Calibration and Reliability in Groundwater Modelling:

Coping with Uncertainty

Edited by

F. STAUFFER, W. KINZELBACH
Swiss Federal Institute of Technology (ETH), Institute of Hydromechanics and Water Resources Management, ETH Hoenggerberg, CH-8093 Zürich, Switzerland

K. KOVAR
National Institute of Public Health and the Environment (RIVM), PO Box 1, NL-3720 BA Bilthoven, The Netherlands

E. HOEHN
Swiss Federal Institute for Environmental Science and Technology (EAWAG), Department of Chemistry, CH-8600 Dübendorf, Switzerland

Proceedings of the ModelCARE'99 conference held in Zürich, Switzerland, from 20 to 23 September 1999. This conference was jointly organized by the:

Swiss Federal Institute of Technology (ETH), Zürich, Switzerland

International Commission on Groundwater (ICGW), of the International Association of Hydrological Sciences (IAHS)

Committee on Groundwater Hydraulics, of the International Association of Hydraulic Engineering and Research (IAHR)

IAHS Publication no. 265
in the IAHS Series of Proceedings and Reports

Calibration and Reliability in Groundwater Modelling (Proceedings of the ModelCARE 99 Conference held at Zürich, Switzerland, September 1999). IAHS Publ. no. 265, 2000.

A hybrid Marquardt-Simulated Annealing method for solving the groundwater inverse problem

A. ALCOLEA, J. CARRERA & A. MEDINA

Technical University of Catalonia, Jordi Girona 1–3, E-08034 Barcelona, Spain

e-mail: andres.alcolea@upc.es

Abstract One of the limitations to the use of inverse problem algorithms is the relatively high frequency of convergence problems. Continuum optimization methods, such as Marquardt, are quite efficient, but not very robust. That is, they normally converge in a small number of iterations but often fail to converge. On the other hand, discrete optimization methods, such as Simulated Annealing, are very robust, but unacceptably expensive. We present a hybrid method aimed at coupling the efficiency of Marquardt's method and the robustness of Simulated Annealing. We explore its possibilities and adjust some of its parameters by applying it to a large number of cases.

INTRODUCTION

The groundwater inverse problem consists of finding model parameters that are close to direct measurements of those parameters and lead to close agreement between measured and computed dependent variables (normally heads and/or concentrations). Closeness can be defined by the so-called objective function, so that optimal parameters are those that minimize the objective function. Hence, the inverse problem becomes an optimization problem.

Optimization problems can be classified as "continuum" and "combinatorial", depending on whether the optimization variables are continuous or discrete, respectively. Optimization methods have been developed accordingly. A large number of methods are available for continuum problems.

In recent years we have concentrated on Marquardt's method, which tends to be quite fast and efficient, but which may lead to local minima or simply fail to converge in highly nonlinear and ill-posed problems (Carrera *et al.*, 1993). On the other hand, combinatorial optimization methods, and specifically Simulated Annealing, tend to be very robust and are purported to yield global minima, but often require unaffordable large amounts of computation time (van Laarhoven & Aarts, 1987).

The objective of this work is to explore the possibilities of a hybrid method aimed at exploiting the advantages of both Simulated Annealing and Marquardt (and hopefully not the opposite). The paper is organized accordingly. We start by outlining Marquardt's method, Simulated Annealing and the proposed hybrid method. We then present the results of the application of the proposed approach to a set of synthetic problems.

FORMULATION OF THE INVERSE PROBLEM

The first step in any modelling effort consists of identifying the dominant processes and expressing them as partial differential equations that are solved numerically. In what follows we will concentrate on saturated flow solved by the finite element method. We assume the governing equations to be well known and skip them. The point, however, is that a number of hydraulic parameters (hydraulic conductivity, storativity, recharge, etc.) need to be specified throughout the spatial and temporal simulation domain. We will assume that each type of parameter can be expressed in terms of a hopefully small number of unknowns, $\mathbf{p}_i$ ($i = 1$ for hydraulic conductivity, $i = 2$ for storativity, etc.).

