluid Transients." *Journal of Hydraulic Engineering*, ASCE, (accepted for publication).

Wylie, E.B., and Streeter, V.L. (1993). *Fluid Transients in Systems*. Englewood Cliffs, New Jersey, USA.

Xu, C., and Goulter, I.C. (1998). "Probabilistic Model for Water Distribution Reliability." *Journal of Water Resources Planning and Management*, ASCE, 124(4), July/August, 218-228.

Yang, J.C., and Hsu, E.L. (1990). "Time-Line Interpolation for Solution of the Dispersion Equation." *Journal of Hydraulic Research*, IAHR, 28(4), 503-523.

Yevjevich, V. (1972). *Probability and Statistics in Hydrology*. Water Resources Publications, Fort Collins, Colorado, USA.

Yu, G., and Powell, R.S. (1994). "Optimal Design of Meter Placement in Water Distribution Systems." *International Journal of Systems Science*, 25(12), 2155-2166.

LIST OF FIGURES

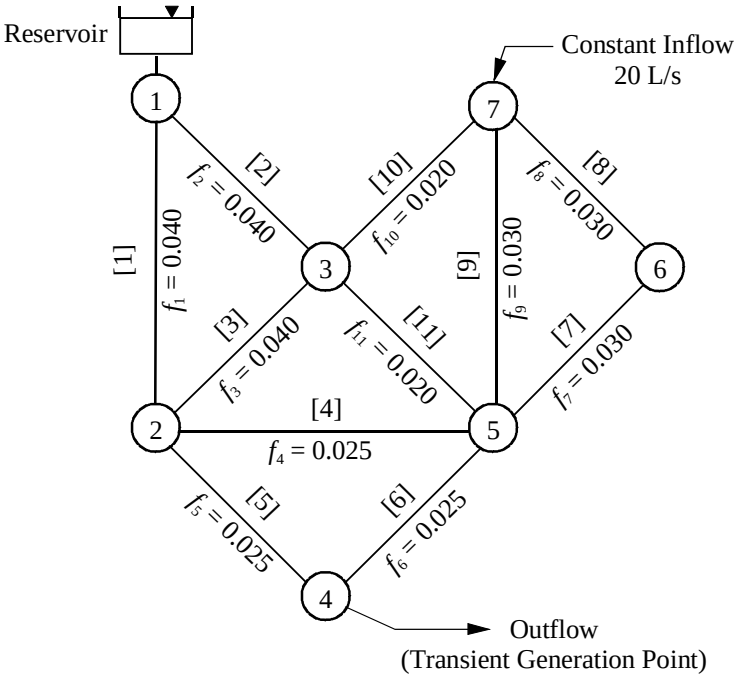

Figure 1. Layout of small example network

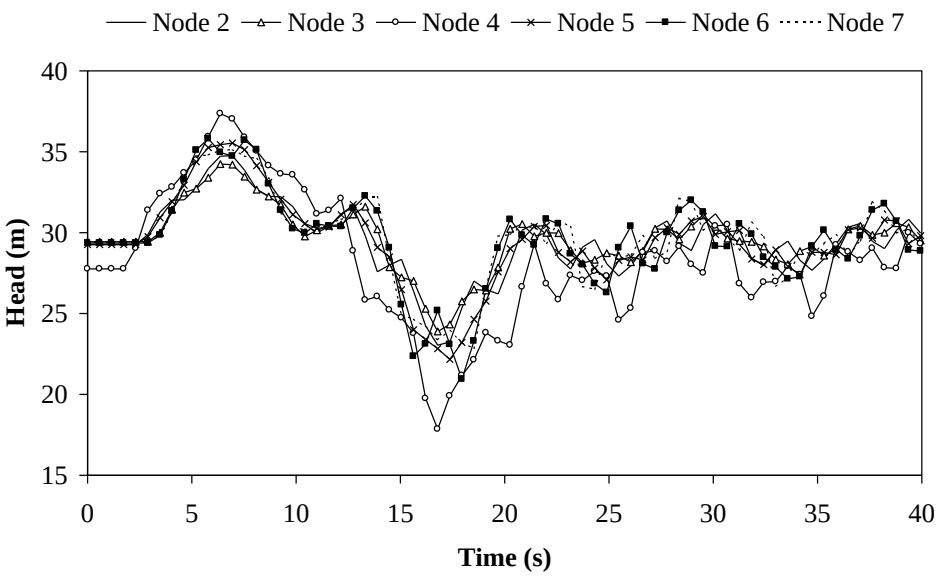

Figure 2. Measured head variation for the small example network

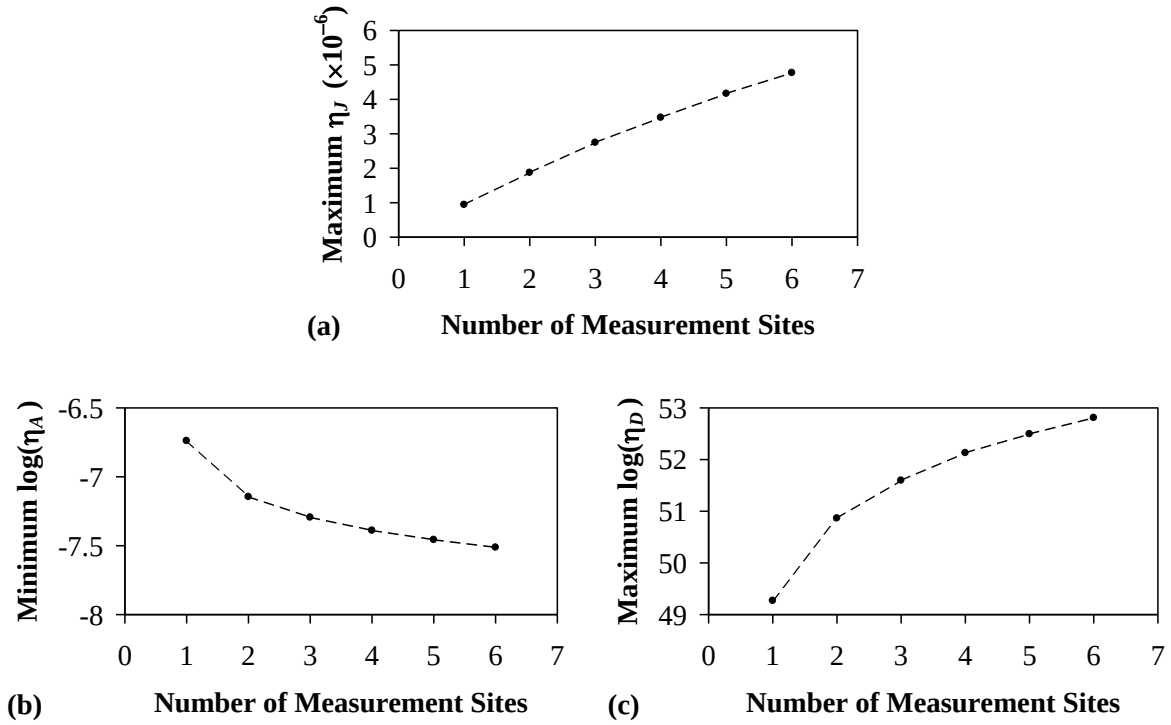

Figure 3. Optimal measurement site location curve (small example network)

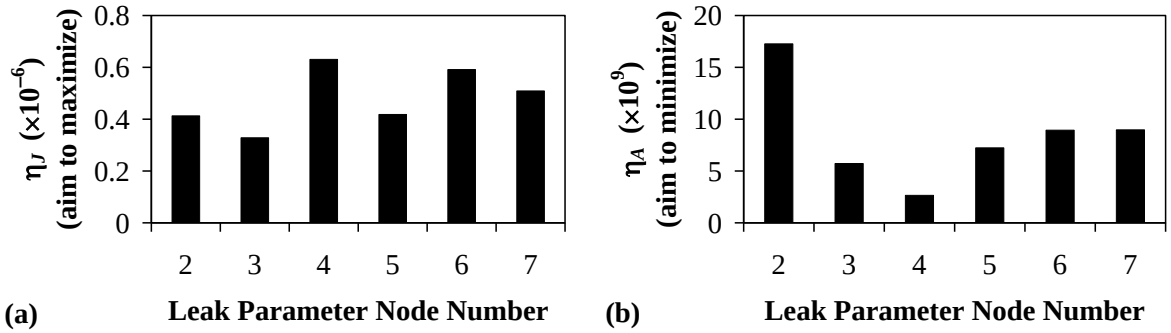

Figure 4. Decomposition of indicators by leak parameter component (small example network)

