Application of Optimization Methods to Calibration of Water Distribution Systems.

Bart G. van Bloemen Waanders

Optimization and Uncertainty Quantification Department, Sandia National Laboratories *

March 9, 2006

Abstract

Inversion algorithms are used to calibrate demands of water distribution systems by minimizing the difference between observed and predicted chemical concentration values. A systematic approach was implemented using a combination of numerical experimentation, parameter studies, global methods, and local gradient based algorithms. An appropriate finite difference step was determined to calculate objective function gradients for the inversion problem. The inversion of demands for a model calibration using field tracer observations resulted in a 26% reduction of the objective function. As a result of the inversion problem, an error in the numerical model was identified that could be responsible for some of the remaining mismatch. Future work includes improving the numerical model and developing more accurate history of concentration observations. Optimization under uncertainty strategies are proposed to help characterize variability in the model calibration problem.

1 Introduction

The ability to accurately predict pressures, velocities and chemical concentrations in a water distribution network through numerical simulation is critical to the management of field operations and to the planning of future expansions. Unfortunately, water distribution models may not produce accurate predictions as a result of 1) a lack of physics, 2) the omission of relevant detail, 3) insufficient fidelity, and 4) errors in the model inputs due to lack of knowledge or inherent variability. To improve the predictive quality of the numerical model, field experiments can be conducted in which a combination of pressures, flow rates, and chemical concentrations are observed and then compared to simulation predictions. Simulation parameters such as friction factors and demands are most commonly perturbed until the mismatch is reduced. Manually manipulating these parameters, however, is a time consuming procedure, and quickly becomes intractable if the inversion space is large. For large number of inversion variables an automated process utilizing optimization algorithms should be considered.

*Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Model calibration has been known to improve the quality of numerical models, but little research effort has been devoted to developing a detailed understanding of the behavior of optimization algorithms on water distribution networks. Some work has focused on reconciling pressure observations with simulated predictions to determine more accurate friction factors and demands. Luvizotto et al [6] present results for demand inversion using pressures and flows. Most investigations have been conducted on small datasets and relatively simple flow patterns, allowing for the use of global optimization methods. Walter et al [13] and deSchaetzen et al [3] use genetic algorithms to invert for friction factors by measuring pressures and flow rates.

Applying optimization algorithms to nonlinearly behaving dynamic systems is a challenging task and is complicated by model uncertainties and field-measurement inaccuracies. In addition to the simulator mathematics not perfectly capturing field behavior, model parameters are often not exactly representative of field conditions. Significant uncertainties are introduced during the data acquisition phase (e.g measurement equipment drift, uncompensated environmental changes) and also during the data post-processing phase (e.g such as reducing the volume of data to a manageable size). To further complicate the optimization problem, inversion formulations can be ill-posed with potentially non-unique solutions. Successfully applying optimization algorithms to model calibration to achieve an exact match can therefore be a complex process. This paper discusses the use of several optimization schemes to arrive at the final solution for both a prototype and real dataset. The goal of this investigation is to take a systematic approach to model calibration, consisting of numerical experimentation of both the prototype and the actual tracer datasets, followed by a strategic use of parameter studies, global and local optimization methods

In the case of the tracer dataset, not all uncertainties can be eliminated and therefore optimization under uncertainty formulation should be considered so that inversion problems can be solved with appropriate statistics.

2 Deterministic Optimization Formulation

We start with a mathematical description of the non-linear least squares problem which is used to minimize the difference between the calculated and target concentrations. The inversion of demands, using chemical concentrations from field experiments, requires minimizing an objective function subject to both the hydraulic and chemical transport equations:

$$\min_{Q_j^e, j\in\mathcal{N}} \quad \frac{1}{2}\sum_{k\in\mathcal{N}}\{\sum_{i\in\mathcal{N}^\star}(\hat{c}_i(t)-\hat{c}_i^*(t))^2+\frac{\beta}{2}\sum_{j\in B}R\} \tag{1}$$

$$\text{s.t.} \quad \frac{\partial\bar{c}_i(x,t)}{\partial t}+u_i\frac{\partial\bar{c}_i(x,t)}{\partial x}=0 \quad i\in\mathcal{P} \tag{2}$$

