PARALLEL PROCESSING FOR PROBABILISTIC DECISION SUPPORT IN WATER DISTRIBUTION SYSTEMS

Bargiela A., Argile A., Hartley J.

Department of Computing,
The Nottingham Trent University,
Burton Street, Nottingham NG1 4BU, U.K.

Abstract

With the development of extensive telemetry schemes in water industry and the significant decrease of the cost of computing hardware, the development of comprehensive decision support systems becomes a realistic proposition. There are, however, some important issues that need to be addressed before such systems become a reality. Firstly, the introduction of decision support systems must integrate well with the existing computer-based systems in water industry. Secondly, they must by flexible enough to support a broad spectrum of control activities, from operational control, through management level control, to system development planning. And thirdly, they must afford closer approximation of the physical reality of water systems by taking into account the probabilistic nature of events that drive the water system operation.

The results of our research, summarised in this paper, have been applied in the industrial context and they appear to provide effective solutions to some of the above issues and indicate their feasibility for industrial implementation.

Keywords: Parallel processing, Decision support systems, Operational control, Water distribution systems, Uncertainty processing, Confidence limit analysis, Probabilistic mathematical models

Introduction

The water industry, together with other public utilities, is rapidly expanding installed telemetry systems and the idea to extend the use of such systems from simple monitoring to computer assisted control of supply and distribution is being incorporated into the development plans of many water companies. Such schemes are also known as Decision Support Systems. They include a full spectrum of processing tasks ranging from data acquisition and operational control, through management level control to system development planning.

Parallel/Distributed Processing Rationale

Operation of water distribution systems requires a variety of decisions to be made. There are system development decisions: where, when and what new elements of the distribution system need to be built. There are system management decisions concerning the regulatory measures such as water pricing principles, water quality standards, legislative measures etc.. There are operational decisions determining such things as water pumping schedules, reservoirs’ operation and pressure control in a distribution network. These decisions are made at different time intervals, with very different time horizons in mind and they are the responsibility of different groups of people. Yet all these decisions are based ultimately on common data reflecting system performance, all be it, on various degrees of abstraction of this data.

In such a situation it seems eminently sensible to develop a flexible distributed computing environment which integrates various aspects of system operation, by allowing exchange of data between applications while maintaining the autonomy of different activities. Indeed, this has been a strategic thinking of some of the water companies and is elaborated on in one of the subsequent sections.

For applications with a common time horizon, and in particular in operational control, the parallel/distributed processing affords the development and refinement of the software components without the disruption of the existing system. This is in sharp contrast to centralised systems where disruption of normal operation on any software upgrade, is almost unavoidable.

Probabilistic Processing Rationale

Almost any operational decision focuses on uncertainty. The predictive model of the system, that is the model which says: 'if your decision will be $\mathbf{u}$, the present state of the system is $\mathbf{x}^k$, and the external influence (eg. variation of consumptions) is $\mathbf{z}$, then the outcome variable will be $\mathbf{x}^{k+1} = F(\mathbf{x}^k, \mathbf{u}, \mathbf{z})$; where $\mathbf{u}, \mathbf{z}$ are time functions and $\mathbf{x}^{k+1}$, $\mathbf{x}^k$ are space distributed states observed in time instances k and $k+1$. Since this model underlies any reasoning by the operator about the future behaviour of the system, there is a need to formalise the treatment of the random factors affecting the outcome variables.

It is clear that the estimation of future system states $\mathbf{x}^{k+1}$, depends criticaly on the quality (accuracy) of measurements $\mathbf{z}$ that are provided as input data. In particular, errors associated with, so called, pseudomeasurements representing consumption estimates can be very large, as these values are at best good guesses, [Rayes and Wood, 1981] and [Walski, 1983]. The uncertainty in measurement data is transferred, through the state estimation process, and results in state estimates that are inaccurate to some degree. This problem has been recognised for some time as a main reason for scepticism about the results produced by the classical state estimators which implicitly assume that the input data is consistent and accurate. Consequently, our research has adopted probabilistic framework for all operational control level data processing.

X11 Based Distributed Multitasking

One of the two major strands of the current research programme addressed the development of a flexible operating environment for the execution of a decision support system. In order to ensure portability of solutions across a wide range of hardware configurations, the research focussed on building a distributed processing harness around the industry standard, X11 window system.