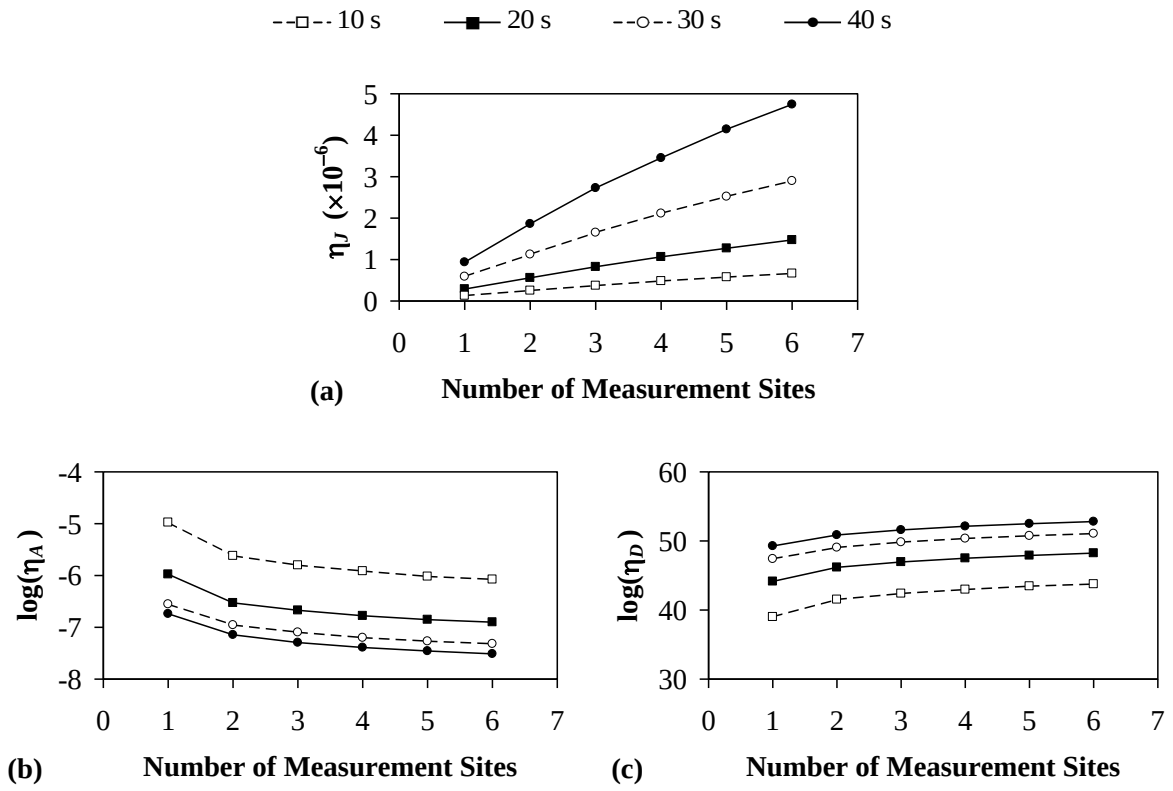

Figure 5. Data length effect on optimal measurement site locations curve (small example network)

