

SCIENTIFIC NOTEBOOK # 664E
A Full-Bayesian Approach For Groundwater Model
Parameter Estimation Using MODFLOW

By:
Allan D. Woodbury
C. Lanczos & Associates Ltd.
Winnipeg, Manitoba, Canada

Prepared For
Southwest Research Institute
Center For Nuclear Waste Regulatory Analyses
San Antonio, Texas

Date of Issue: June 7, 2004
Valid Dates: June 7, 2004 to September 27, 2004

Contents

1 Initial Entries	1
2 Objectives and Summary	2
2.1 Computers, Computer Codes, and Data Files	2
3 Introduction	2
3.1 Objectives	3
4 Literature Review	4
4.1 Functional Relationships and Governing Equations	5
4.2 Construction	6
5 Optimization Techniques	7
5.1 Constrained Simplex	9
5.2 Controlled Random Search	11
6 Bayesian Methods and Monte-Carlo Methods	12
6.1 Metropolis Algorithm	15
6.2 Analysis	16
7 Stopping Criteria	16
7.1 Computation of Covariances	17
8 Verification	18
9 Summary to Date	20
10 REFERENCES	20
10.1 Additional Readings and Bibliography	22

1 Initial Entries

Scientific Notebook: #664E

Issued to: Allan D. Woodbury

Issue date: June 7, 2004

Printing Period: September 30, 2004 (final printout)

Project Title: A Full-Bayesian Approach For Groundwater Model Parameter Estimation Using MODFLOW

Project Staff: James Winterle and (CNWRA).

By agreement with the CNWRA QA program, this notebook is to be printed at approximately quarterly intervals. This computerized Scientific Notebook is intended to address the criteria of CNWRA QAP-001. [Allan D. Woodbury, September 30, 2004]

2 Objectives and Summary

Within the Amargosa Desert and Fortymile Wash regions adjacent to Yucca Mountain, Nevada, vast areas exist along the projected radionuclide flow path for which little hydrogeologic and geologic data are available. As a result, groundwater flow and mass transport models are poorly constrained within this region.

This study will evaluate the feasibility of implementing a full-Bayesian approach for groundwater model parameter estimation using the MODFLOW groundwater modeling software. The initial phases of this work will include literature reviews to evaluate different options for implementing this approach, and assistance with the initial development of routines for executing the MODFLOW code and updating input files.

It is the hypothesis of this report that the full-Bayesian techniques mentioned can be successfully adapted to the inverse problem of groundwater flow and mass transport. In the following sections the basic concepts of fluid transfer in porous media are reviewed, followed by an adaptation of a "full-Bayesian" approach for nonlinear inversion of hydraulic head data.

A program was developed to illustrate the Bayesian solution to the regression of x against y for a linear relationship. The correspondence to least squares is near perfect and shows verification of the technique.

This notebook documents aspects of the work performed by C. Lanczos & Associates Limited (Dr. A. Woodbury) and CNWRA staff on this project.

2.1 Computers, Computer Codes, and Data Files

The computer codes used in this study are based on a suite of FORTRAN 77 and FORTRAN 90 codes developed or acquired by Dr. Allan Woodbury. The data analyses were carried out using computer versions of Windows 2000 operating system. Processed data files, FORTRAN code and output will be included on a floppy disk with the hard copy of this report.

3 Introduction

Difficulties associated with direct measurement of the hydrologic parameters needed for physically-based mathematical models are well known. Equally well known are the difficulties in the calibration procedure when trying to adjust parameters within preconceived limits until model output at selected points matches observed values. Quite often questions are raised as to the uniqueness and optimality of these models. A major focus of research over the last decade has been directed towards inversion techniques and parameter estimation as a way of both automatic calibration and as a statistical procedure to quantify the reliability of parameter estimates (see reviews by Ginn and Cushman, 1990; and McLaughlin and Townley, 1996, 1997; Kitanidis, 1997). The understanding of the problem has improved, and while it is generally considered as yet unsolved (in the sense that no panacea has yet been developed) there are clear ideas of what the weak points are and what might be the remedies. In the inverse approach measurements of hydraulic head, hydraulic conductivity (or transmissivity), seepage flux, and the like are inputs to an inverse algorithm, and fitted hydraulic conductivity (or other parameters) become the output, along with the parameter covariance structure. Mathematically, the model takes the form of a vector \mathbf{m} which consists of all unknown parameters (for instance, hydraulic conductivities, boundary conditions, etc.).