The basic methodology underlying this development is that of creating a virtual distributed shared memory. Being primarily a graphics software, X11 does not provide directly such facility, but it does provide means of attaching data structures, called properties, to graphics windows maintained by the X11 server. What had to be developed, was a structured way of accessing properties by distributed applications.

A protocol was developed, which uses the fact that the X11 server has the capability to grant the ownership of properties to windows. However, in order to overcome the limitation of the X11 server, which does not require the owner to free a property before granting ownership to another window, it was necessary to develop an efficient mutual exclusion algorithm for granting ownership of properties to the requester. Full description of the shared task memory for distributed computer systems based on X11 windows is given in a separate publication, [Argile and Bargiela, 1993].

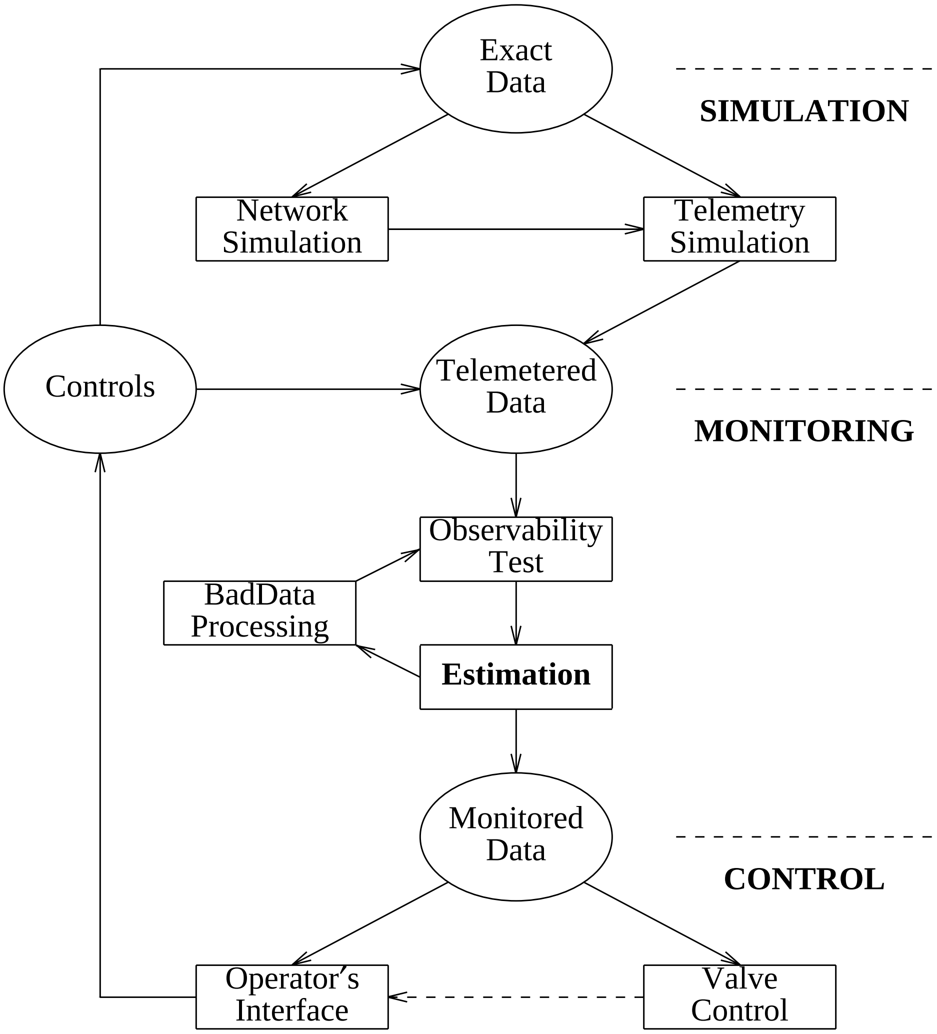

Figure 1 Water network monitoring suite

The distributed processing harness outlined above has been applied to the water network monitoring system described in detail in [Bargiela, 1984]. The overall structure of this software suite is given in Figure 1. The software essentially consists of three subsystems concerned with simulation, monitoring and control of water networks. The subsystems are configured so as to provide a classical feedback control loop. Each subsystem is composed of self-contained modules performing specific tasks.