$$\hat{c}_i(t)=\frac{\sum_{j\in P_k}(Q_j\bar{c}_j(L_j,t)+Q_{j,ext}c_{j,ext}(t))}{\sum_{j\in P_k}(Q_j+Q_{j,ext})} \quad i\in\mathcal{N} \tag{3}$$

$$\sum_{j\in\mathcal{N}}Q_{ij}-Q_j^e=0 \quad i,j\in\mathcal{N} \tag{4}$$

$$\sum_{j\in\mathcal{N}}h_f-\sum_{j\in\mathcal{N}}H_{ij}=0 \quad i,j\in\mathcal{N} \tag{5}$$

Here the constraints are a set of differentiable algebraic equations, consisting of an ordinary differentiable equation for chemical transport and algebraic equations for nodal mixing rules and hydraulic behavior. Given the sensor measurements $\hat{c}^*$ at times t, we need to determine the unknown demand terms Q_j^e. The calculated pipe concentrations, $\hat{c}_i(t)$, are at the end position of a pipe. The last term in equation (1) is a regularization parameter to address the non-uniqueness where R can be defined as $(Q_j^e)^2$ or $(\nabla Q_j^e)^2$. The penalty variable, β, is set to an arbitrary low value. $\mathcal{N}$ is the space of nodes or junctions, $\mathcal{N}^*$ the space of nodes with sensors, B is the space of inversion parameters, and $\mathcal{P}$ is the space of pipe segments. The external flow rates Q_{ext} and concentrations $c_{ext}(t)$ are specified for reservoirs. Q_j and u_i are flow rate and velocity in a pipe segment. The variable u_i is bulk flow and is related to the flow rate through pipe diameters. P_k is the set of all pipes flowing into node i. The last two equations (4, 5) represent the dynamics of a network system [10] consisting of continuity and energy conservation, where Q_{ij} represents flow between node i and j. H_{ij} is the head between node i and j and h_f is the head-loss term. The head loss is a nonlinear function of flow rates and friction factors. Although the chemical transport and hydraulic equations are explicitly written as constraints in this formulation a non-intrusive interface is used that eliminates these equations at each optimization cycle [12] [11]. The primary reason for this explicit format is for exposition of the underlying simulation mechanics.

3 Deterministic Optimization Numerical Results

Numerical experiments were conducted by executing a forward simulation, extracting concentration values at sparsely-selected locations and using these sensor values in a nonlinear least squares formulation to solve an inverse problem. Numerical experiments were conducted on two datasets. The first dataset was developed to provide relatively simple flow behavior so that certain aspects of the inversion algorithms could be conveniently tested. Figure 1(a) shows a 10x10 grid consisting of 100 nodes used in the prototype network. A reservoir supplies the necessary flow rate to meet the total demand of the network. Without the variability associated with real data, the flow patterns and the chemical transport can be easily verified by inspection and thereby allowing some qualitative validation of the simulation and inversion algorithms. Complexities were systematically

incorporated to evaluate the quality of the inversion. For instance, increasing the number of demands and decreasing the number of sensor points provided verification that in concept the inversion algorithm could successfully handle the demand and sensor ratio of the tracer dataset.