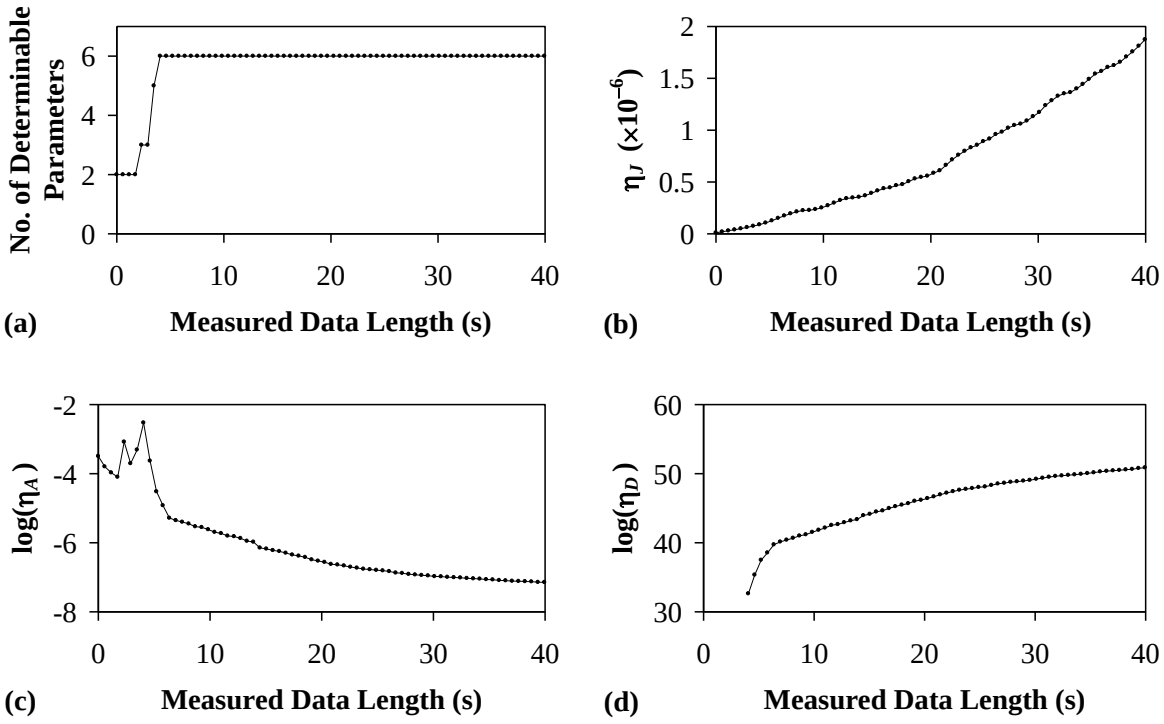

Figure 6. Data length effect on the number of determinable parameters and performance indicators (small example network)

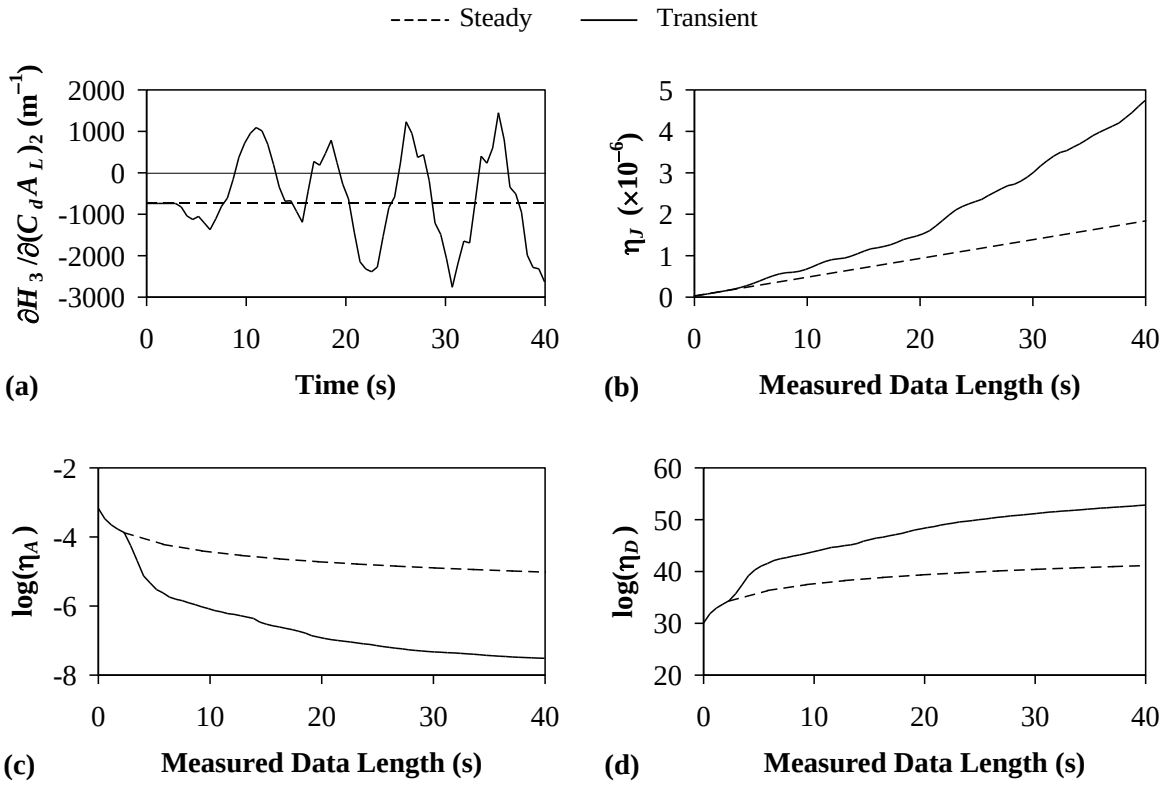

Figure 7. Steady and transient comparison of sensitivity and performance indicators