Traditionally, inverse techniques in hydrogeology rely on measurements of hydraulic conductivity and hydraulic heads, and they employ the groundwater flow equation for interpretation. Relatively few works have gone beyond this approach and introduced additional information such as tracer data (cf., Carrera et al., 1993), or geophysical measurements (Woodbury and Smith, 1988; Rubin et al., 1992; Hyndman et al., 1994; Copty and Rubin, 1995; Hubbard et al., 1997). The quest for diversifying the types of information stems from the recognition that sophistication of inverse algorithms cannot replace information and data. Along these lines, Abriola et al (1992) noted that “for most specialties, it was generally felt that the state-of-the-art has surpassed the ability to utilize the results in a practical scenario” and that “..applications of [mathematical models] is hindered by the lack of data required to implement or verify them”. This recognition is well demonstrated in Carrera and Neuman (1986) where it is shown that the instability and non-uniqueness of solutions to the inverse problems can only be eliminated by introducing additional measurements and information. The challenge of course, is to find inexpensive and reliable sources of information.

3.1 Objectives

The purpose of this work is to:

1. Review the pertinent literature related to linear and non-inverse methods,
2. Identify gaps in knowledge,
3. Identify those methods that may work best in the nuclear regulatory framework of Yucca Mountain,
4. Set up a series of small test problems in which simple regression methods can be used to verify results,
5. Set up a realistic two-dimensional problem with roughly the same number of parameters and scale as the Yucca site. Of course, this would be a generic site and therefore would satisfy the “control” we would need to verify techniques,
6. Scale up to a fully three dimensional model of Yucca Mountain flow system and transport of radionuclides to a compliance boundary.
7. Apply the inverse approach to the field site.

Note that the following features of any inverse method are important:

1. The objectives and methodology may change depending on the nature of the parameter estimation problem, or the inversion.
2. The technique should be flexible enough to be able to wrap-around code or codes that have been developed to solve the forward problem.
3. Be able to incorporate soft geologic data and other types of constraints.
4. Handle potentially large numbers of unknowns.

5. Be adaptable so that the technique can aid in the selection of alternative models.
6. Should be publically defensible, understandable, and reliable.

To do these things as mentioned above, a series of different objective functions need to be minimized under a set of parameter constraints. Several powerful optimization techniques are detailed that can be utilized to minimize a complicated nonlinear functional without derivative computations. As well, Markov-chain Monte Carlo methods will also be investigated. These inverse techniques allows us to solve problems with more than one system of equations, or problems with uncertain boundary conditions, such as a free-surface. The methods are similar to the Random Search Method (Price, 1977; Silva and Hohmann, 1983), and the Complex Technique (Box, 1965). The algorithms produce a simplex figure of an objective function in parameter space by randomly distributing a set of search points within constraint boundaries. For each search point, a value of the objective function is calculated. These search points are adjusted within the constraints in a series of iterations until the expected value of the objective function reaches a threshold value which is determined by the level of noise in the data. A statistical analysis involving all the sampled points will be carried out when the parameter estimates are found. With the techniques described, the parameter space is randomly sampled to produce a covariance matrix; the only assumption being that the random sample be representative. The mean of all samples is taken as the best parameter estimate. In addition, the covariance structure of the data is also provided, thus avoiding the effort of further conditional simulations.

4 Literature Review

(These references are by no means exhaustive and serve only to indicate the importance of the overall problem of site characterization). In the inverse approach measurements of hydraulic head, hydraulic conductivity or transmissivity, seepage flux, and the like are inputs to the inverse algorithm, and fitted hydraulic conductivity (or other parameters) become the output, along with the parameter covariance structure. Some authors (for instance, Clifton and Neuman, 1982) use hydraulic conductivity values (with their associated covariances) determined from the inverse in conditional simulations. These authors showed that predicted hydraulic-head variances from the inverse/conditional simulations are greatly reduced over estimates of the variance obtained from kriging alone.

As mentioned, mathematically, the model takes the form of a vector \mathbf{m} which consists of all unknown parameters (for example, hydraulic conductivities, boundary conditions) in a functional relationship that predicts physical data \mathbf{d} (for example, hydraulic heads). By a functional relationship, we mean some appropriate form of the groundwater flow equation.

Currently there are two broad classes of inverse methods documented in the hydrogeologic literature: a parametric approach, in which a continuous model is replaced by a finite number of parameters (e.g., Neuman and Yakowitz, 1979; Neuman et al., 1980; Neuman, 1980; Clifton and Neuman, 1982; Carrera and Neuman, 1986a; Cooley, 1977, 1979, 1982, 1983; Yeh et al., 1983), and the cokriging-geostatistical approaches (e.g., Kitanidis and Vomvoris, 1983; Hoeksema and Kitanidis, 1985; Dagan, 1985). Hybrids of both methods have also been developed (de Marsily et al., 1984). Fundamental to both of these broad cases of inverse techniques is the adoption of a model with a relatively small number of unknown parameters in relation to the number of available data points.

The parameters are then determined through some form of maximum likelihood framework and/or a weighted least squares minimization.