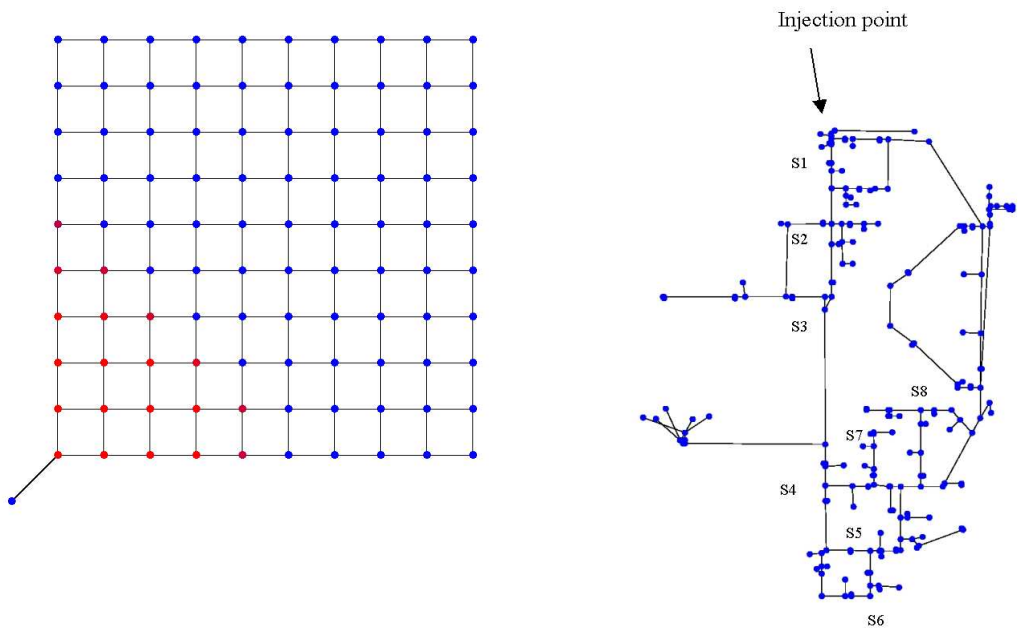

Figure 1: (a) Prototype network 10x10 (b) Tracer study network

The hydraulic behavior and chemical transport in both the prototype and tracer networks were simulated using EPANET [9]. EPANET first solves the hydraulic system for a time sequence given some boundary conditions and a time period, followed by the solution of the chemical transport within the hydraulic time period using a Langrangian approach. The prototype and tracer studies were simulated for 6 and 24 one hour hydraulics time steps respectively, and both datasets were simulated with 3 minute chemical time steps within each hydraulic time step. Parameter and optimization studies were performed using the DAKOTA framework [5], which is a general toolkit capable of uncertainty quantification, sensitivity analysis, design of experiments and other iterative studies for high performance computers.

3.1 Prototype Network Results

As a standard procedure for most optimization problems the design space is investigated by inspecting the general behavior and sensitivity of the specified objective function with respect to the design variables. If the general behavior is differentiable, efficient gradient-based methods can be used to find the optimum. On the other hand, if the response function is non-differentiable or multiple minima are suspected, then non-gradient based methods may be preferred. Figure 2(a)

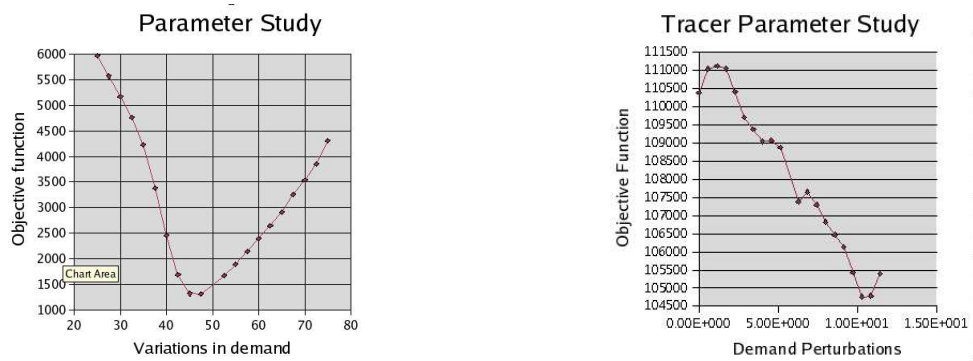

Figure 2: Parameter study for (a) prototype network (b) tracer network

shows a smoothly behaving objective function for the prototype network using a relatively large perturbation step.

Sensitivity analyses were conducted to investigate the performance of the inversion algorithm. One of the important parameters investigated was the finite difference step which is used to calculate the gradient of the objective function. A standard guideline is usually applied by estimating how many reliable digits are available in the objective function $F(x)$ and then only relying on half the digits in the perturbed objective function $F(x + he_j)$, where h and e_j are the step size and the unit vector, respectively [4]. There is an underlying assumption however that the entire simulation operates with sufficient precision in every component and consistently propagates the appropriate precision to the final output values. This is difficult to inspect but easily tested by trying different finite difference steps on the subject dataset. For the prototype network, a finite difference step of approximately 0.01 % resulted in the best inversion solution. This required the precision of EPANET's output values to be increased to a maximum number of digits.

