

# PRESSURE AND FLOW UNCERTAINTY IN WATER SYSTEMS

Andrzej Bargiela and Graham D. Hainsworth\*

## ABSTRACT

*In the monitoring of water distribution systems, the inaccuracy of input data contributes greatly to the inaccuracy of system state estimates calculated from them. It is important, therefore, that the system operators are given not only the values of flows and pressures in the network at any instant of time but also that they have some indication of how reliable these values are. The quantification of the inaccuracy of calculated flows and pressures caused by the input data uncertainty is called here confidence limit analysis. Several confidence limit analysis algorithms are presented. These include a Monte Carlo simulation method, an optimization method, and a sensitivity matrix technique. The performance of these algorithms is assessed in terms of their suitability for real-time control or design stage applications. Results are presented for a realistic test network.*

## INTRODUCTION

Recent developments in telemetry and telecontrol hardware have made efficient computer control of certain classes of remote or distributed systems a realistic proposition. One such class is that of water distribution networks. Typically, water networks cover large areas, which means that these systems will benefit particularly from the new telemetry facilities. For a water network, the telemetry system consists of a collection of meters and monitoring devices that relay information about flows and pressures at some discrete points of the distribution network. However, on their own, such telemasurements have rather limited usefulness to the operators since they only give local views of the system's operational parameters. To gain a global view of the system, the telemetry data needs to be combined with a mathematical model of the water network so that the readings can be cross-referenced and all variables of interest that are not directly measured can be calculated. This process is known as state estimation and has been used in the water industry for a number of years (Coulbeck 1977; Piotrowski 1978; Rao *et al.* 1974; Sterling and Bargiela 1984a; Stimson and Brameller 1981). Depending on the mathematical model, the state estimates may represent nodal pressures or pipe flows or a combination of the two; however, the minimal number of state estimates is always the same for a given network. For this reason, state estimates are a convenient means of describing water systems.

Generally the number of telemasurements available to the state estimator is restricted by their cost. The cost of installing a meter in the network and setting up the necessary communication link with the control room limits the number of meters that can be used. As a result the telemetry data by itself is insufficient to calculate the state of the system and will have to be augmented by predictions of consumptions or nodal pressures, which are referred to as "pseudomeasurements".

Associated with the measurements and pseudomeasurements are inaccuracies which can lead to error in the state estimates. In particular, the errors associated with pseudomeasurements can be very large since these values are at best good guesses (Rayes and Wood 1981; Walski 1983). The uncertainty in measurement data is transferred through the state estimation process and results in inaccurate state estimates. This problem has been recognized for some time as a main reason for scepticism about the results produced by the classical state estimators, which implicitly assume that the snapshot measurement data

is consistent and accurate, (Coulbeck 1977; Piotrowski 1978; Rao *et al.* 1974; Sterling and Bargiela 1984a; Stimson and Brameller 1981).

There are also other causes for inaccuracy of the water system state estimates. Inaccurate network modeling and inaccurate values for pipe parameters are examples of two other important sources of error. Walski (1983) discusses the problem of inaccurate pipe parameters, C-values in particular, and suggests ways of calibrating a network model. Eggener and Polkowski (1976) examine the problems caused by invalid modeling assumptions. These problems are considered separate to that of quantifying the effects of pseudomeasurement and real measurement inaccuracy and are not addressed in this paper. The most distinctive feature of our analysis is that it concerns the errors of computer simulations caused by random variations of consumptions while the problems tackled in Walski (1983) and Eggener and Polkowski (1976) concern a deterministic inadequacy of a network model leading to systematic errors. Further, the results of our confidence limit analysis indicate that even with exactly known pipe parameters, the calculated flows and pressures can be significantly different from their actual values. This seems to be in agreement with the intuitive feelings of many system operators who are skeptical about the applicability of present day simulators for real time monitoring and control.

In "Water Distribution Research" (1974), the AWWA Research Committee remarks that "... the major source of error in simulation of contemporary performance will be the assumed loading distributions and their variations," and encourages research in this area. However, the writers are not aware of any publication proposing a solution to this problem. The confidence limit analysis is a natural extension of the earlier research into topological observability of water networks (Bargiela 1985a) which aimed at determining a set of measurements and pseudomeasurements that make the estimation process mathematically feasible regardless of the accuracy of input data.

This paper addresses the problem of the interrelation between the quality of measurement data and the quality of the results and presents several confidence limit analysis algorithms. The paper is

\* Dept. of Computing, Trent Polytechnic, Burton St., Nottingham, NG1 4BU United Kingdom. Reprinted by permission from ASCE, *Journal of Water Resources Planning and Management* (ISSN 0733-9496). Vol. 115, No. 2, March, 1989. Copyright 1989 by American Society of Civil Engineers.

organized as follows. The rationale behind the computation of confidence limits is discussed in the next section, and the section after that gives an overview of deterministic state estimation. This is followed by a description of the Monte Carlo simulation technique and an analysis of the results it produces for a medium-sized, realistic water distribution network. Although computationally inefficient, Monte Carlo simulation gives a useful reference point for comparison of further results. Following this, new confidence limit analysis algorithms are introduced that use a linear approximation of the network model. The results obtained using these algorithms are compared to those of Monte Carlo simulation. An account of the computational performance of all the methods discussed is also given in this section. Some practical aspects of confidence limit analysis are then discussed, after which conclusions are reported.

## UNCERTAINTY IN WATER SYSTEMS

The relationship between the quality of measurement data and the quality of simulation results is a function of several factors. These include the meter type and accuracy, the location of the metering sites, and the topology of the water network. In this section the interrelation of all these factors is discussed.