In the cokriging approach, a simple functional form for both the drift and the covariance of the parameters is adopted, thus yielding a small number of unknown parameters to be determined. This approach requires a data base in hydraulic head and hydraulic conductivities or transmissivities sufficient to establish covariograms and variograms. However, in many geoenvironmental problems (for example, stability of landslides, design of open-pit mines) it is usual that few hydraulic conductivity data are available and, as well, most of these problems are conceptualized on cross sections. Boundary conditions may also be uncertain. These facts make the application of geostatistical (cokriging) approaches difficult for these problems. Therefore, a geostatistically based method such as cokriging will not be pursued in this study.

In the parametric approach it is typical to determine the parameters in a relatively small number of zones to ensure a mathematically unique solution is found, whether or not there is any physical justification for the parameterization (for example see Carrera and Neuman, 1985c, p 233-236 Birtles and Morel, 1979). In some applications (for example, cross-sectional problems) the modeler may not possess sufficient geologic information to establish physically-based arguments for zoning sections of a finite difference or finite element grid. There are a number of possible methods for resolving this problem in particular cases. One approach is to allow the inverse method to identify zones of similar parameter structure.

In the solution of boundary value problems, one can face uncertainties in assigning a spatial representation for the hydraulic conductivity field, \mathbf{K} , identifying sources/sinks, major structural features (i.e. faults) and boundary conditions. This situation requires that an inverse scheme should be robust and capable of incorporating these types of conditions as unknowns.

4.1 Functional Relationships and Governing Equations

A solution of the inverse problem requires that both the forward and inverse problems must be clearly defined. A forward problem is set up by constructing a functional relationship to predict physical data, given a set of input parameters to a physical model. The goal of inverse theory is to use a finite set of inaccurate observations to extract information about the model \mathbf{m} (adapted from Oldenburg, 1984). For many physical problems the data and the model are related through a linear functional. However, in groundwater hydrology, when a parametric approach and a numerical scheme are used to solve the governing equations, a non-linear functional results:

$$h = \mathfrak{S}[\mathbf{x}, \mathbf{m}] \quad (4.1)$$

where \mathfrak{S} is a non-linear functional relating h , the data (the actual values of hydraulic head), \mathbf{x} , a vector of grid coordinates in a numerical scheme and \mathbf{m} , the actual model parameters, which could consist of hydraulic conductivities, boundary fluxes, sources and sinks.

In applications of the inverse involving observed data, hydraulic heads have uncertainties associated with their values, resulting from interpolation or measurement errors. In these cases (4.1) takes the form:

$$h^* = \mathfrak{S}[\mathbf{x}, \mathbf{m}] + \nu \quad (4.2)$$

where h^* is the data (interpolated or measured values of hydraulic head), and ν is a vector of residuals. Usually, ν is assumed to have the following properties (a mean of zero and some covariance

structure):

$$\begin{aligned} E(\nu) &= 0 \\ E(\nu\nu^T) &= \sigma_h^2 V_h \end{aligned} \quad (4.3)$$

where σ_h^2 is a scaling factor used in defining the magnitude of the covariances. If kriging is used to determine V_h , then σ_h^2 is taken as equal to one.

In this study the functional $\mathfrak{S}[\mathbf{x}, \mathbf{m}]$ is given by the two-dimensional form of the steady-state groundwater flow equation:

$$\nabla \cdot \mathbf{K} \cdot \nabla \Psi = q \quad (4.4)$$

subject to

$$-\mathbf{K} \cdot \nabla \Psi \cdot \mathbf{n} = q_f \quad (4.5a)$$

on Γ_1 and

$$\Psi = g(x, y) \quad (4.5b)$$

on Γ_2 . Here $\nabla = (\partial/\partial x, \partial/\partial y)$, $\nabla \cdot = (i\partial/\partial x + j\partial/\partial y)$, Ψ is the hydraulic head, q represents sources and sinks, \mathbf{K} is the hydraulic conductivity tensor, q_f is a specified flux term on boundary Γ_1 , and $g(x, y)$ is a function specifying Dirichet boundary conditions on Γ_2 . When a numerical scheme is used to solve (4.4), the functional (4.1) takes the form:

$$\mathbf{h} = \mathbf{A}^{-1} \mathbf{f} \quad (4.6)$$

where \mathbf{h} is the approximate value of Ψ due to the discretization, \mathbf{A} is a global stiffness matrix which is a non-linear function of \mathbf{K} , and \mathbf{f} represents a loading vector of the appropriate boundary conditions.