Considering first the subsystems: the monitoring tasks expect to receive telemetry data and are entirely decoupled from the software that actually provides these readings. In one case, as illustrated in Figure 1, the data might be provided by telemetry simulation software, which corresponds to ’operator training’ or ’off-line assistance’ scenario. But in ’on-line’ operation the telemeasurements would be provided by telemetry hardware. This decoupling of subsystems enables different frequency of access to the telemetered data by the simulation and monitoring software, subject only to the mutual exclusion of data update operations.

Within the subsystems, the individual tasks are also largely autonomous. For example the observability test which is responsible for verifying whether the measurement set is sufficient for subsequent computations, may run with a frequency different to that of the state estimator which utilizes the verified measurement set. Indeed, the system allows for more than one state estimator to run concurrently (e.g. to achieve better error rejection capability), or alternatively, temporarily there may be no state estimator running.

The modular structure of the software, quite apart from its software engineering rigour, emphasizes the concurrent nature of processing tasks that occur within the context of water network monitoring and control systems. In the original implementation [Bargiela, 1984], the concurrent processing tasks were mapped onto pseudo-concurrent processing of a single CPU running under a multitasking operating system. Each task communicated with others through shared memory areas with specified access privileges and the timing of task execution and synchronization was achieved by reference to semaphores and event keys in shared data.

While the conceptual design of the system is not compromised by a single processor implementation, it is clear that the gradual increase of the complexity of individual tasks (historic trending of telemetered data, data management, optimization of control etc.) leads to the specialization of processing and the introduction of distributed processing nodes. Furthermore, the distributed processing concept promises ease of expansion into areas which are complementary to on-line operations such as medium and long term system planning, system management etc.

The distributed processing harness, developed in the course of this research, offers a flexible framework for mapping individual tasks onto the most appropriate network of processing nodes and, most importantly, it provides a physically distributed logical shared memory. The underlying mechanisms for creating and accessing such shared data areas are hidden from the user. Therefore the application program’s view of such areas is, as in a single processor implementation, a ’common’ block. The only demonstration of a distributed implementation of the shared memory is reduced efficiency of access to data, due to the need for explicit network transfers. However, in the case of a water network monitoring system, it has been found that the reduction of data access efficiency is not critical since the interprocess communication is relatively infrequent and involves small amounts of data while, at the same time, the distribution of processing benefits the computationally intensive modules.

Using a network of 4 Sun SLC Sparcstations running the simulation, telemetry, estimation, and operator interface modules respectively, the cycle time achieved for a 65 node network was approximately 10 secs. This is at least an order of magnitude better than would be expected in real life. Projecting the results for larger networks, it is expected that the communications overhead will, at worst, increase linearly with network size, so the communication to computation time ratio will actually decrease.

The main addition to the system, though not involving any change to the original subsystem concept, is the provision of a data pool task. This holds the shared data areas in one of its windows. This task also monitors the ownership status of each area, and periodically saves the data areas to disk. It also has system control options to save the areas to disk, to lock all areas and reinitialize all shared data. The subsystem tasks all have a similar visual representation, presenting the user with a standard graphical interface.

Parallel State Estimation

State estimation is a pivotal component of any operational control software suite. It enables cross-referencing of system measurements by relating the actual meter readings, to the values calculated from the mathematical model of the system. By doing so, it is often possible to identify and localise erroneous measurements and to advise the operator on appropriate remedial action.

Unfortunately, state estimation is a computationally intensive task. Although the technique has been refined by the application of sparse matrix processing methods [Bargiela, 1984] it remains quadraticaly dependent on the size of the physical network. Early attempts at overcoming this computational complexity problem focused on hierarchical structuring of the estimation algorithm [VanCutsem, 1983], but the recent advances in computing hardware point towards the distributed, parallel state estimation as a cost-effective scheme.

The second strand of the current research programme addressed the development of a parallel, nonlinear state estimator that is suitable for use in a decision support role in large water distribution systems. The proposed state estimator, builds on our earlier diakoptical simulation algorithm for nonlinear networks [Bargiela, 1992]. Full details of this algorithm have been given in [Hartley and Bargiela, 1993] and a brief summary of it is given below.

The estimator progresses in two phases. Firstly a solution to a minimum norm measurement set (simulation) problem is obtained and, subsequently, additional measurements are processed, one-at-a-time, through the application of the recursive least squares technique, yielding a state estimate for the whole system. In our current implementation, the processing of additional measurements is accomplished by the coordinating task, so the algorithm is particularly suitable for systems with low measurement redundancy (defined as a ratio of the number of all measurements to the number of measurements in the minimum measurement set). This situation is typical of water distribution networks. For systems with high measurement redundancy, the algorithm might be implemented on a non-homogeneous set of processing nodes, where the coordinating task is allocated a high performance computing node. Low data transfer volume between the processing nodes, render the algorithm relatively insensitive to the topology of the processing system [Bargiela, 1992].