The measurement data come from two sources. Flow and pressure meters placed throughout the network relay observations via the telemetry system. These provide the most accurate type of data but unfortunately their numbers are often restricted. As a result, estimates of nodal consumption are heavily relied upon. It is these data that introduce much of the uncertainty to the system. It is possible to obtain accurate demand predictions for the network as a whole (Bargiela and Sterling 1985) but estimating nodal consumptions for nodes where the population is low is more difficult. If possible, the number of meters used in the telemetry system should be increased. In this way, a greater level of accuracy can be achieved reducing the reliance on the less accurate pseudomeasurement data.

Generally, the addition of a meter to the network will increase the accuracy of the state estimates (in some special circumstances this may not be the case; this situation is discussed later in this section). As an example, consider the effect of adding a pressure meter at a node in the network. It is clear that the head at this node can now be estimated more accurately, provided, of course, the meter is accurate enough. The head estimates in the neighboring nodes will also show an improvement but the amount of improvement will decrease as distance from the meter increases. The exact extent of the meter's region of influence will be dependent on, among other things, the accuracy of the nodal consumption estimates in the neighboring nodes, the accuracy of the flow estimates in this area, pipe sizes, and the presence of any other meters nearby. The flow variables close to the new head meter may also show an improvement of accuracy, though this will be less marked. Similarly, the addition of a flow meter to the network will reduce the potential error in the neighboring flow and head variable estimates, although the region of influence may have a different shape and the flow and pressure variables will be affected to a different degree.

These effects are demonstrated on a small network illustrated in figure 1. Full network data are given in tables 1, 2, and 3. Figures 2 and 3 show the effect of adding a head meter. On these diagrams, the values by the nodes represent error bounds for the pressure variables, that is, the maximum amount by which the state estimate for these variables may be in error. Similarly, the values by the pipes represent the flow error bounds. In both cases the error bounds were calculated by enumerating the feasible (satisfying mass balance constraints) measurement values to obtain maximum and minimum values for each variable. In the first diagram (figure 2), the network has just a minimal measurement set consisting of one reference head measurement at node 1, inflow measurements at nodes 1 and 7 and nodal consumption estimates for the remaining load nodes. In figure 3, a second, more accurate head meter is added at node 8. This gives a measurement accuracy of  $\pm 0.10$  m. The new meter has a significant effect on the head errors, the improvement being strongest in nodes closest to the new meter site. With the new, more accurate meter

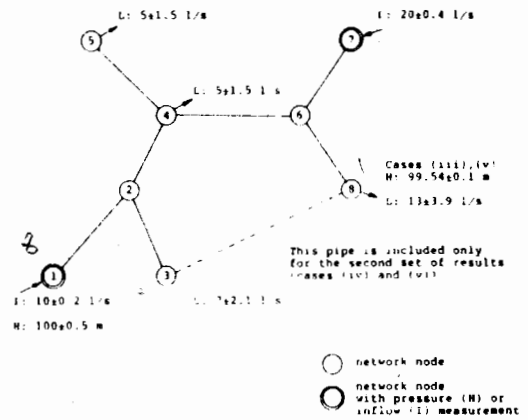


Figure 1. Pipe network used for results for cases (ii)-(v). Length, diameter and C-value of each pipe are given in table 3

Table 1. State estimates for figures 2, 3, 4, and 5

Node (1)	Figs. 2 and 3			Figs. 4 and 5		
	Head (2)	Accuracy Fig. 2 (3)	Accuracy Fig. 3 (4)	Head (5)	Accuracy Fig. 4 (6)	Accuracy Fig 5 (7)
1	100.00	0.50 m	0.32 m	100.00	0.50 m	0.22 m
2	99.79	0.51 m	0.31 m	99.79	0.51 m	0.21 m
3	99.15	0.86 m	0.67 m	99.38	0.62 m	0.15 m
4	99.75	0.56 m	0.26 m	99.70	0.54 m	0.22 m
5	99.60	0.64 m	0.34 m	99.54	0.62 m	0.31 m
6	99.78	0.59 m	0.23 m	99.72	0.53 m	0.21 m
7	99.98	0.60 m	0.24 m	99.92	0.53 m	0.22 m
8	99.54	0.72 m	0.10 m	99.42	0.62 m	0.10 m

Table 2. Network data

Pipe (1)	Fig. 2 and Fig. 3			Fig. 4 and Fig. 5		
	Flow (2)	Accuracy Fig. 2 (3)	Accuracy Fig. 3 (4)	Flow (5)	Accuracy Fig. 4 (6)	Accuracy Fig. 5 (7)
6-4	7.00	4.30 L/s	4.30 L/s	5.38	2.74 L/s	2.74 L/s
4-2	-3.00	2.30 L/s	2.30 L/s	-4.63	0.94 L/s	0.94 L/s
2-3	7.00	2.10 L/s	2.10 L/s	5.35	0.88 L/s	0.88 L/s
2-1	-10.00	0.20 L/s	0.20 L/s	-10.00	0.20 L/s	0.20 L/s
4-5	5.00	1.50 L/s	1.50 L/s	5.00	1.50 L/s	1.50 L/s
6-8	13.00	3.90 L/s	3.90 L/s	14.65	2.94 L/s	2.94 L/s
6-7	-20.00	0.40 L/s	0.40 L/s	-20.00	0.40 L/s	0.40 L/s
3-8				-1.57	2.28 L/s	2.28 L/s

Note: Pipe 3-8 is only included in the network shown in figures 4 and 5.

Table 3. Pipe data for figures 1-5

Pipe (1)	Length (m) (2)	Diameter (m) (3)	C-value (4)
6-4	1,000.0	0.30	170.0
4-2	650.0	0.30	50.0
2-3	110.0	0.15	60.0
2-1	400.0	0.20	145.0
4-5	500.0	0.20	100.0
6-8	400.0	0.25	100.0
6-7	1,000.0	0.30	170.0
3-8	500.0	0.20	100.0

Note: Pipe 3-8 only appears in figures 4 and 5.