The inverse can be posed as an optimization problem. For example J below is a functional to be minimized, with respect to other constraints. A generalized L_2 norm is introduced as (after Neuman and Yakowitz, 1979):

$$J = (\mathbf{h} - \mathbf{h}^*)^T \mathbf{V}_h^{-1} (\mathbf{h} - \mathbf{h}^*) + \lambda (\mathbf{m} - \mathbf{m}^*)^T \mathbf{V}_m^{-1} (\mathbf{m} - \mathbf{m}^*) \quad (4.7)$$

where the above terms are defined as (Neuman and Yakowitz, 1979): V_h and V_m are head and model covariance matrices, \mathbf{m} is a vector of log conductivities determined by the inverse method, \mathbf{m}^* is a vector of observed or estimated parameters, \mathbf{h}^* is a vector of observed or estimated hydraulic heads, and λ is a scaling factor, which may be unknown. V_h and V_m are defined based on the characteristics of the data set and the finite element mesh. If \mathbf{h}^* and \mathbf{m}^* are estimated by kriging, along with V_h and V_m , then $\lambda = 1$. With an unknown λ we recognize that we may have knowledge about the structure of the covariance matrices but not their magnitudes. Equation (4.7) can be derived from a maximum likelihood consideration for a Gaussian distribution (Carrera and Neuman, 1986a), but one does not have to assume any underlying statistical distributions of \mathbf{h} or \mathbf{m} to apply the norm (Schweppe, 1973). The objective function can also be viewed as a weighted sum of L_2 prediction error (heads) and L_2 solution simplicity.

4.2 Construction

Basic to most of the work in groundwater hydrology is the determination of a model which fits the data. The mathematical foundation for model identification has been developed in a number

of disciplines: control theory in electrical engineering (e.g. Schweppe, 1973), history matching in petroleum reservoir engineering (e.g. Gavals et al., 1976), aquifer simulations in groundwater hydrology (e.g. Neuman and Yakowitz, 1977), and in geophysics (e.g. Parker, 1977 Oldenburg, 1984).

A fundamental problem in determining a model which fits the data is that the ‘true’ model is a continuous function of the data. Unfortunately, there are an infinite number of functions that can reproduce a finite number of observations. In the present context, this could be called *geologic non-uniqueness*, in that a decision must be made, based on the available geologic information, as to how to represent or parameterize the model. For instance, a decision must be made on the number of zones (layers) of similar hydraulic conductivity that exist in the flow domain.

Work in this area has been carried out by Carrera and Neuman (1986a) using a maximum likelihood criterion to aid in optimal selection of zones to which single values of the parameters are assigned. Generally, there is no unique method of resolving this type of non-uniqueness, unless strong geologic evidence exists to aid in the parameterization. As Oldenburg (1984) states: “ This problem becomes more acute when data are inaccurate and sparse, and models which do not fit the data precisely are considered acceptable ”. Woodbury et al. (1987) suggested that a reasonable approach to the inverse problem may be to construct a variety of acceptable models which fit the data under different norms. These norms are based on the L_1 and L_2 lengths. This approach may provide insight into the non-uniqueness of the parameterization, and models which are geologically parsimonious, or the ‘simplest’ in some sense, can be selected as best approximations for the model.

The underlying philosophy proposed in this work is to introduce a series of objective functions (norms) which can be used for model construction. The norms are based on Bayesian and Maximum Entropy viewpoints of probability.

In setting up an inverse problem, an interpretational model is formulated for which parameter estimates are sought. If insufficient geologic information exists to distinguish zones, a large number of parameters is incorporated in the model and the simplest geologic model can be chosen as the best estimate of the ‘true’ model. Simplicity can be defined as either the homogeneous-isotropic model, the smallest, the shortest deviatoric length from some estimate, the flattest, or the smoothest. The smallest model has the shortest Euclidean length of vector \mathbf{m} . The SVD method (singular value decomposition) can be used to find smoothest and flattest models can be found by minimizing the first and second spatial derivatives, respectively between physically adjacent model parameters (see Menke, 1984). Assuming that the data are free of gross errors, a difference in model parameters estimated using these different norms suggests a possible inconsistency in the model structure. Conversely, differences in constructed models may also point to gross errors in the data if the parameterization is correct. Silva and Hohmann (1983) used such an approach in treating a non-linear magnetic inversion problem in geophysics. The idea of using different norms for construction purposes is well established (see Oldenburg, 1984 for a review).

5 Optimization Techniques

Equation (4.2) represents a constrained non-linear optimization problem and a large number of optimization techniques can be used to solve it. Some algorithms use quadratic programming techniques to solve a linearized form of (4.2). These programming techniques involve the solution of a set of simultaneous linear equations, and are sensitive to rank deficiency. Rank deficiency

occurs when one or more of the unknown parameters are linearly dependent. This situation can occur if a large number of unknown conductivities are estimated in an overparameterized model with inadequate prior information. If rank problems are encountered in a conventional quadratic programming scheme, the matrix solution routines fail to yield a solution. Cooley and Naff (1986) describe instances of this type of problem.