The algorithm can be summarised as follows:

1. Read-in the system description data.
2. Form subsystem data packets and send them to individual solvers.
3. Calculate the uncoordinated subsystem solutions.
4. Coordinate partial solutions.
5. If the coordinated corrections from Step 4 are less than a given threshold value then STOP otherwise repeat from Step 2.
6. Calculate the initial state covariance matrix.
7. Update the state estimate by processing an additional measurement.
8. Update the state covariance matrix.
9. Modify measurement weights.
10. Repeat 7,8,9 for every additional measurement.

The structure of the estimator is depicted in Figure 2.

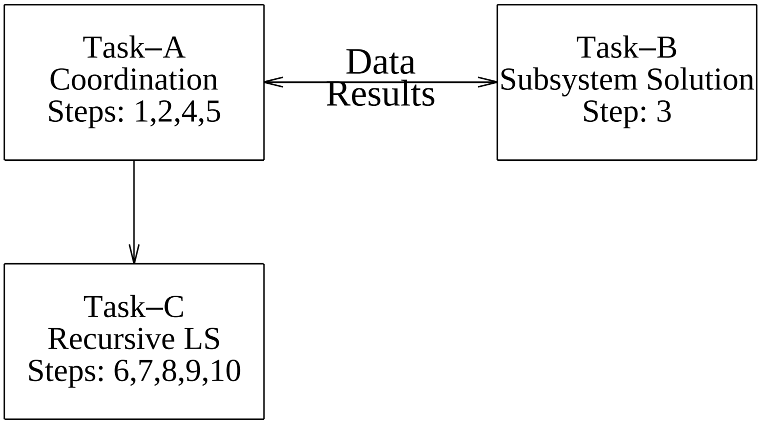

Figure 2 Paralel state estimator structure

Confidence Limit Analysis

It is very important that, in a decision support system, the level of uncertainty associated with water system state estimates is quantified. If these state estimates are to be used for control purposes it is necessary to know by how much they may be in error. The process of quantifying the uncertainty about the state estimates must however take into account also factors other than the measurement accuracy. The number of meters, their configuration and network topology all play an important role in this relationship.

Generally the addition of a meter to the telemetry system will increase the accuracy of the state estimates. For example, if a pressure meter is added to the system, the head at the relevant node will be estimated more accurately, provided, of course, the meter is accurate enough. The head estimates in the neighbouring nodes will also show some 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 depend on, among other things, the accuracy of the nodal consumption estimates in the neighbouring nodes, the accuracy of the flow estimates in this area, pipe sizes and the presence of any other meters near by.

The number of meters in the measurement system and their accuracy are not the only factors involved. The distribution of these meters throughout the network is also an important factor. If all of 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 influence of 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.

A quantitative analysis of the influences of the meter accuracy, meter distribution and network topology has been accomplished as part of our earlier research [Bargiela and Hainsworth, 1988] and has been termed 'Confidence limit analysis'.

In this work, the method chosen for representing the uncertainty was to define for each measurement value, an error bound and to 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_i and an error bound e_i, then its true value is no more than x_i+b_i and no less than x_i-b_i. This approach provides the necessary flexibility while sufficiently defining the confidence limits required for control purposes.

With measurements $\mathbf{z}$ no longer representing a point-value but a range $[\mathbf{z}-\mathbf{z}_l,\mathbf{z}+\mathbf{z}_u]$, the state estimation problem became equivalent to $2n$ minimisations, each subject to $2m$ non-linear constraints. Clearly a relatively complex computational task. Although the state estimation problem can be linearised and re-formulated as an upper- lower-bounding linear programme, it has been found, [Bargiela and Hainsworth, 1988], that the approach based on the use of sensitivity matrix provided a more effective solution.

Current research addresses the development of an analogous scheme, compatible with parallel state estimation outlined in the previous section.

Industrial Application

The Strategy

East Worcester Water, in common with many others in the Public Utilities, have been forced to re-examine their Management Information Systems. This has come about by the improvement in management practices not being matched by Management Information Systems. Applying accountability to managers has to be supported by the information the manager needs in the right form and at the right time.