The inversion of demands was initially tested on the prototype network. The goal of this experiment was to determine the minimum number of sensor points for achieving a successful inversion. For a 100 node prototype network simulated for six one hour hydraulic time steps, the inversion resulted in a near perfect match with only 25% of the total sensors. Less than 25% of total sensors resulted in a lower quality solution.

An important issue for this inversion problem is how the quality of the inversion is affected by the character of the demand variables. A simple procedure was implemented through which one randomly placed demand variable was added to the inversion problem and the quality of the inversion was measured. For the 100 node network, the maximum number of demands that could recover the original demands was approximately ten. More than ten parameters required a close starting point including the use of nearly all the observation points to perform a

successfully inversion.

If demands at every node are assigned a random value but within a relatively tight range for the forward problem, then the target demands are not recovered in the inverse problem even if all the nodes are used as observation points. The main reason is that the finite difference based objective function calculation cannot resolve sufficient detail to distinguish between neighboring nodes producing very small concentration differences. Theoretically, this can be corrected if exact sensitivity information exists (machine precision derivatives). Previous work demonstrated the same phenomenon [7] and accordingly it can be concluded that the inversion can be significantly improved from accurate first order information.

Three different gradient based optimization algorithms were applied to the inversion problem. The first algorithm was a Broyden-Fletcher-Goldfarb-Shanno (BFGS) sequential quadratic programming (SQP) method with an augmented-Lagrangian line search. The second algorithm was also a BFGS-based SQP method but with an L1 line search method. Finally, a Gauss Newton methods was evaluated which theoretically should be the most efficient algorithm for a nonlinear least squares problem since no Hessian calculations are required. However, it appears that the Hessian updates are not expensive and the BFGS-SQP methods are faster. The BFGS SQP L1 line search method was the most efficient because no gradient calculations are required (within the line search), whereas the augmented-Lagrangian line search requires a finite difference calculation for each line search iteration. If more precise sensitivity information was available, the augmented Lagrangian method may be more efficient.

3.2 Tracer Test Results

A second network represents a small section of an industrial park and consists of approximately 200 nodes (figure 1-b). A simple tracer test was performed on this network in which CaCl was injected in one location and conductivity was monitored at eight different locations (indicated in the figure by the S labels). Although some demands were accurately measured during the test, eight nodes were unknown. Some historical data provided a guide for the initial guesses for these unknown nodal demands.

Unlike the smooth behavior of the prototype model, the objective function of the tracer test dataset showed non-smooth behavior at small scales that could cause difficulties with the finite difference calculation of the objective function gradient. Figure 2(b) shows the objective function behavior versus perturbations in demand. The possibility of multiple minima can also be detected for the tracer dataset. To address the multiple minima and non-smooth behavior of the network, a global optimizer, specifically a genetic algorithm (GA), was applied for a reduced set of parameters by assuming that the eight demands are constant during the 24 hours simulation period. The GA was able to find a significant improvement from the initial guess, reducing the objective function by approximately 21%.