Other optimization techniques involve the minimization of an objective function in its original, non-linear form. Some methods require that derivatives of the objective function with respect to model parameters be calculated, others do not. The derivative or gradient based methods as they are called are iterative in nature. These methods are usually faster than non-gradient methods but may diverge (fail to find an answer), or converge to a local rather than a global minimum (Reklaitis, et al., 1983). A non-linear inverse problem may have a complex functional surface, and there is no guarantee that any technique will converge to a global minimum. Gradient methods work best if the initial guess is linearly close to the solution. To investigate non-uniqueness it would be desirable to repeat the procedure many times with different initial guesses. Unfortunately, with gradient based methods it is sometimes difficult to arrive with a new initial guess for the parameters, especially if the problem involves constraints. Carrera and Neuman (1986b) evaluate several gradient based methods and show examples of non-convergence using examples in groundwater hydrology.

Model covariances are useful end-products of an inverse scheme. However, because the distribution for m^{est} may be non-Gaussian, the covariance of the estimated model parameters may be difficult to interpret, especially in terms of confidence intervals. The covariance of the parameter values can be estimated as (Menke, 1984):

$$[cov\ m^{est}] \cong G_n^{-g}[cov\ h^*]G_n^{-gT} + [\mathbf{I} - R_n][cov\ m^*][\mathbf{I} - R_n]^T \quad (5.1)$$

where $R_n = G_n^{-g}G_n$, \mathbf{I} is the identity matrix, and the final n 'th iteration is used. Also:

$$[G_n]_{ij} = \frac{\partial \mathfrak{S}(x, m)_i}{\partial m_j} \quad (5.2)$$

G^{-g} is called the *generalized* inverse of G . The reason for the non-Gaussian nature of the J surface has been examined at a theoretical level by Tarantola and Valette (1982). Because (5.1) is non-linear the probability distribution function of J will *in general* be non-Gaussian, although fortunately the experience in groundwater hydrology with aquifer inverses is of Gaussian or near-Gaussian behavior (e.g.. Cooley, 1977). Nevertheless, the maximum likelihood point of a non-Gaussian objective function may not yield the most sensible parameter estimates. Gaussian distributions are symmetric, so the maximum likelihood point always coincides with the mean value. For an arbitrary non-linear surface (for example, multi-modal, skewed) the maximum likelihood point can be quite far from the mean value. Figure 2 shows schematic diagrams of objective functions for several non-linear problems. Figure 2 (after Mosegaard and Tarantola, 2002) shows an objective function that has several local minima, or points of non-convexity with a well defined global minimum.

Plots (a) and (b) of this figure show linear or linearizable problems with one solution, (c) shows a finite range of solutions. Plot (d) shows a problem with (possibly) an infinite range of solutions. It is the goal of inverse theory to condition the norm in such away that will yield features like (a), with a well defined solution. The techniques proposed in this review sample the objective function surface close to the minimum to find an average value of the model parameters. These values may be more appropriate than a single estimate obtained by most algorithms.