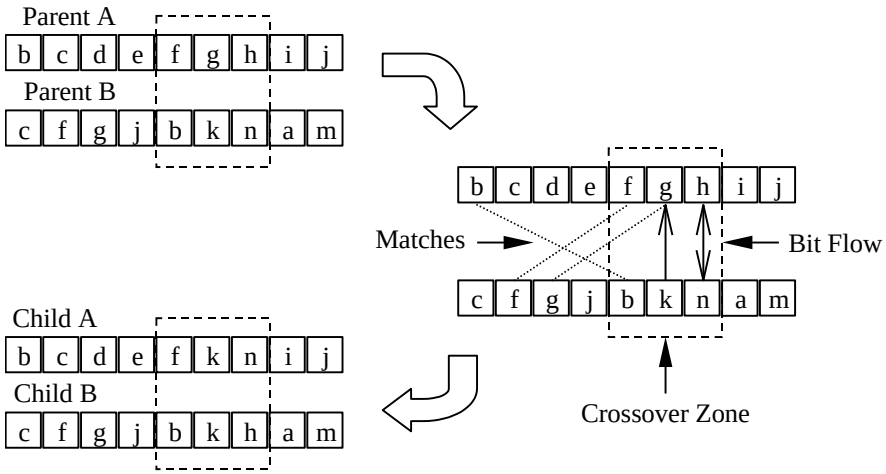

Figure 8. Non-ordered, extended-set, combinatorial crossover (NOESCX)