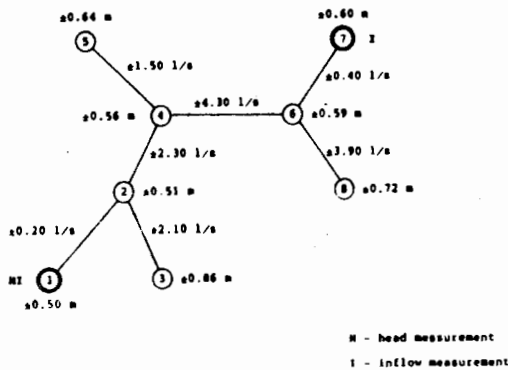


Figure 2. Error bounds for tree network with minimal measurement set

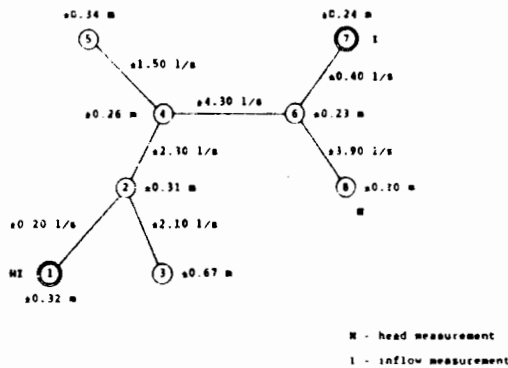


Figure 3. Error bounds for minimal measurement set with addition of head meter at 8

used in node 8, the original meter at node 1 is effectively redundant. This is because the pressure at node 1 can be estimated more accurately using the pressure measurement at node 8 than it can be measured by the original meter.

In a tree-like network, as is shown in figures 2 and 3, the error bounds are quite simple to calculate but in the more realistic situation of a looped network, computations are much more involved. The addition of a single pipe from node 3 to node 8 forming a loop in the system means that all of the measurement data have to be considered simultaneously. This is because the pressure drop around the loop must be zero. For this augmented network, the Monte Carlo method was used to obtain the error bounds and the results are presented in figures 4 and 5. The same trend is again apparent - the errors increase as the distance from the more accurate meters increases - but the exact values of the improvements obtained are less obvious. Comparing figures 4 and 5 with 2 and 3 shows that the topology of the network has a great influence on

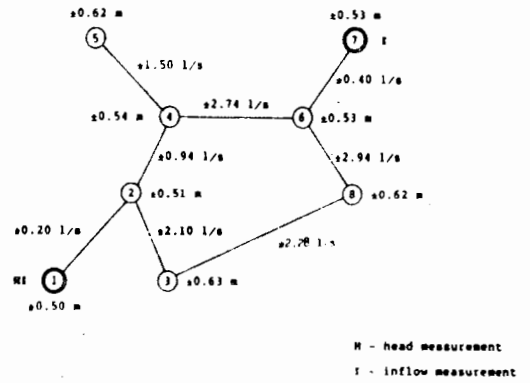


Figure 4. Error bounds for minimal measurement set on loop network

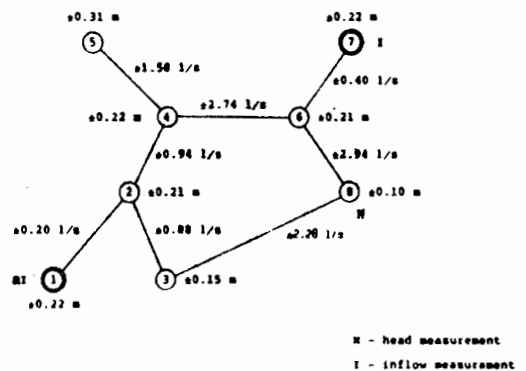


Figure 5. Error bounds for minimal measurement set with addition of head meter at node 8 on loop network

the accuracy of calculated state variables. The two networks shown only differ by one pipe but the error bounds are significantly changed.

The number of real meters in the measurement system and their accuracy are not the only factors involved. The distribution of these meters throughout the network is also very important. If all the meters are placed in one region of the network, then it will be possible to obtain accurate state estimates for variables in this region; but the accuracy of variables in other regions may be poor. This distribution effect is complicated further by the network's topology. For example, a meter placed in a weakly-connected region of the network will have little influence on the accuracy of state variables elsewhere in the network. Similarly, a flow meter may have a greater impact if it is placed in a pipe that has a large flow, since this will mean that a larger proportion of the water flowing through the network is being measured. A combination of meter distribution, meter accuracy, and network topology may mean that certain meters are redundant. That is to say, the accuracy of system variables will not be adversely affected by the removal of such a meter. This situation occurs when a meter is measuring a variable that can be estimated more accurately using data from other meters in its vicinity.

The operating state of the system can have a large effect on the way in which the measurement uncertainty is passed on to the state estimates. In many water networks, the flow pattern can change considerably throughout a day's operation. The accuracy of nodal consumption estimates may

similarly alter as a result of these changes. Also, some flow meters have an accuracy that is dependent, in absolute terms, on the size of the flow they are measuring. Consequently, one particular measurement configuration may provide differing levels of accuracy under different operating conditions or at different times of the day.

## STATE ESTIMATION AND NETWORK MODEL

This section gives a brief explanation of the state estimation procedure employed in processing telemetered data.

A water distribution network consists of a collection of reservoirs, pipes, pumps, and valves connected to form a network. Depending on the required accuracy of the modeling, an individual link in the network may represent either a single hydraulic element (pipe, pump, or valve, for instance), or a group of elements lumped together. Similarly, network nodes can represent either physical junctions between the various elements or sub-networks across which there is a relatively small pressure drop. The end result is a network of nodes connected by hydraulic elements for which the following two rules are satisfied.

### Mass balance rule

The sum of inflows, outflows, and a relative change of the stored volume of water is equal to zero for every node and for every sub-network of the original network.