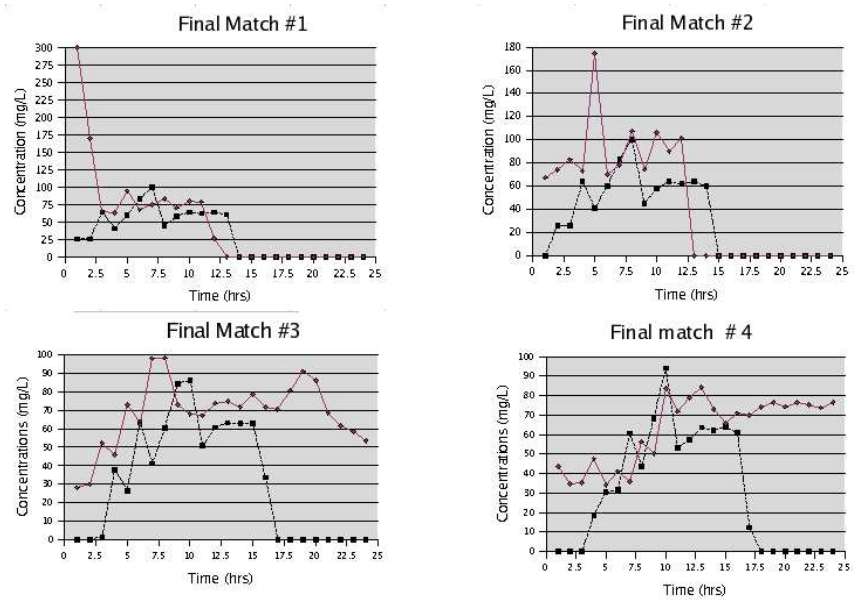

Figure 3: Comparison of concentrations and prediction for sensors no. 1-4; red curves represent observations and black curve represent simulated values

An additional 24 inversion parameters were added to the formulation in an attempt to capture temporal consumption behavior for each demand variable over the 24 hour period. Starting from the previous solution obtained by the global optimizer a gradient based method was able to reduce the objective function an incremental 5% with the temporal-based demand variables. Because of the non-smooth observations of the parameter study, the additional progress toward a better match was not anticipated, but apparently the behavior within the neighborhood of the last solution point was smooth enough for the gradient based algorithm. The final match of the sensor points are shown in figure 3 and 4.

During the initial phases of the study a simple gradient based inversion identified one sensor location as an outlier. The reason for the erroneous result is still inconclusive but is most likely caused by a misrepresentation of field conditions in the numerical model. Some of the error associated with the final match can most likely be attributed to this inaccuracy in the numerical model.

.

4 Optimization Under Uncertainty

Very little work has been done on calibration of water distribution systems under uncertainty. Lansey et al [8] and Xu et al [2] use uncertainty as a constraint in their optimization problem, and Bayan et al [1] discuss the need to incorporate uncertainty associated with demand to perform long term design analysis. Some work has been done on reliability analysis for water distribution networks but are applied to long term hydraulic planning design problems and not model cal-

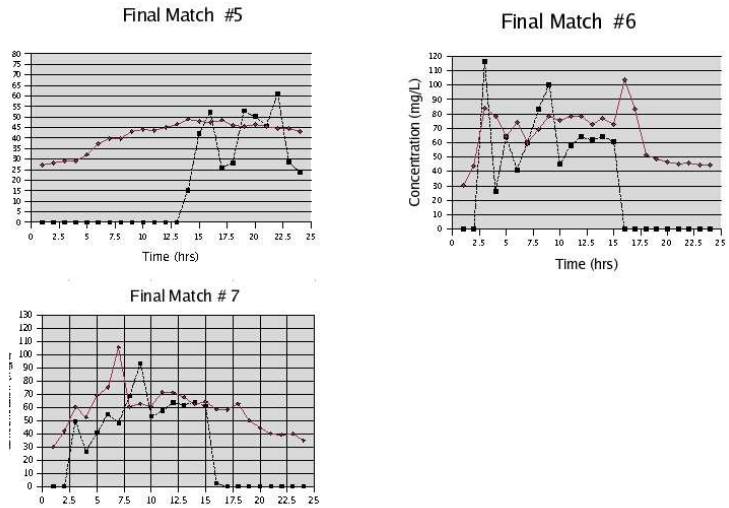

Figure 4: Comparison of concentrations and prediction for sensors no. 5-7

ibration. Xu et al [2] present a comprehensive review of previous efforts in this area and they present a reliability-based optimization approach that uses a first order reliability method (FORM) and is focused on failure to meet demands at a minimum pressure.