East Worcester Water believes that its business can be supported by a MIS providing a number of strategic databases, all linked together and maintained by a 'once-off' entry system. The strategic databases are then available to be used by application software packages. Figure 3 shows the approach.

The compatible data bases will be centered on geographical, financial, operational and customers data. The complete system will be used for a variety of management information and decision making tasks. The information will be gathered both automatically and manually and user applications will include the familiar Customer Service, Operations, Works Management, Job Control, Design, Water Quality etc. including network analysis and network modelling.

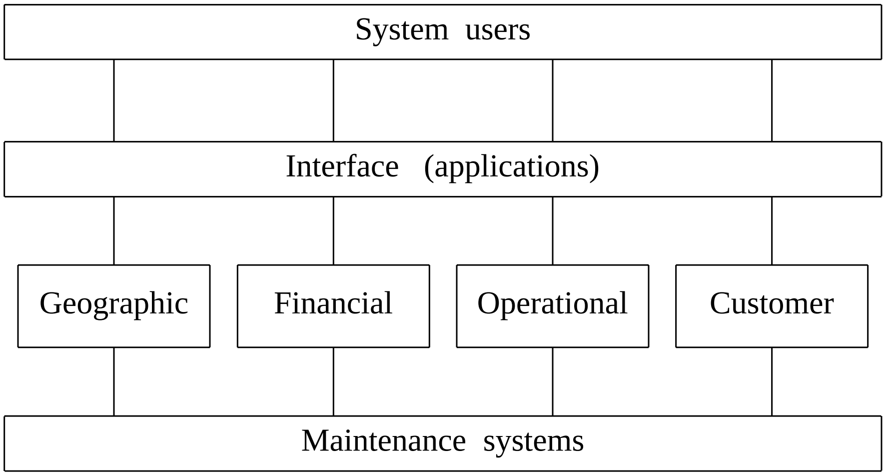

Figure 3 E.W.W. Integrated Management Information System

The work being done at Nottingham Trent University correlates well with the Integrated Management System approach adopted by East Worcester Water. The probabilistic decision support system, developed through the SERC research grant, fits the overall structure of figure 3.

The System

User applications for the water industry have traditionally been provided using mainframe systems for multi-user access, or personal computers if the application was run by a single user. E.W.W. found that mainframe applications were becoming difficult to support, and not particularly cost effective. The PC provided a multitude of standard application packages but lacked full multiple user access. The advent of the RISC workstations running standard operating system (UNIX), provided an opportunity to engineer the system East Worcester Water required.

They were fortunate in that a major application system, the SCADA Master Control System was scheduled for replacement in 1991/2. This replacement was to become the basic building block for the Management Information System.

The new SCADA system was built around RS6000 workstations and included a number of X-terminals all connected to a ’token-ring’ network (figure 4). The software would be operating under AIX and the application packages would, in the main, be standard commercially available software with only specialist software being written to interrogate remote outstations. The new system would also provide the Operational Database shown in figure 3.

With the Operational and Geographical databases available, the Network Analysis/Confidence Limit Analysis applications can be implemented with the minimum of application-specific data and provide the opportunity to implement the results produced by the Nottingham Trent research team.