### Head loss rule

The total pressure drop around a loop of links in the network is equal to zero.

Using these rules and the functional relationships between flow and pressure drop for every link of the network, both measurements and pseudomeasurements can be expressed in terms of nodal pressures and system supplies. This takes the form of a series of network equations which can be represented by

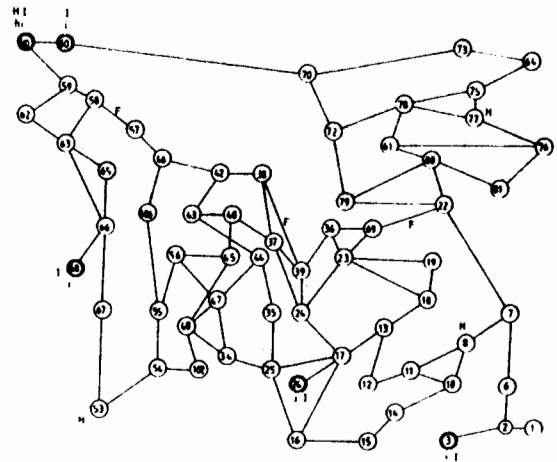
$$g(x) = z \quad (1)$$

where  $x$  = the state vector representing pressures in all network nodes and inflows from all source nodes;  $z$  = the measurement vector representing both the real measurements and pseudomeasurements, and  $g(\cdot)$  is a vector function combining all of the network equations. Equation 1 is called the network model.

The nonlinearity of  $g(\cdot)$  means that the network model cannot be solved directly and an iterative technique must be employed. This starts with an initial guess for the state vector which is progressively refined until a vector  $x$  is found that satisfies Eq. 1 as closely as possible. The iterative procedure requires at each step:

1. A linearization of the nonlinear vector function  $g(\cdot)$  around the current estimate for the state vector. This takes the form

$$z = g(x^k) + J \cdot dx \quad (2)$$



H.I. - head and inflow meters for NS1  
H.I.F. - head, inflow and flow meters for NS2

Figure 6. Test network

where  $k$  = the index of steps in the iteration procedure;  $x^k$  = the current estimate for the state vector;  $z$  = the measurement vector;  $g(x^k)$  = the vector function  $g(\cdot)$  evaluated at  $x^k$ ;  $J$  = the Jacobian matrix evaluated at  $x^k$ ;  $J = \partial g(x)/\partial x$ ; and  $dx$  = the correction vector.

2. A solution of Eq. 2, which is a system of linear equations, to obtain the correction vector  $dx$ . The next estimate for the state of the system can be calculated from Eq.3:

$$x^{k+1} = x^k + dx \quad (3)$$

There are various state estimators, each using a different numerical technique to solve equation 2. A common feature they all share is that the computation is done for a fixed measurement vector  $z$  which does not change throughout the computation. This is a serious weakness of these deterministic estimators since it implies that the calculated flows and pressures will not hold if, as is usually the case, some of the nodal consumptions are in error. The situation may become critical if the network contains pressure-controlled or flow-controlled devices such as pressure- or flow-reducing valves, one-way valves or altitude-controlled valves. A deterministic state estimate, based on inaccurate measurement data, may indicate that these devices are in one operational state, when they are, in fact, in the opposite state. Consequently, an operator relying on the deterministic estimates would be misled about the topology of the system that is implied by the closed/open status of the valves.

The algorithms presented in this paper overcome the limitations of deterministic state estimators by calculating ranges of feasible values attainable by flows and pressures within the network under a given variation of real measurements and pseudomeasurements. These results have a dual application. First, they provide guidance on the required accuracy of measurements and second, they enable the real-time identification of all operational states in which the operator could be misled by the results of by the deterministic algorithms.

## MONTE CARLO SIMULATION

In normal use, deterministic state estimators produce one state estimate for one measurement vector. Used in this way, they give no indication of how a state estimate may be affected by the fuzziness of input data. Alternatively, if a deterministic state estimator is used repeatedly for a whole range of measurement vectors, then some indications of state estimate variance are provided. This idea forms the basis of the Monte Carlo approach to confidence limit analysis in water systems.

The uncertainty in the measurement data means that the measurement vector, instead of being single valued, can take any one of a whole range of feasible states. In the Monte Carlo method, a sequence of measurement vectors is chosen randomly from within this range. Each of these is then tested against the balance constraint to ensure that the amount of water leaving the network, as defined by the nodal consumption measurements, is equal to the amount of water entering the network, as defined by the inflow measurements. Using this randomly generated measurement vector, a deterministic state estimate is calculated. Each state estimate,  $x$ , would then be checked to see that it satisfies all of the constraints specified by the mathematical model of the system (equation 1); if not, then it must be rejected. Otherwise it is a feasible state estimate and can be used to update the set of feasible state estimates. At the start of the Monte Carlo simulation, the state error vector is zero. As soon as a feasible state estimate is found, it can be used to define the current maximum and minimum feasible values for each state variable. Any new feasible state estimate found is compared against the maximum and minimum values; if, for any state variable, either of these bounds is violated, new maxima can be defined. In this way the error bounds for the state variables can be gradually increased until, after many trials, their limits are asymptotically reached. In a system with many variables, as is the case for water distribution networks, many trials are required. Since each trial requires the calculation of a state estimate, this method is computationally slow. When the total number of measurements and pseudomeasurements exceeds the number of state variables, a large proportion of the trials has to be rejected due to measurement inconsistency. This further adds to the computation time.