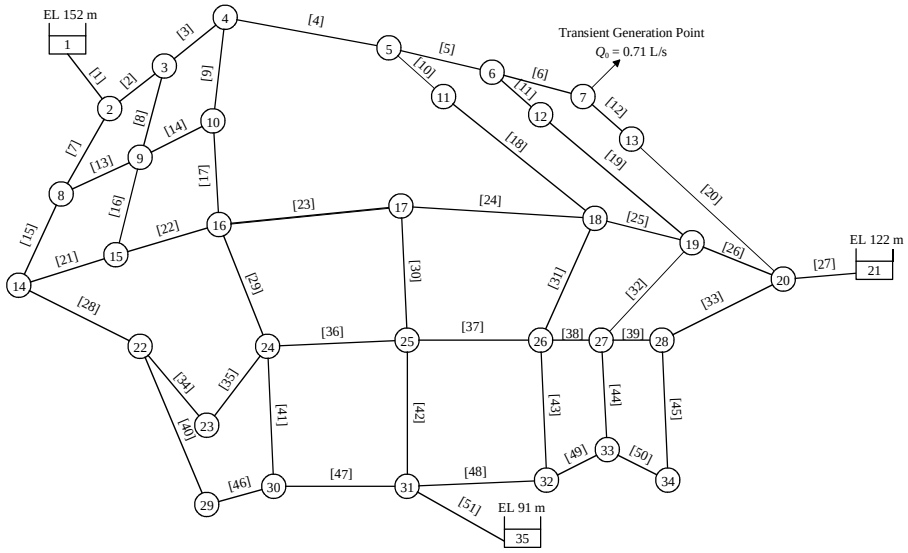

Figure 9. Layout of the large example network

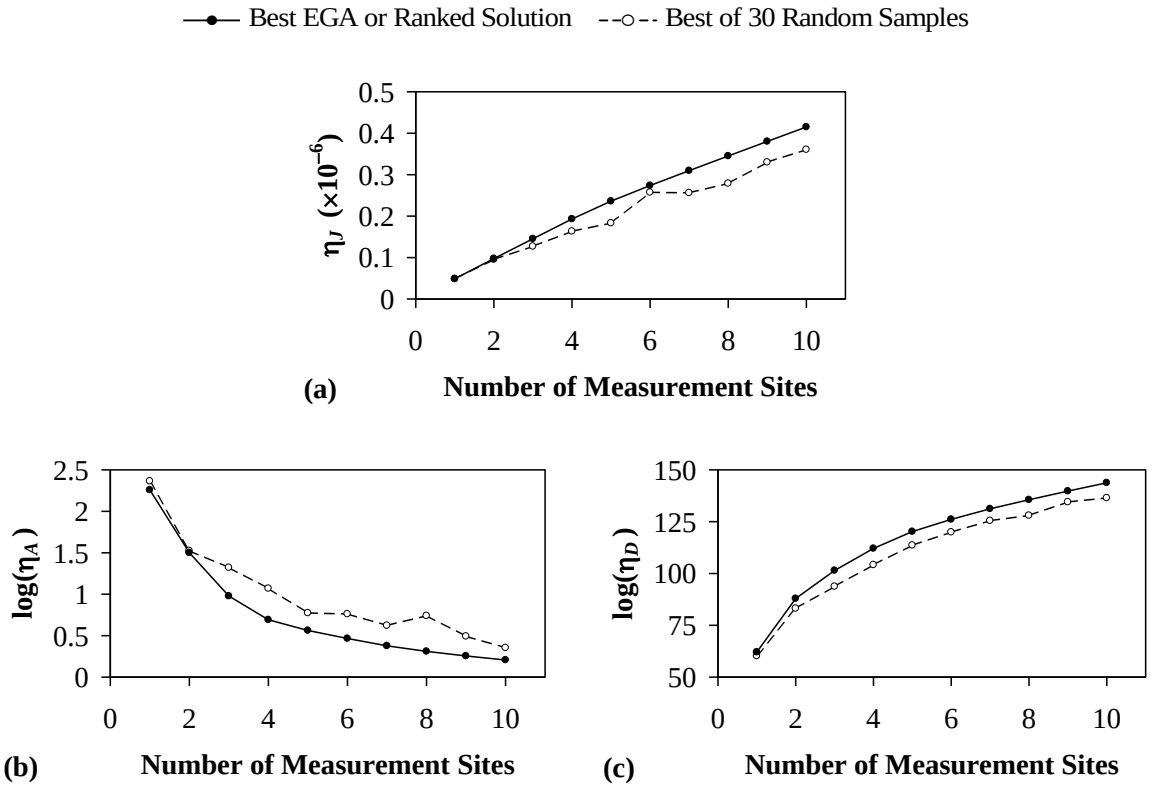

Figure 10. Optimal measurement site location curves (large example network)

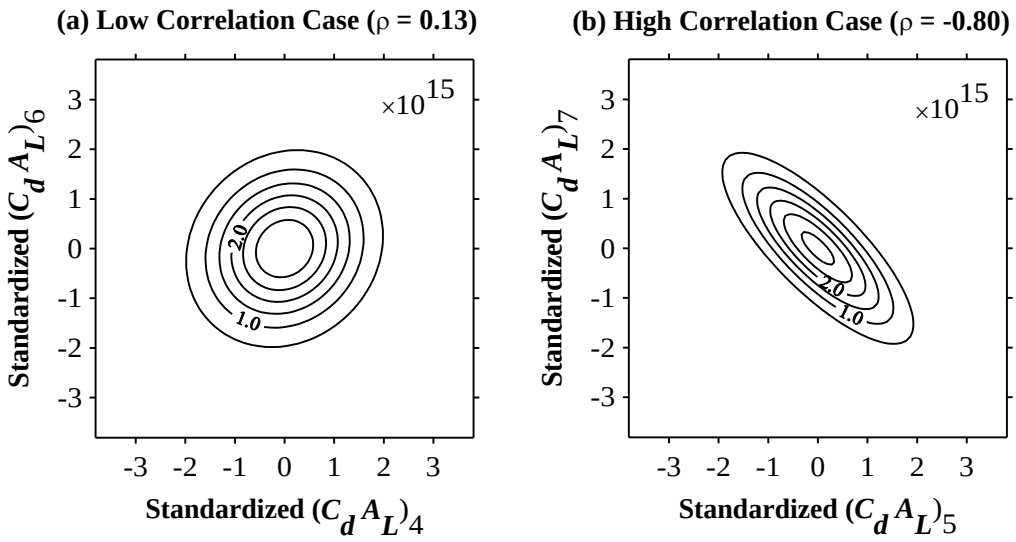

Figure 11. Bivariate probability distribution functions. Leak areas are standardized by subtracting their mean and dividing by their standard deviation. Probability density function multiplier is shown in top right-hand corner of plots.