Our formulation of the inverse problem consists of assuming that we know some prior estimates, $\mathbf{p}_i^*$, of $\mathbf{p}_i$; that the errors $(\mathbf{p}_i - \mathbf{p}_i^*)$ have a covariance matrix $\sigma_i^2\mathbf{V}_i$ (where $\mathbf{V}_i$ is usually assumed known, but σ_i^2 is a possibly unknown scalar); that data heads $\mathbf{h}^*$ have been measured at a discrete number of points in space and time, and that the residuals $(\mathbf{h} - \mathbf{h}^*)$ ($\mathbf{h}$ being computed heads, hence a function of all $\mathbf{p}_i$) have a covariance matrix $\sigma_h^2\mathbf{V}_h$. With these definitions, Maximum Likelihood Estimation leads to an objective function of the form (Carrera & Neuman, 1986):

$$S = S_h + \sum_i \lambda_i S_i + \beta \tag{1}$$

where S_h, S_i and λ_i are given by:

$$S_h = (\mathbf{h} - \mathbf{h}^*)^t \mathbf{V}_h^{-1} (\mathbf{h} - \mathbf{h}^*) \tag{2}$$

$$S_i = (\mathbf{p}_i - \mathbf{p}_i^*)^t \mathbf{V}_i^{-1} (\mathbf{p}_i - \mathbf{p}_i^*) \tag{3}$$

$$\lambda_i = \frac{\sigma_h^2}{\sigma_i^2} \tag{4}$$

and β is a function of the statistical parameters (σ_h^2, σ_i^2, etc.). In what follows, we will assume β to be constant. Therefore, the solution to the inverse problem simply becomes a minimization problem. Our objective is to minimize equation (1) with respect to $\mathbf{p}_i$ in an efficient manner.

MARQUARDT'S METHOD

This method belongs to the Gauss-Newton family. Basically, it consists of linearizing the dependence of $\mathbf{h}$ on $\mathbf{p}$ and minimizing the resulting quadratic function, while imposing that the parameter change $\Delta\mathbf{p}^k$ at iteration k is constrained. This leads to a linear system of equations (Marquardt, 1963):

$$(\mathbf{H}^k + \mu^k \mathbf{I})\Delta\mathbf{p}^k = -\mathbf{g}^k \tag{5}$$

where $\mathbf{H}^k$ is the first-order approximation to the Hessian matrix at $\mathbf{p}^k$, vector of parameters at iteration k; $\mathbf{I}$ is the identity matrix; $\mathbf{g}^k$ is the gradient vector and μ^k is a positive scalar. While μ^k is the Lagrange multiplier of the constrained optimization, it is updated sequentially according to a set of rules and is usually referred to as Marquardt's parameter.

Actual implementation of the method is quite sensitive to the problem to be minimized. Variations include whether equation (5) is scaled before solving, whether specific constraints are imposed upon parameter changes per iteration, and how μ^k is updated. Details of our implementation are described in Medina (1993). We scale equation (5), and increase μ^k whenever $\Delta\mathbf{p}^k$ becomes nearly orthogonal to $\mathbf{g}^k$, when the objective function is poorly approximated by its quadratic expansion, or when $\Delta\mathbf{p}^k$ fails to yield an improvement in the objective function.

In order to understand the limitations of the method, it is convenient to understand its behaviour. When things work properly, μ^k remains zero (or close to) and the method displays a second order convergence. Increasing μ^k leads to two effects. First, the magnitude of $\Delta\mathbf{p}$ is reduced and, second, its direction becomes closer to that of the gradient vector, which is known to yield poor convergence. It is clear that, in highly nonlinear problems, the search direction computed by this method does not seem to be the most appropriate one. Hence, it is not surprising that Marquardt's method often fails to converge, which motivates this work. Before describing our proposed approach, we need to describe Simulated Annealing.