Results produced by a Monte Carlo simulation on the test network, shown in figure 6, are presented in tables 4 and 5. The full description of the network

Table 4. Typical state estimates and errors--heads

Node (1)	Head (m) (2)	Measurement Set 1 <sup>a</sup>			Measurement Set 2 <sup>a</sup>	
		Error 1 <sup>b</sup> (m) (3)	Error 2 <sup>c</sup> (m) (4)	Error 3 <sup>d</sup> (m) (5)	Error 2 <sup>c</sup> (m) (6)	Error 3 <sup>d</sup> (m) (7)
1	140.11	2.26	2.14	2.15	0.47	0.47
8	140.02	2.07	1.95	1.96	0.10	0.10
22	140.29	1.29	1.28	1.30	0.52	0.58
38	140.33	1.27	1.27	1.27	0.53	0.63
45	140.02	1.36	1.37	1.37	0.63	0.67
53	140.92	1.20	1.22	1.22	0.10	0.10
62	142.84	0.57	0.59	0.59	0.37	0.34
77	140.37	1.33	1.29	1.30	0.10	0.10
78	140.35	1.32	1.28	1.29	0.16	0.15
160	144.77	0.03	0.03	0.03	0.03	0.03

<sup>a</sup>See Fig. 6.

<sup>b</sup>The state error bounds produced for this measurement set using the Monte Carlo method.

<sup>c</sup>The state error bounds produced using the sensitivity matrix method.

<sup>d</sup>The state error bounds produced using the optimization method.

Table 5. Typical state estimates and errors--flows

Pipe (1)	Flow (L/s) (2)	Measurement Set 1 <sup>a</sup>			Measurement Set 2 <sup>a</sup>	
		Error 1 <sup>b</sup> (L/s) (3)	Error 2 <sup>c</sup> (L/s) (4)	Error 3 <sup>d</sup> (L/s) (5)	Error 2 <sup>c</sup> (L/s) (6)	Error 3 <sup>d</sup> (L/s) (7)
1-2	4.77	1.91	1.91	1.91	1.91	1.91
11-10	-2.10	1.55	1.48	1.47	1.31	1.33
18-19	-3.49	5.91	6.05	6.05	4.39	4.34
22-69	13.40	19.33	20.25	20.48	0.50	0.50
38-39	0.00	17.09	21.52	21.55	0.50	0.50
36-69	-12.25	18.59	19.38	19.38	1.00	1.17
61-76	-1.45	1.00	1.12	1.12	1.23	1.24
58-57	-24.06	6.91	7.03	7.05	0.50	0.50
46-42	-15.89	11.21	11.74	11.85	4.98	4.98
60-160	-6.63	4.88	4.72	4.72	3.53	3.53

<sup>a</sup>See Fig. 6.

<sup>b</sup>The state error bounds produced for this measurement set using the Monte Carlo method.

<sup>c</sup>The state error bounds produced using the sensitivity matrix method.

<sup>d</sup>The state error bounds produced using the optimization method.

parameters can be found in Bargiela (1985b). It should be noted that because of small number of real meters (five inflow meters and a single reference pressure), this example illustrates mainly the effect of uncertainty in nodal consumption predictions. The error bounds for these data ranged from 30-100 per cent. The potential inaccuracy of the state estimates calculated using such data was found to be up to 2.0 m for head variables and 15.0 L/s for flow variables. This means that for many pipes not only the value but also the direction of flow is uncertain. It should be noted, however, that these results will clearly improve when extra meters are added. Conversely, if the variations of nodal consumptions are assumed to be wider, the uncertainty of the system state will also increase.

Although the Monte Carlo method is effective in providing realistic state error vectors, its computational complexity tends to be a major drawback. Even for a medium-sized system, as discussed in the paper, the number of feasible measurement vectors is enormous, thus rendering this approach impractical for routine use. In view of these limitations, two alternative methods have been developed. In both of these, an accurate linearization of the system model is used to reduce the mathematical complexity. The first uses the linearized network equations as constraints in a mathematical optimization problem and the second uses them to construct a sensitivity matrix.

## UNCERTAINTY QUANTIFICATION METHODS

The first problem of quantification is expressing the uncertainty in the measurement data and in the state variables. The accuracy of each variable or measurement value should be assessed independently, since a particular meter configuration will mean that some variables can be estimated accurately and others poorly. So it is not possible to describe how accurate a meter configuration is by just one value. Also, certain measurement values are more reliable than others; for instance, telemetry data are generally more accurate than are nodal consumption estimates. The method chosen for representing the uncertainty is to define for each measurement value an error bound and calculate the corresponding error bounds for each state variable. To explain the use of these error bounds: if a particular state variable is estimated to have a value  $(x)_j$  and an error bound  $(e)_j$  then its true value is no more than  $(x)_j + (e)_j$  and no less than  $(x)_j - (e)_j$ . This approach provides a degree of flexibility while sufficiently defining the confidence limits required by control applications (Sterling and Bargiela 1984b).

The deterministic network model, represented by equation 1, must be altered to account for the uncertainty of the measurements. In the deterministic situation the measurement vector,  $z$ , is assumed to be single valued. In the nondeterministic model, it is assumed to have a whole range of values from within the region bounded by  $z_0 + z_u$  and  $z_0 - z_l$ , where  $z_0$  is a deterministic measurement vector and  $z_u$  and  $z_l$  are the vectors of upper and lower error bounds for the measurements. Mathematically, this can be expressed as the region  $Z$ , where

$$Z = \{z: z_0 - z_l \leq z \leq z_0 + z_u\}$$

In this format the network model, Equation 1, becomes

$$z_0 - z_l \leq g(x + \Delta x) \leq z_0 + z_u \quad (4)$$

where  $\Delta x$  = a state displacement vector which reflects the changes in supplies and consumptions and relates to the state error bounds in the following way:

$$-e \leq \Delta x \leq e$$

To define these error bounds for the state vector, each component of  $\Delta x$  must be maximized and minimized over the whole region of feasible operating states. For the  $i$ th state variable, the error bound can be calculated as follows:

$$(e)_i = \max \{|\max [(\Delta x)_i]|, |\min [(\Delta x)_i]|\} \quad (5)$$

where  $\max [(\Delta x)_i]$  and  $\min [(\Delta x)_i]$  = the maximum positive and negative changes of the  $i$ th variable calculated over the set of all feasible measurements and pseudomeasurements,  $Z$ .