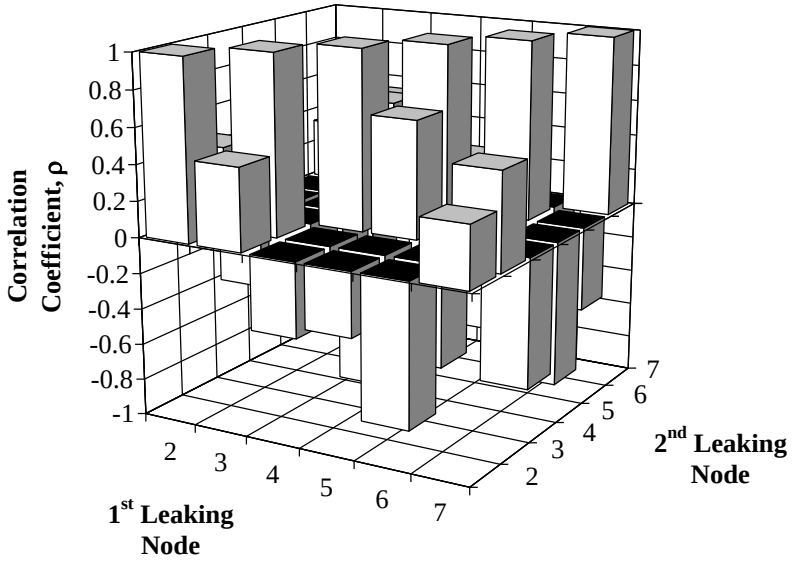

Figure 12. Correlation matrix for all parameters

LIST OF TABLES

Table 1. Measurement site combinations for the small example network

N_S	N_C*	Possible Measurement Site Configurations*
1	6	{2}, {3}, {4}, {5}, {6}, {7}
2	15	{2,3}, {2,4}, {2,5}, {2,6}, {2,7}, {3,4}, {3,5}, {3,6}, {3,7}, {4,5}, {4,6}, {4,7}, {5,6}, {5,7}, {6,7}
3	20	{2,3,4}, {2,3,5}, {2,3,6}, {2,3,7}, {2,4,5}, {2,4,6}, {2,4,7}, {2,5,6}, {2,5,7}, {2,6,7}, {3,4,5}, {3,4,6}, {3,4,7}, {3,5,6}, {3,5,7}, {3,6,7}, {4,5,6}, {4,5,7}, {4,6,7}, {5,6,7}
4	15	{2,3,4,5}, {2,3,4,6}, {2,3,4,7}, {2,3,5,6}, {2,3,5,7}, {2,3,6,7}, {2,4,5,6}, {2,4,5,7}, {2,4,6,7}, {2,5,6,7}, {3,4,5,6}, {3,4,5,7}, {3,4,6,7}, {3,5,6,7}, {4,5,6,7}
5	6	{2,3,4,5,6}, {2,4,5,6,7}, {2,3,5,6,7}, {2,3,4,6,7}, {2,3,4,5,7}, {3,4,5,6,7}
6	1	{2,3,4,5,6,7}

* Head measurement was not allowed at the reservoir (node 1)

Table 2. Optimal measurement site locations for the small example network

N_S	Best Meas. Site Config. Using η_J	Best η_J	Best Meas. Site Config. Using η_A	Best η_A	Best Meas. Site Config. Using η_D	Best η_D
1	6	9.4×10^{6}	4	1.8×10^{-7}	4	1.9×10^{49}
2	4, 6	18.7×10^{6}	4, 6	7.2×10^{-8}	4, 6	7.3×10^{50}
3	4, 6, 7	27.5×10^{6}	3, 4, 6	5.1×10^{-8}	4, 6, 7	3.9×10^{51}
4	4, 5, 6, 7	34.7×10^{6}	3, 4, 6, 7	4.1×10^{-8}	3, 4, 6, 7	1.3×10^{52}
5	2, 4, 5, 6, 7	41.7×10^{6}	2, 3, 4, 6, 7	3.5×10^{-8}	3, 4, 5, 6, 7	3.1×10^{52}
6	2, 3, 4, 5, 6, 7	47.7×10^{6}	2, 3, 4, 5, 6, 7	3.1×10^{-8}	2, 3, 4, 5, 6, 7	6.5×10^{52}

Table 3. Number of measurement site combinations (large example network) in which N_s is the number of sites and N_c is the number of configurations.