To characterize variability in the model and field data, optimization under uncertainty algorithms can be implemented. However, a variety of formulations are possible and are dependent on the problem description. The following strategies can be considered for future application to the model calibration problem:

$$\min_{Q^e} \sum_i (\mu - \bar{c})^2 \quad i \in \mathcal{L} \tag{6}$$

$$\min_{Q^e} \sum_i \beta^2 \tag{7}$$

$$\min_{Q^e} \sum_i \gamma \tag{8}$$

where μ is the mean of a distribution associated with the simulation predictions, $\bar{c}$ are the least squares terms, and $\mathcal{L}$ is the space of nonlinear least squares terms. The mean values for the simulation predictions are generated by the uncertainty quantification process. Equation 7 incorporates more statistic, where β is defined as $\frac{(\mu-\bar{c})^2}{\sigma}$. The standard deviation σ enters in the objective function. The function

γ is some mathematical expression that relates standard deviation and mean of both the simulation and the field measurements. The exact form for γ is the subject of future research. Equation 6 and 7 have been implemented for the model calibration problem. Equation 6 has been executed on the prototype network for a basic validation test. Although it is not anticipated that these formulations will improve the deterministic match, these formulations can help characterize the variability of the model and field measurements.

5 Conclusions and Future Work

Global and local optimization techniques were applied to a model calibration problem which successfully inverted for eight demands by using time histories of concentration measurements. The prototype model used gradient based methods to invert for 10 demands in a 100 node network using as little as 25 sensors whereas a combination of global methods and local gradient based methods were applied to the tracer test dataset to realize a reduction in the objective function of 26%. The calibration process determined that a portion of the network was not being modeled properly (i.e. bad pipe diameters, missing valves, etc).

During the initial evaluation stage of the investigation, a parameter study showed smooth general behavior in the prototype but some small scale non-smoothness in the actual tracer test simulation. Accordingly, a global method was selected to perform the initial inversion on a reduced set of inversion variables. Additional improvement of the inversion was realized by using a local gradient methods starting from the last global method solution and using time dependent inversion parameters. Various parameters were evaluated during the parameter study phase including finite difference step, grid resolution, regularization, initial conditions, and solvers.

Inversion under uncertainty formulations are proposed to help characterize variability in the model and field measurements. The exact formulation and application to the tracer test remains for future work.

References

[1] A Babayan, D. Savic, and G. Walters. Design of water distribution networks under uncertainty demand. In *Pumps, electromechanical DEvices and Systems Applied to Urban Water Mangement.*

[2] Goulter I.C. Chengchao Xu. Reliability-based optimal design of water distribution network. *ASCE Journal of Water Resources Planning and Management*, 1999.

[3] W. de Schaetzen, G.A Walters, and D. Savic. Optimal sampling design for model calibration using shortest path genetic and tropy algorithms. *Urban Water*, 2000.

[4] J.E. Dennis and R.B. Schnabel. Numerical methods for uncontrained optimization and nonlinear equations, SIAM. 1996.

[5] M.S. Eldred, A.A. Giunta, B.G. van Bloemen Waanders, S.F. Wojtkiewicz, W.E. Hart, and M.P. Alleva. Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis. version 3.0 users manual. Technical report SAND2001-3796, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, April 2001.

[6] E. Luvizotto Jr and H. DA Silva Pizzo. Properties and analysis of searching techniqies for calibrating a water distribution network with a model. *ACTA Universitatis Carolinae - Geologica*, 2002.

[7] C.D. Laird, L.T. Biegler, B.G. van Bloemen Waanders, and R.A. Bartlett. Time dependent contamination source determination for water networks. *Submitted to A.S.C.E. Journal of Water Resources Planning and Manag.*, 2004.

[8] Mays L.W. Yeouu-Kung Tung Lansey, K.E. Ning Duan. Distribution system design under uncertainties. *ASCE Journal of Water REsources Planning and Management*, 1989.

[9] L.W. Rossman. EPANET 2. Technical report, EPA, 2000.

[10] E. Todini and S. Pilati. "a gradient method for the analysis of pipe networks. In *International Conference on Computer Applications for Water Supply and Distribution.*

[11] B. G. van Bloemen Waanders, R. A. Bartlett, L. T. Biegler, and C. D. Laird. Nonlinear programming strategies for source detection of municipal water networks. *Presented at EWRI Conference, Philadelphia*, 2003.

[12] B.G van Bloemen Waanders, R.A. Bartlett, K.R. Long, P.T. Boggs, and A.G. Salinger. Large scale non-linear programming for PDE constrained optimization. Technical report, Sandia National Laboroatories, 2002.

[13] G. Walter, D.Savic, M. Morley, W. deScaetzen, and R. Atkinson. Calibration of water distribution network models using genetic algorithms. 1998.