SIMULATED ANNEALING

The method is based on establishing an analogy between optimization and annealing. The physical process of annealing consists of raising the temperature of a solid so as to increase the vibration of its molecules to a chaotic state, in which the solid loses its structure (maximum reachable energy state). Slowly reducing the temperature should allow the particles to reach their minimum energy state, provided that the initial temperature was sufficiently high and the cooling process sufficiently slow.

Simulated annealing consists of representing this process by assimilating the value of the objective function to the energy state, the set of parameter values to the molecular state of the solid, and perturbations on the parameters are equated to vibrations of the molecules. Temperature can be viewed as a dummy parameter controlling both the probability of accepting a given state and the size of the perturbations on the parameters.

The method consists of the following steps. First, it is initialized by defining an initial temperature and parameter state (configuration). Second, given a parameter configuration, $\mathbf{p}^{k-1}$, a new one is generated randomly by a single perturbation of the previous one. Third, the objective function for the new parameters is computed. If the change ΔS, is negative, then the new configuration is accepted, otherwise, the probability of accepting $\mathbf{p}^k$ is given by the Boltzmann distribution:

$$P\ (\text{accept}\ \mathbf{p}^k) = \frac{1}{RT}\exp\left(-\frac{\Delta S}{RT}\right) \tag{6}$$

where R is Boltzmann's constant and T is temperature.

One of the key features of the method is that it can accept worsening the objective function. This insures that the algorithm will not get stuck in a local minimum and is purported to lead to global minimization, provided that cooling is sufficiently slow and the initial temperature high enough. In fact, implementations of Simulated Annealing differ from each other in the way the parameters are perturbed and on the cooling rules.

While the possibility of avoiding local minima makes the method very attractive, the need to evaluate thousands of configurations (each one requiring the solution of a finite element problem) makes the computational effort of the method unacceptably high. It must be stressed that this problem is shared by all combinatorial optimization methods (e.g. genetic algorithms), so it should also be acknowledged that these methods are better suited for optimizing discrete variables, rather than continuous ones.

THE HYBRID METHOD

The method we propose explores the possibility of coupling the robustness of Simulated Annealing with the efficiency of Marquardt's method. In essence, it consists of extending Simulated Annealing so as to test the value of the objective function not at each parameter configuration, but at its minimum within a neighbourhood. This is achieved by modifying each configuration by Marquardt's method. Therefore, after initialization, the algorithm flows as follows:

1. Generate a parameter configuration, $\mathbf{p}_{SA}{}^{k}$, through a random perturbation of $\mathbf{p}^{k-1}$.
2. Simulate the flow problem to find the objective function S ($\mathbf{p}^{k}_{SA}$).
3. Perform a Marquardt's iteration to obtain a new configuration $\mathbf{p}^{k}_{M}$.
4. Simulate again the flow problem to compute the new objective function S ($\mathbf{p}^{k}_{M}$).
5. Compare S ($\mathbf{p}^{k}_{M}$) and S ($\mathbf{p}^{k}_{SA}$) and choose $\mathbf{p}^{k}$ as the one leading to smaller S.
6. If ΔS is negative (a better configuration has been found) then accept $\mathbf{p}^{k}$, else, use equation (6) to decide whether the new configuration is accepted.
7. Repeat steps 1 to 6 until convergence.

A number of issues require testing. Specifically, the resulting algorithm should be faster than conventional Simulated Annealing (otherwise the algorithm would be overly expensive). Also, parameter perturbations may take advantage of the knowledge of the objective function derived from Marquardt's iteration. We have implemented this algorithm in TRANSIN III (Galarza *et al.*, 1996), a code that solves the inverse problem using Marquardt's method.

We have implemented two schemes for perturbing the parameters. In the first, parameters are randomly perturbed according to the approximation of the Hessian matrix derived from Marquardt's method. In the second one, they are perturbed according to the prior variances of the parameter estimates ($\mathbf{p}^{*}$). We will denote this method as ellipsoids. Cooling is performed geometrically (that is, temperature is reduced by a constant factor after a fixed number of successful iterations).