The input data required by a confidence limit analysis algorithm consists of the same topological and measurement data as used by the deterministic state estimator, plus additional data on measurement error bounds  $z_l$  and  $z_u$ . The meters in the telemetry system will have a specified accuracy, within a certain operating range; this may take the form of a percentage error or an absolute error. The value of this error is directly derived from the manufacturer's specifications provided with the meter. In the case of pseudomeasurements, the error bounds are calculated from a knowledge of past operating states of the network and from the factors which determined the value of a particular consumption estimate such as the population distribution, the type of industrial consumption, etc.

### Optimization

In general, the values  $|\min [(\Delta x)_i]|$  and  $|\max [(\Delta x)_i]|$  are different due to the nonlinearity of network equations. If  $m$  = the number of elements in the measurement vector  $z$ ; and  $n$  = the number of state variables, then the problem described by Equations 4 and 5 represents  $2n$  maximizations and minimizations all subject to  $2m$  nonlinear constraints. Clearly, this is a relatively complex computational task. However, if the model (Equation 1) is linearized, then  $|\min [(\Delta x)_i]|$  and  $|\max [(\Delta x)_i]|$  are equal and the problem simplifies to

$$(e)_i = \max [(\Delta x)_i] \quad i = 1, 2, \dots, n \quad (6)$$

subject to

$$z \leq J \cdot \Delta x \leq z \quad (7)$$

Equations 6 and 7 can be converted to the standard linear programming form with  $2m$  inequalities and solved using one of the readily available sparse implementations of the Simplex algorithm (Reid 1975). However, this approach has not been adopted here since it means maintaining an inverse of an unnecessarily large basis matrix, thus sacrificing computational efficiency. Instead, an algorithm proposed by Roberts and Ben-Israel (1970, which capitalizes on the special structure of equation 7, has been employed. The set of inequalities Equation 7, is subdivided in the following way:

$$-z_l^i \leq J_a \Delta x \leq z_u^i \quad (8)$$

$$-z_l^i \leq J_b \Delta x \leq z_u^i \quad (9)$$

where  $J^T = [J_a^T | J_b^T]$ ;  $z_u^T = [z_u^{aT} | z_u^{bT}]$ ;  $z_l^T = [z_l^{aT} | z_l^{bT}]$ ;  $J_a$  is a nonsingular matrix of dimension  $n \times n$  formed by  $n$  linearly independent rows of  $J$ ; and  $J_b$  is a rectangular matrix formed by remaining  $m - n$  rows of  $J$ . The solution to Equations 6 and 7 is then obtained by solving a finite number of auxiliary problems of the form

$$(e)_i = \max [(\Delta x)_i] \quad i = 1, 2, \dots, n$$

from Equation 6, subject to

$$-z_l^i \leq J_a \Delta x \leq z_u^i$$

from Equation 8, and

$$-z_l^i \leq r \cdot \Delta x \leq z_u^i \quad (10)$$

where  $r$  = a row vector taken from the  $J_b$  matrix.

In the first auxiliary problem, some of the original constraints are ignored. Therefore the optimal solution to this restricted problem needs to be checked as to whether it satisfies the remaining constraints (whether it is feasible). If all the constraints are satisfied, then the optimal solution of Equations 6, 8 and 10 is also an optimal solution of Equations 6 and 7 (Roberts and Ben-Israel 1970). If some of the constraints are violated, then another auxiliary problem is formulated by deleting one of the nonbinding constraints in  $J_a$  of the previous auxiliary problem and including one of the violated constraints. This process is continued until it converges with the optimal solution.

By comparing the results obtained using the optimization method with the results obtained by Monte Carlo simulations (tables 4 and 5), it can be seen that the linearization of the mathematical model of the system does not significantly affect the values of calculated error bounds. Consequently, the method appears to be an acceptable substitute for the full Monte Carlo simulation, especially if the computational performance is also taken into account. The optimization method produced results in approximately five minutes whereas the Monte Carlo simulation requires many hours on a VAX-11/785 minicomputer under the VMS operating system.

Although the optimization method performs satisfactorily in off-line meter placement studies, it is

felt that it is not efficient enough to be used in real time for online decision support. The next section presents a more efficient algorithm which is also based on the linearized network model but makes better use of the specific structure of the problem.

### Sensitivity matrix

In the section on state estimation, the Jacobian matrix was introduced. At that time it was used to calculate the correction vector at each iteration of the state estimation process through Equation 2. Alternatively, Equation 2 can be viewed as an equation relating the sensitivity of the measurement vector to changes in the state vector. In this way, the Jacobian becomes a sensitivity matrix, forming the basis of the sensitivity matrix approach to confidence limit analysis.

Let  $\Delta z$  represent the difference  $z - g(x^t)$ . If  $x^t$  in Equation 2 is replaced by the true state vector for the system,  $\Delta z$  will then represent the difference between the measured vector and the true values of the measured variables. In this way, measurement uncertainty can be introduced into the system model.

Of course, the true state of the system is not known. The best estimate available is obtained by assuming the measurement vector to be true and using a deterministic state estimator. For the rest of the paper,  $J$  represents the Jacobian matrix calculated from this state estimate. The symbol  $\Delta z$  will represent the discrepancy between the measurement vector and the true value of measured variables and  $\Delta x$  will represent the discrepancy between the calculated state vector and the true state of the system. With these symbols Equation 2 becomes

$$\Delta z = J \cdot \Delta x \quad (11)$$

It is reasonable to assume that  $\Delta z$  is bounded by the measurement error vectors  $z_i$  and  $z_u$ . The aim of confidence limit analysis is to find a state error vector  $e$  which bounds  $\Delta x$  in the same way. Equation 11 can be used to do this, but first it must be inverted, to give  $\Delta x$  in terms of  $\Delta z$ .