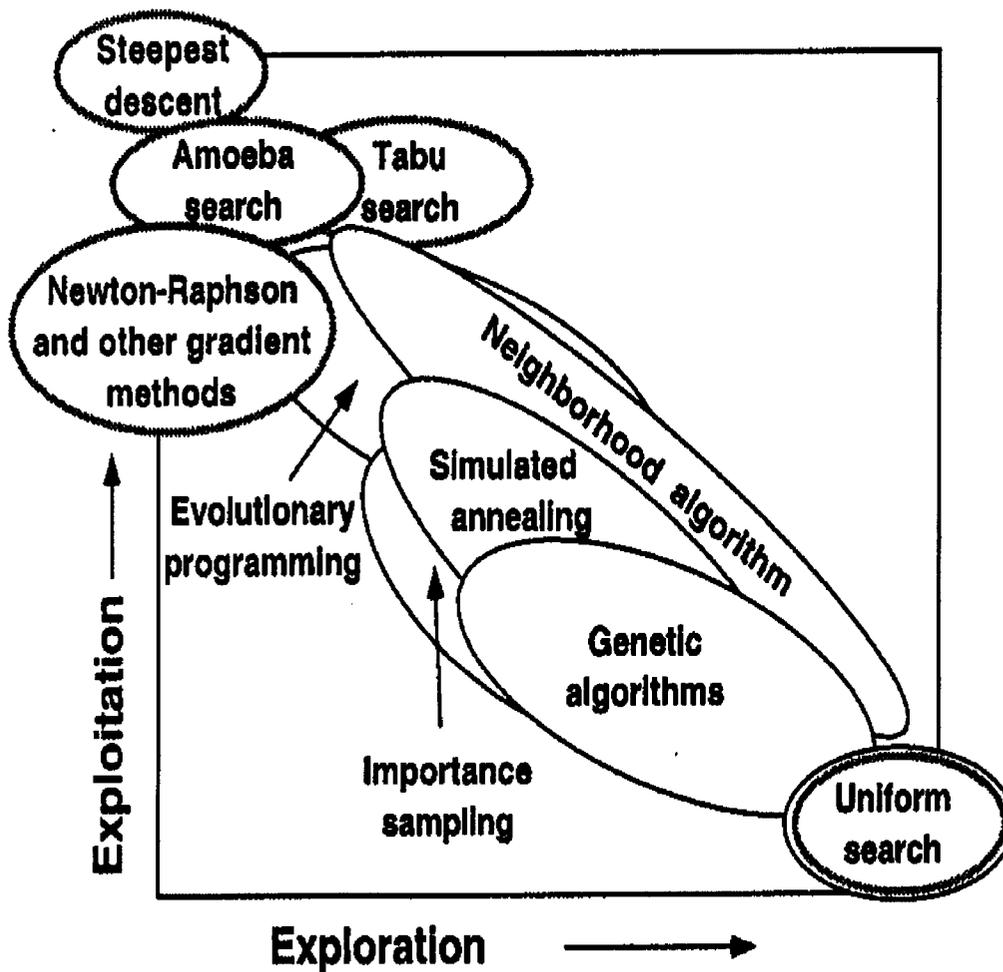


Figure 1: Overview of solution methods, after Sambridge and Mosegaard, 2002)

The inverse procedure produces estimates of the data as:

$$\mathbf{h}^{est} = \mathfrak{S}(\mathbf{x}, \mathbf{m}^{est}) \quad (5.3)$$

where h^{est} is the estimated hydraulic head data produced from the inverse-parameter estimates. Ordinarily the correlation structure of h^{est} cannot be obtained from the inverse approach even though the prior covariance structure of the data may be known. In order to generate this matrix ($cov h^{est}$) one has to use the Monte Carlo approach (Clifton and Neuman, 1982), or use a perturbation technique (Townley and Wilson, 1983) after obtaining the estimated model covariances, ($cov m^{est}$).

5.1 Constrained Simplex

In order to deal with solution divergence and false convergence, and to avoid having to compute gradients, inflexibility of starting guesses, and linear approximations in covariance calculations, we