APPLICATION TO SYNTHETIC EXAMPLES

The code was applied to 10 synthetic examples. In all of them, the conceptual model was known *a priori* and was used to derive "true" heads. Measurements were derived from these by adding a white noise (Gaussian with zero mean and known variance). The same was done with the parameters to generate prior estimates. The conceptual model was also perturbed so that, in most cases, the "true" model could not be reproduced (exactly) by the calibrated one, in order to simulate conceptual errors. In

some of the examples, the zones of the original model were melted. In these cases, prior estimates were taken as the geometric average of the corresponding parameters in the true model.

The base model consists of a 25 × 16 m^2 domain. The upper and lower boundaries are taken as impervious, while flow is prescribed as 1 $m^3 day^{-1}$ at the left-hand side and the prescribed head equals 3 m at the right-hand side. A concentrated pumping rate of 5.5 $m^3 day^{-1}$ is imposed at a point inside the domain and a mixed (river) boundary condition is set at a line of nodes. The setup is outlined in Fig. 1. A varying recharge is imposed throughout the domain. Transmissivity is constant in each of the five zones in which the domain has been subdivided. One of them is taken as anisotropic. Storativity is constant.

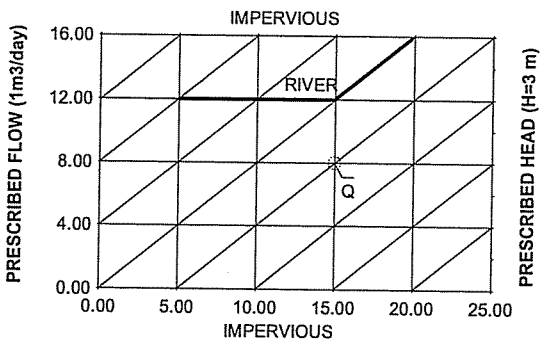

Fig. 1 Flow domain and boundary conditions.

DISCUSSION OF RESULTS

One of the main problems with the proposed method is the large number of optimization parameters that have to be specified. Hence, our objective was two fold. On one hand, we aimed at obtaining standard values for these parameters. On the other, we wanted to select an appropriate scheme for perturbation. In all, over 512 models were run. Only the main results can be summarized here. Regarding initial temperature and search radius, convergence is not very sensitive to the initial choice, provided that parameters are perturbed according to the ellipsoids method. However, CPU time (number of iterations) is sensitive. The evolution of the objective function versus the number of iterations for one example in which Marquardt's method failed to converge is shown in Fig. 2 for three initial temperatures.

The behaviour of the cooling rate is more complex and it is linked to the initial temperature. What we found reliable and relatively fast was to start from a high temperature, and decrease it after every successful iteration. A coefficient of reduction of 0.9 behaved correctly in most cases.

Perturbing the parameters according to the Hessian matrix was not appropriate because it displayed a very similar behaviour to that of Marquardt's method. When the latter converged, it was faster than the hybrid method. Therefore, little was gained by using a method that behaved similarly to Marquardt's.

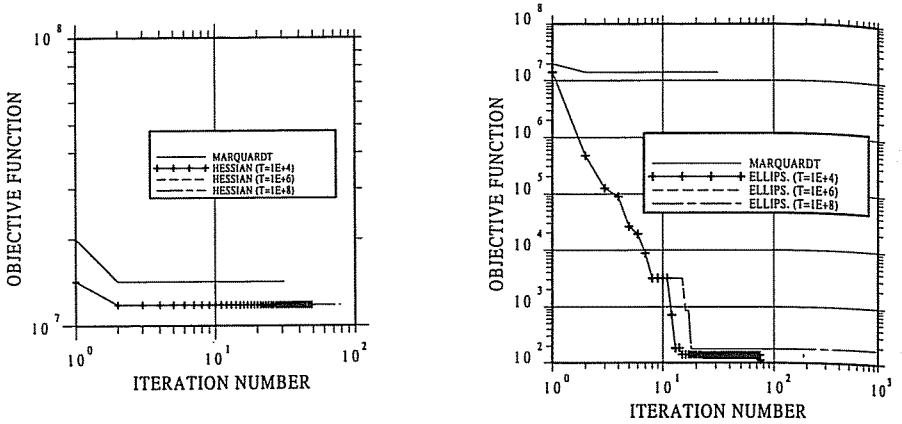

Fig. 2 Comparison between Marquardt's and the hybrid method using (left) Hessian matrix and (right) ellipsoids perturbation schemes, for three initial temperatures.