Since the total number of measurements and pseudomeasurements is greater than the number of state variables, the  $J$  matrix is not square. For this reason  $J$  has no proper inverse and a pseudoinverse has to be used. The inverted form of Equation 11 is

$$\Delta x = (J^T J)^{-1} J^T \cdot \Delta z \quad (12)$$

Where  $J^T$  is the transpose of  $J$  and  $(J^T J)^{-1}$  is the inverse of  $(J^T J)$ . Provided that the system is observable (Bargiela 1985),  $(J^T J)^{-1}$  exists and Equation 12 is well defined. The matrix  $(J^T J)^{-1} J^T$  can now be used as a sensitivity matrix, relating changes in the state vector to changes in the measurement vector. For one state variable ( $x_i$ ), calculating its error bound is just a matter of maximizing  $a_i \cdot \Delta z$ , where  $a_i$  is the  $i$ th row of the sensitivity matrix,  $(J^T J)^{-1} J^T$ . The maximization is done with respect to the condition that  $\Delta z$  is bounded by the measurement error vectors  $z_i$  and  $z_u$ . That is

$$(e)_i = a_i \cdot b \quad (13)$$

where

$$(b)_i = (z_u)_i \quad \text{if } (a_i)_i > 0; \quad (b)_i = (z_l)_i \quad \text{otherwise} \quad (14)$$

In this way, the state error vector  $e$  can be calculated by repeating the maximization of Equation 13 for each state variable.

The maximization equation (Equation 13) can be refined to take account of other system constraints. Perhaps the most important of these is the balance constraint, which specifies that the total amount of water entering the network at any one time is equal to that leaving it. But the procedure presented here represents the core of the sensitivity matrix method. First a state estimate is produced on the assumption that the measurement vector is correct. Then the possible error of the measurement set is considered and it is used, together with the sensitivity matrix, to predict the possible error in the state vector.

This approach requires significantly fewer mathematical operations required to calculate confidence limits than other methods. The nature of water network equations means that the matrix  $J$  is sparse. This means that its proportion of non-zero elements is low. This leads to further savings in the amount of computation required. The use of sparsity exploiting methods in matrix factorization (Duff 1980) also increases numerical stability, reducing the risk of errors building up during the calculations.

Computational results can again be compared to the results obtained with Monte Carlo simulation (table 3). The linearization of the network model does not significantly affect the values of calculated error bounds. Furthermore, the results are almost identical to the ones produced by the optimization method and the minor differences are explained by the round-off in numerical processing. To produce results for the network of figure 6, the sensitivity method required 10 s of CPU time on a VAX-11/785 minicomputer.

## PRACTICAL ASPECTS

To realize the potential of the confidence limit analysis, the algorithms described in this paper have been implemented as a graphical interactive software tool called TCLAS (Telemetry Confidence Limit Analysis Software). Although the full description of the software exceeds the scope of this paper, it will be useful to enlist here its main application areas. These can be grouped in the following categories.

### Optimal telemetry design

This function of TCLAS software is concerned with the optimal positioning of telemetry outstations and the selection of required accuracies of pressure and flow meters in the water network. The software helps to ensure that any investment in telemetry hardware will produce the best returns in terms of accuracy of system monitoring. The minimization of the number of meters employed to achieve a given monitoring accuracy results in savings in the following: the cost of extra meters, the cost of their installation, the cost of associated telemetry, and the maintenance costs.

### Leakage detection

The confidence limit analysis of flows and pressures in the network gives an indication of the feasibility of detecting leakage with a given level of instrumentation. Since the TCLAS quantifies the

uncertainty about the flows in all pipes in the network as a function of the accuracy of telemetering, it can be conveniently used for evaluating an upper limit on the volume of a potential leakage in any given area of the network as a sum of uncertainties of supplies into this area. By comparing the cost of potential water losses with the cost of telemetering, the user can decide on an appropriate balance of these two concerns.

#### Operator training and decision support

This function concerns the online execution of TCLAS software where the system operator is presented, in real time, with envelopes of feasible flows and pressures rather than with their "average" numerical values. Consequently, the TCLAS approach overcomes the shortcomings of deterministic simulations which are often mistrusted by system operators, since the "average" flows or pressures rarely, if ever, agree with the actual meter readings. By providing better reference values, in the form of confidence limit envelopes, TCLAS promotes better understanding of the water system's behaviour. This in turn translates into cost savings stemming from the avoidance of, for example, unstable operation of pressure reducing valves producing pressure surges and consequent pipe bursts, prevention of undesirable low or high pressure profiles etc.

#### CONCLUSIONS

This paper examines the problem of telemetry-related uncertainty in water systems, its causes, and its consequences.

Present-day deterministic state estimation techniques are very efficient, having small computational requirements and producing results of an acceptable level of accuracy. But no state estimator can give accurate results from inaccurate data. Due to the cost of metering, the water industry is and will be, in the foreseeable future, making use of relatively inaccurate pseudomeasurements. For this reason, the computationally accurate results of state estimators can be very inaccurate when compared to the actual system state. Monte Carlo simulation has shown that this uncertainty can have a very significant influence on the reliability of the state estimates derived from the measurement data. The degree of confidence that can be put in these results must be calculated and presented with the state estimates themselves. Only then can the computation results be used in operational control.

There are many factors involved in the interdependence between measurement accuracy and state estimate accuracy. These include: the type, the number, and the accuracy of meters used in the system; the placement of those meters; the network's topology; and the operational state of the distribution system. These complex and interacting factors mean that any confidence limit algorithm must tackle the problem from a global viewpoint, considering all of the measurements simultaneously. This is the way the three confidence limit analysis algorithms presented in this paper have been constructed.