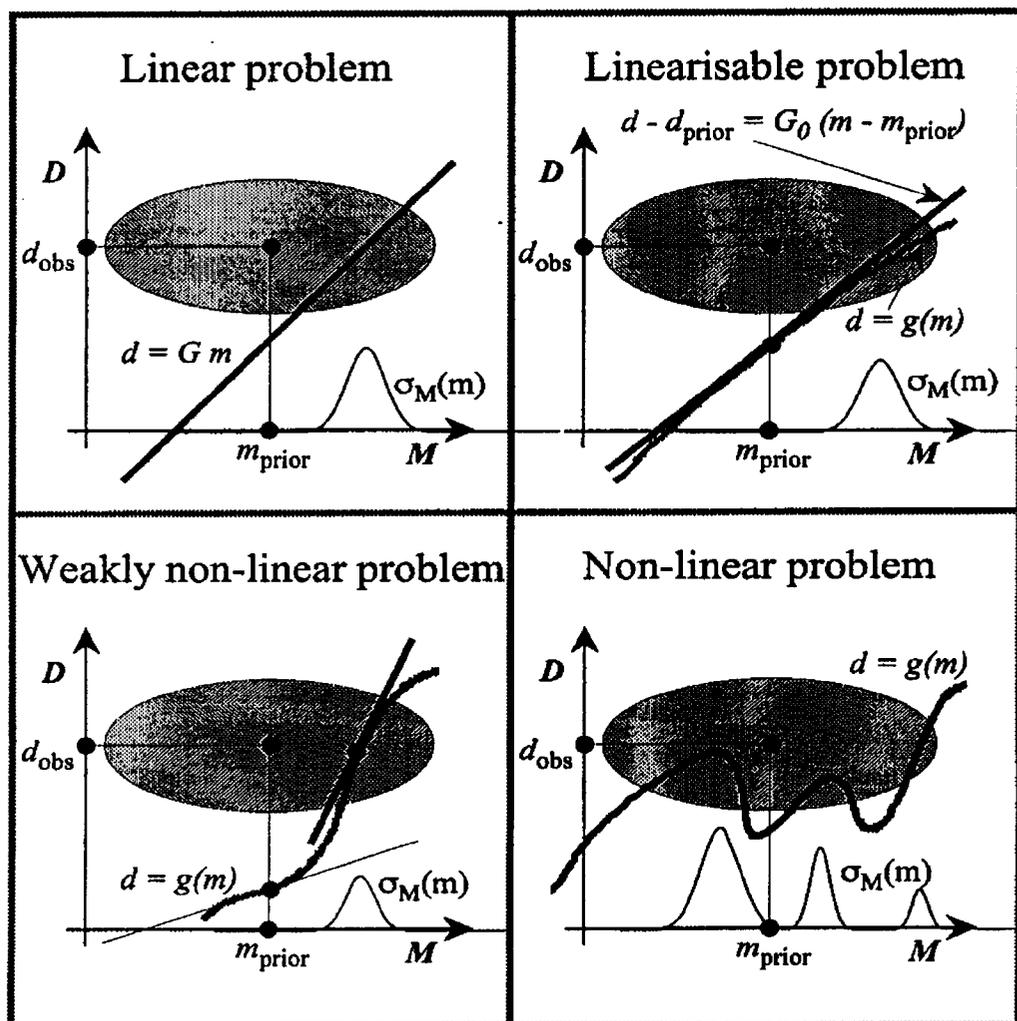


Figure 2: Illustration of four domains of linearity

can adopt alternate methods of solution. Silva and Hohmann(1983) used a controlled random search technique to optimize a non-linear functional in magnetic anomaly interpretation.

A constrained simplex approach can be used (Complex technique: Box,1965) for global optimization. The algorithm easily handles constraints, does not require partial derivatives of the functional with respect to the parameters, and can be employed to minimize complicated non-linear functionals which may contain more than one system of equations. In addition, convergence to a global minimum is enhanced (even in the presence of local minima) provided the global minimum exists within the parameter constraints. Besides finding the global minimum, the algorithm produces a sampling of the parameter space around the global minimum. This sampling produces the parameter and estimated data covariances matrices.

The constrained simplex technique is well known in the chemical engineering and optimization literature (Reklaitis et al., 1983). The algorithm consists of choosing L search points ($L \geq N + 1$ parameters) randomly within the constraint boundaries. This initial task can be accomplished with

a random number generator. These points set up the initial simplex figure. Each search ‘point’ is a vector of model parameters. Thus, L points form an $N \times L$ matrix P^* . The objective function is then evaluated at each search point and the search points with the minimum and maximum objective function values are identified. The parameter estimates and function value for each point are stored in computer memory. At the start of each iteration, the point with the maximum function value is reflected and expanded about the centroid of the remaining points according to the relation:

$$B = S + \alpha(S - Q) \quad (5.4)$$

where B is a vector of new parameter values, S is a vector of parameter values with each entry calculated as the average of the parameter estimates using the other $(L - 1)$ search points, Q is the previous position vector of the maximum J value, and α is an expansion factor. Box (1965) recommend a value of $\alpha = 1.3$, based on experience. If the reflected point violates a constraint for one or more of the parameter values it is relocated just slightly within the appropriate constraint boundaries. The objective function is then reevaluated at B . If the point B persists in having the worst function value it is contracted one-half way towards the center of the remaining points. If the point B has a smaller objective function value than before, then new minimum and maximum values are sought. This series of steps constitutes one iteration. In this way the centroid of all the values proceeds to the global minimum, and is not affected by local minima or saddle shaped troughs in the objective function surface (Schwefel, 1981). If the minimum of all trial points remains the same for a number of iterations (usually set to 100), then the entire simplex figure is moved one half way towards the minimum. This modification to the constrained simplex procedure is made to speed convergence. The iteration procedure is terminated when all search points have objective function values less than a specified noise level. Under ideal (noise-free) conditions the objective function (4.2) behaves as a quadratic surface and the minimum objective function value corresponds to the true parameter values. However, when the data are contaminated by noise, the minimum does not correspond to the true parameter values. Its expected value is equal to the noise level in the data. This behavior is discussed in more detail in a subsequent section.

5.2 Controlled Random Search

In order to deal with solution divergence and false convergence, and to avoid having to compute gradients, inflexibility of starting guesses, and linear approximations in covariance calculations, we can adopt alternate methods of solution. For example, Silva and Hohmann(1983) used a controlled random search technique to optimize a non-linear functional in magnetic anomaly interpretation.

The basic CRS procedure (Price, 1977), is similar to the CS approach noted above. Here, initially a number of points N are chosen at random over some parameter space. Suppose also that the dimension of the parameter space is n . For each value chosen and objective function value for each is stored in a matrix \mathbf{A} . Next, determine the greatest function value of the set. At this point $n + 1$ values of the parameter set are chosen randomly from \mathbf{A} . Produce a new trial point by a reflection through the centroid of the $n + 1$ values. If this new value has a greater function value than the first found, replace in \mathbf{A} the coordinates and function value of those of the newly found point. If not, go back and find $n + 1$ new values from the set \mathbf{A} .

Price (1977) has numerous examples in his original paper, and the technique works well, particular on problems that has a few or many, local minima. Other extensions to the algorithm can

be found in Ali et al. (1997). Other examples in Geophysics can be found in Smith and Ferguson (2000).