As shown in Fig. 3, the method is not very sensitive to the limits on the size of parameter change per iteration.

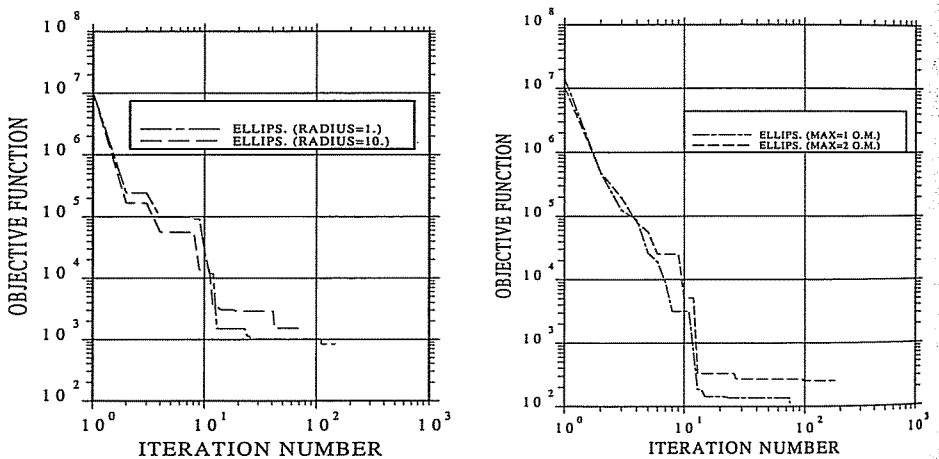

Fig. 3 Sensitivity of the hybrid method to the initial search radius, using the ellipsoids perturbation scheme (left), and to the limits on the size of parameter change per iteration (OM denotes order of magnitude).

The summary of the exercise is that the proposed approach is much more robust than Marquardt's method. However, when the latter converges, it is much faster. Therefore our results suggest that a possible strategy is to first try Marquardt and, if it fails, try the proposed approach. However, a number of issues remain open, including the possible parallelization of the code, different cooling schemes and perturbation methods, testing on a wider spectrum of cases, etc.

Acknowledgement This work has been performed with funding from ENRESA (Spanish Nuclear Waste Disposal Company)

REFERENCES

Carrera, J. & Neuman, S. P. (1986) Estimation of aquifer parameters under transient and steady state conditions. I. Maximum Likelihood Method incorporating prior information. *Wat. Resour. Res.* 22(2), 199–210.

Carrera, J. , Medina, A. & Galarza, G. (1993) Groundwater inverse problem. Discussion on geostatistical formulations and validation. *Hydrogéologie* 4, 313–324.

Galarza, G., Medina, A. & Carrera, J. (1996) TRANSIN III: Fortran code for solving the coupled non-linear flow and transport inverse problem. *E.T.S.E.C.C.P.B., UPC. El Berrocal Project, Topical Report 17, ENRESA.*

Marquardt, D. W. (1963) An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Industr. Appl. Math.* 11(2), 431–441.

Medina, A. (1993) Estimación conjunta de los parámetros de las ecuaciones de flujo y transporte (Coupled estimation of the parameters of flow and transport equations). PhD Thesis, School of Civil Engineering, Technical University of Catalonia, Spain.

Van Laarhoven, P. J. M. & Aarts, E. H. L. (1987) *Simulated Annealing: Theory and Applications.* Kluwer Academic Publishers, Dordrecht, The Netherlands.