The Monte Carlo simulation technique generates a set of feasible state estimates and from these calculates upper and lower error bounds for each state variable. To guarantee the validity of the Monte Carlo results, a massive number of state estimates must be used. For this reason, the Monte Carlo method is an unrealistic proposition for real-time application. Conversely, the results it produces are very reliable mathematically.

This is because any error bound or confidence limit produced by the Monte Carlo programme is attainable, i.e., there is a feasible state estimate that will reach this bound. This mathematical reliability means that this approach provides a yardstick against which other algorithms can be tested.

The optimization and the sensitivity matrix methods presented in this paper are both based on a linearized system model. However, the results produced by these methods compare very well with the results of the Monte Carlo simulations. The computational efficiency of the sensitivity algorithm presented in the sensitivity matrix section renders it suitable for online decision support applications.

#### REFERENCES

- Bargiela, A. (1985a). "An algorithm for observability determination in water system state estimation." *IEE Proc.*, Part D, 132, 245-249.
- Bargiela, A. (1985b). "Real time network analysis and leakage detection-final report" Dept. of Computing, Trent Polytechnic, Nottingham, United Kingdom.
- Bargiela, A., and Sterling, M. (1985). "Adaptive forecasting of daily water demand." *Comparative models for electrical load forecasting*. D.W. Bunn and E.D. Farmer, eds., John Wiley and Sons, New York, N.Y.
- Coulbeck, B. (1977). "Optimisation and modelling techniques in dynamic control of water distribution systems," thesis presented to the University of Sheffield, at Sheffield, Eng., in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
- Duff, I.S. (1980). "MA28 - a set of Fortran subroutines for sparse unsymmetric linear equations. Report R 8730. AERE Harwell, United Kingdom.
- Eggerer, C., and Polkowski, L. (1976). "Network models and the impact of modeling assumptions." *J. AWWA*, 68(4).
- Piotrowski, J. (1978). "State estimation in water systems." *Water Distribution Systems Conference*, Ustron, Poland.
- Rao, H.S., Bree, D.W., and Benzvi, R. (1974). "Extended period simulation of water distribution networks." Report PB-230 148, Systems Control Inc, Palo Alto, California.
- Rayes, A.J., and Wood, D.J. (1981). "Reliability of algorithms for pipe network analysis." *J. Hydr. Div.*, ASCE, 107(HY10).
- Reid, J.K. (1975). "A sparsity-exploiting variant of the Bartel-Golub decomposition for linear programming bases." Report CSS 20, AERE Harwell.
- Roberts, P.D., and Ben-Israel, A. (1970). "A sub-optimisation method for interval programming: A new method for linear programming." *Linear algebra and its applications*, 3, 383-405.
- Sterling, M. and Bargiela, A. (1984a). "Minimum norm state estimation for computer control of water distribution systems." *IEE Proc.*, 131(2), 57-63.
- Sterling, M., and Bargiela, A. (1984b). "Leakage reduction by optimised control of valves in water networks." *Trans. Inst. Meas. Control*, 6, 293-298.
- Stimson, K.R., and Brameller, A. (1981). "An integrated mesh-nodal method for steady state water distribution network analysis." *J. Inst. Water Engrs. and Sci.*, 35(2).
- Walski, T.M. (1983). "Techniques for calibrating network models." *J. Water Resour. Plng. and Mgmt.*, ASCE, 109(4).
- "Water distribution research and applied development needs." (1974). American Water Works Association Research Committee on Distribution Systems, *J. AWWA*, 66(6).

## NOTATION

*The following symbols are used in this paper:*

- $a_i$  =  $i$ th row of the Jacobian matrix;
- $dx$  = correction vector in state estimation;
- $e$  = state error vector;
- $g(\cdot)$  = vector function combining all network equations;
- $J$  = Jacobian matrix  $J = (\partial g(\mathbf{x})/\partial \mathbf{x})$ ;
- $J_a$  = matrix consisting of first  $n$  rows of  $J$ ;
- $J_b$  = matrix consisting of remaining rows of  $J$ ;
- $n$  = number of independent state variables;
- $m$  = number of measurements and pseudomeasurements;
- $\mathbf{r}$  = row vector taken from  $J_b$ ;
- $\mathbf{x}$  = state vector;
- $\mathbf{x}^k$  =  $k$ th estimate in state estimation procedure;
- $Z$  = range of feasible measurement vectors;
- $\mathbf{z}$  = measurement vector;
- $\mathbf{z}_0$  = average measurement vector;
- $\mathbf{z}_l$  = lower limit on measurement error;
- $\mathbf{z}_u$  = upper limit on measurement error;
- $\mathbf{z}_l^a$  = vector consisting of first  $n$  elements of  $\mathbf{z}_l$ ;
- $\mathbf{z}_u^a$  = vector consisting of first  $n$  elements of  $\mathbf{z}_u$ ;
- $\mathbf{z}_l^b$  = vector consisting of remaining elements of  $\mathbf{z}_l$ ;
- $\mathbf{z}_u^b$  = vector consisting of remaining elements of  $\mathbf{z}_u$ ;
- $z_l^r$  = element of  $\mathbf{z}_l$  corresponding to vector  $\mathbf{r}$ ;
- $z_u^r$  = element of  $\mathbf{z}_u$  corresponding to row  $\mathbf{r}$ ;
- $\Delta \mathbf{x}$  = state discrepancy vector;
- $\Delta \mathbf{z}$  = measurement discrepancy vector.

### Subscripts

- $u$  = upper, with respect to measurement error vector;
- $l$  = lower, with respect to measurement error vector;
- $i, j$  =  $i$ th or  $j$ th elements of a vector.

### Superscripts

- $k$  = current estimate in iteration;
- $a$  = relates to  $J_a$  matrix;
- $b$  = relates to  $J_b$  matrix;
- $T$  = transpose of matrix;
- $-1$  = inverse of matrix.