N_S	N_C	Analysis Type*
1	32	E
2	496	E
3	4,960	E
4	35,960	E
5	201,376	GA
6	906,192	GA
7	3,365,856	GA
8	10,518,300	GA
9	28,048,800	GA
10	64,512,240	GA

* E = Full Enumeration, GA = Genetic Algorithm

Table 4. Best of sampled sites and optimal ranked sites using η_J (large example network)

N_S	Best of Sampled Measurement Site Configurations	η_J	Optimal Ranked Measurement Site Configurations	η_J
1	5	4.9×10^{4}	11	4.9×10^{4}
2	6, 12	9.6×10^{4}	5, 11	9.7×10^{4}
3	5, 6, 19	1.3×10^{5}	5, 11, 12	1.4×10^{5}
4	5, 11, 14, 19	1.6×10^{5}	5, 6, 11, 12	1.9×10^{5}
5	4, 5, 14, 18, 33	1.8×10^{5}	4, 5, 6, 11, 12	2.4×10^{5}
6	4, 5, 6, 8, 12, 14	2.6×10^{5}	4, 5, 6, 10, 11, 12	2.7×10^{5}
7	3, 4, 5, 10, 15, 18, 19	2.7×10^{5}	4, 5, 6, 9, 10, 11, 12	3.1×10^{5}
8	3, 5, 9, 10, 12, 13, 16, 18	2.8×10^{5}	4, 5, 6, 9, 10, 11, 12, 15	3.5×10^{5}
9	3, 4, 6, 11, 12, 14, 16, 27, 28	3.3×10^{5}	4, 5, 6, 9, 10, 11, 12, 14, 15	3.8×10^{5}
10	4, 6, 8, 9, 11, 15, 19, 22, 26, 34	3.6×10^{5}	4, 5, 6, 8, 9, 10, 11, 12, 14, 15	4.2×10^{5}

Table 5. Best of sampled sites and optimal EGA sites using η_A (large example network)

N_S	Best of Sampled Measurement Site Configurations	η_A	Optimal EGA Measurement Site Configurations	η_A
1	5	231	11	181
2	10, 11	33.4	11, 23	31.7
3	11, 14, 23	21.0	10, 12, 23	9.51
4	4, 10, 12, 16	11.8	10, 11, 12, 23	4.92
5	5, 7, 12, 15, 23	5.94	10, 11, 12, 23, 34	3.66
6	3, 11, 19, 23, 28, 34	5.78	10, 11, 12, 23, 24, 34	2.92
7	7, 10, 11, 13, 23, 24, 32	4.20	6, 10, 11, 12, 23, 24, 34	2.39
8	6, 7, 13, 14, 15, 18, 23, 28	5.51	6, 10, 11, 12, 16, 23, 24, 34	2.05
9	5, 7, 10, 11, 23, 26, 29, 32, 33	3.11	4, 6, 10, 11, 12, 16, 23, 24, 34	1.80
10	6, 7, 10, 12, 15, 16, 23, 26, 27, 32	2.26	4, 6, 10, 11, 12, 16, 23, 24, 33, 34	1.61

Table 6. Best of sampled sites and optimal EGA sites using η_D (large example network)

N_S	Best of Sampled Measurement Site Configurations	η_D	Optimal EGA Measurement Site Configurations	η_D
1	7	1.6×10^{60}	11	1.3×10^{62}
2	10, 11	1.5×10^{83}	7, 11	8.6×10^{87}
3	11, 14, 23	5.7×10^{93}	7, 10, 11	3.0×10^{101}
4	4, 10, 12, 16	1.6×10^{104}	10, 11, 12, 23	1.2×10^{112}
5	5, 7, 12, 15, 23	4.1×10^{113}	7, 10, 11, 12, 23	1.7×10^{120}
6	3, 11, 19, 23, 28, 34	9.5×10^{119}	7, 10, 11, 12, 23, 34	1.3×10^{126}
7	7, 10, 11, 13, 23, 24, 32	3.1×10^{125}	4, 7, 10, 11, 12, 23, 34	1.6×10^{131}
8	3, 6, 7, 8, 11, 14, 17, 22	1.1×10^{128}	4, 7, 10, 11, 12, 16, 23, 34	3.8×10^{135}
9	5, 7, 10, 11, 23, 26, 29, 32, 33	3.4×10^{134}	4, 7, 8, 10, 11, 12, 16, 23, 34	5.8×10^{139}
10	6, 7, 10, 12, 15, 16, 23, 26, 27, 32	2.8×10^{136}	3, 4, 7, 8, 10, 11, 12, 16, 23, 34	5.9×10^{143}