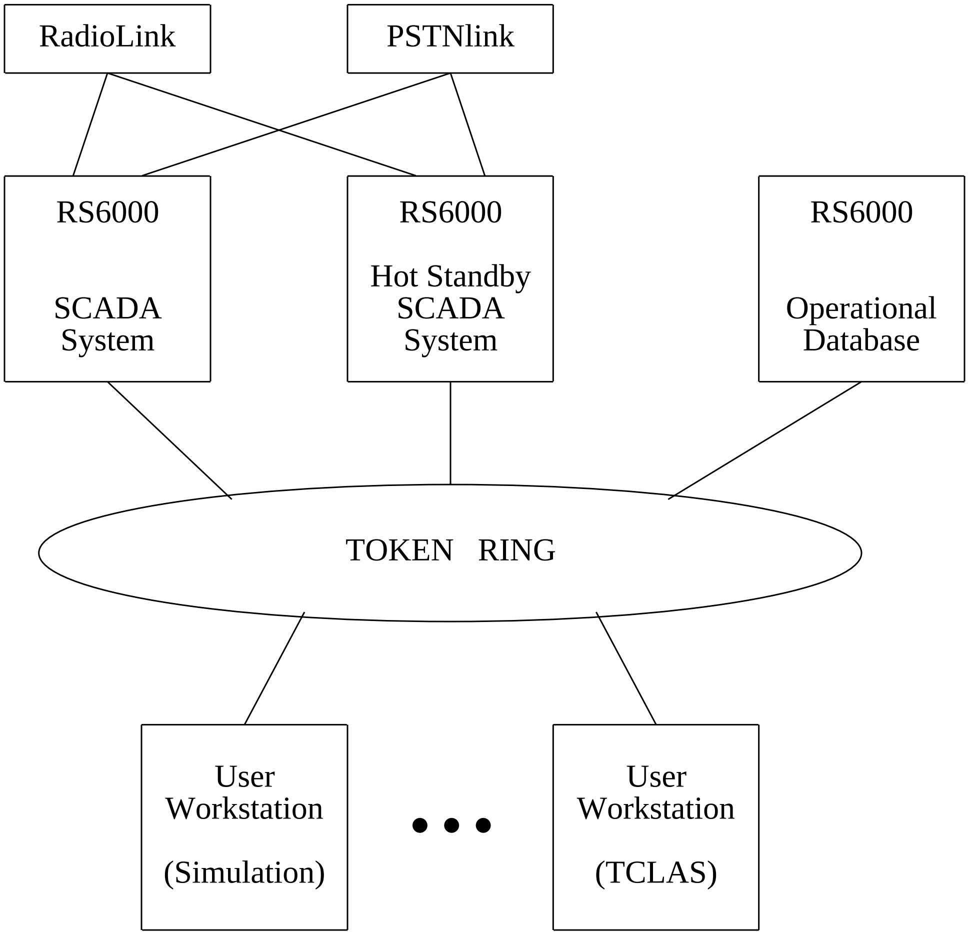

Figure 4 Implementation of the E.W.W. SCADA System

Summary and Conclusions

The research into Parallel Processing for Probabilistic Decision Support answers the real need in water industry for dependable and flexible Management Information Systems. The piecemeal approach to the development of various data processing systems, frequently adopted in the past, often implied the need for duplication of information (with inherent consistency problems), and difficulty of system modification/extension.

The proposed distributed processing framework, which integrates various commercial and custom software, shows a promise of overcoming both of these difficulties. A further result of our research has been the development of a probabilistic state estimator which, unlike the commercially available deterministic estimators, relates the accuracy of the estimates to the accuracy of the input data. The developed methodology, termed Confidence Limit Analysis, is seen as having applicability extending beyond the public utility systems.

Acknowledgements

The authors wish to acknowledge the financial support of the SERC for the research and the support of the EWWC in facilitating the industrial implementation of some of the research results.

References

Argile A, Bargiela A, 1993, Using X11 Windows to provide shared task memory in distributed computer systems, Proc. of Int. Conf. on Computer Applications, Leicester, U.K.

Bargiela A, 1984, On-line monitoring of water distribution networks, PhD thesis, University of Durham

Bargiela A, 1992, Nonlinear network tearing algorithm for transputer system implementation, Proc. Transputers and Parallel Applications, TAPA ’92

Bargiela A, 1993, Uncertainty processing in water distribution systems, Proc. Int. Conf. on Computer Applications, Leicester, U.K.

Bargiela A, Hainsworth G, 1988, Confidence limit analysis in water systems, in Coulbeck B. and Orr C.H. (eds.), Systems Analysis and Simulation, Research Studies Press, John Wiley, 1988

Bargiela A, Hartley J, 1993, Orthogonal linear regression algorithm based on augmented matrix formulation, Computers and Operations Research

Bargiela A, Hartley J, 1993, Diacoptical state estimation of nonlinear networks using transputers, submitted for publication.

Boulos A, Wood D, 1991, Explicit algorithm for calculating operating parameters for water networks, Civil Engineering Systems, Vol 8

J Hartley, A Bargiela, 1993, Utilisation of smoothing formulae for describing hydraulic relationships in water networks, Proc. Int. Conf. on Computer Applications, Leicester, U.K.

Schweppe F, et al., 1979, Power system static state estimation, Parts I,II,III, IEEE Trans, PAS-89.

Van Cutsem Th, Ribbens-Pavella M, 1983, Critical survey of hierarchical methods for state estimation, IEEE Transactions PAS-102.

Walski T, 1983, Techniques for calibrating network models, ASCE Journal of Water Resources Planning and Management, 109, 4