Information potentially subject to copyright protection was redacted from this location. The redacted material is a figure from the reference information listed below in the caption.

Figure 3: Illustration of multiple minima and successful (CRS) solution, from Price (1977). This is a two parameter case and the solution is at the top left hand corner of parameter space.

6 Bayesian Methods and Monte-Carlo Methods

In recent publications (cf. Woodbury and Sudicky, 1992; Woodbury and Rubin, 2000) a full-Bayesian approach was used to obtain parameter estimates and variances. The "full-Bayesian" approach signifies that both parameter and hyperparameter determination is involved.

As mentioned in the introduction, it is the goal of this work to reconstruct a vector of hydrogeologic model parameters from observations of hydraulic heads. This is a non-linear inverse problem, and although for the cases presented in this report there are more data points than unknown parameters, the problem may be ill-posed and potentially lead to non-unique solutions. This is the

case, say if one attempts an inversion to determine hydraulic gradients, hydraulic conductivity, porosity and retardation factors from measurements of a tracer cloud. The reader will note that the above parameters are all coupled and their unique determination is very difficult. To so called 'Bayesians', inverse problems are problems of inference and this is the philosophy adopted in this work to circumvent the aforementioned concern.

Much has been written on the subject of Bayesian inference and different points of view apply (for review see Ulrych et al., 2000). The reader will note that we refer to a "Full-Bayesian" approach and this is to signify that the inference problem will consist of both primary parameter and hyperparameter estimation (Mohammad-Djafari, 1996; Woodbury and Rubin, 2000; Woodbury and Ulrych, 2000).

Bayesian inference supposes that an observer can define a prior probability-density function (pdf) for some random variable \mathbf{m} . This pdf, $p(\mathbf{m})$, can in principle, be defined on the basis of personal experience or judgment. However, applications of Bayesian probability theory have been hampered by the precise meaning and interpretation of probabilities and controversy surrounding the appropriate choice of prior pdfs. An orthodox view of probabilities dictates that frequencies measured in an experiment are equated to probabilities and 'prior' information is not allowed. An alternative viewpoint of probability, denoted as the Jaynes-Cox viewpoint (Jowitt, 1979), is one in which probabilities are equated with the degree of plausibility of a proposition and may have no frequency interpretation whatsoever. This viewpoint is essentially Bayesian and is readily applicable to the questions that scientists and engineers typically ask. A necessary component of the Jaynes-Cox view is the 'principle of maximum entropy' (PME) which replaces the need for subjective prior information in the Bayesian approach and forces all observers who possess common information to produce consistent results (Woodbury and Ulrych, 1998).

Woodbury and Ulrych (1993), Woodbury et al. (1995) and Woodbury (1997) deal with the estimation of appropriate prior pdf's for hydrogeologic applications. As shown by Woodbury and Ulrych (1993), $p(\mathbf{m})$ may have the form of a multivariate-truncated exponential distribution. This pdf preserves the statistical independence of the parameters. That is, if no correlation is known beforehand the maximum entropy principle does not inject any correlation into the result. In this manner $p(\mathbf{m})$ has the most freedom in assigning realizations of the process. It is important to note that the above approach (PME) of determining $p(\mathbf{m})$ is the one which is the most uncommitted with respect to unknown information.

Bayes' rule (for example; Press, 1989) quantifies how the prior pdf can be changed on the basis of measurements. Simply stated, Bayes' rule is

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Consider a vector of observed data \mathbf{d}^* . If the conditional pdf of \mathbf{d}^* given \mathbf{m} and some prior information I , is given by $p(\mathbf{d}^* | \mathbf{m}, I)$, then Bayes' rule states that

$$p(\mathbf{m} | \mathbf{d}^*, I) = \frac{p(\mathbf{d}^* | \mathbf{m}, I)p(\mathbf{m} | I)}{\int p(\mathbf{d}^* | \mathbf{m}, I)p(\mathbf{m} | I)d\mathbf{m}} \quad (6.1)$$

In the above, $p(\mathbf{m} | I)$ is the prior probability density of the model parameters, given some form of prior information, I , and $p(\mathbf{d}^* | \mathbf{m}, I)$ is the likelihood of observing \mathbf{d}^* given the model parameters and the prior information. This latter term is often referred to as a 'direct' as opposed to a subjective pdf. The term on the left hand side is called the posterior probability (after measurements are

taken into account). Finally the term in the denominator is a constant that ensures the posterior is normalized, but is also the actual pdf of observing a set of data, with the uncertainty in the model parameters taken into account. In the sections below we will outline how the various conditional pdfs and the prior information are defined and show how we can use Bayes' rule to reconstruct a vector of model parameters from heads.

Consider a finite element model for the hydraulic head predictions in an aquifer. Equation (6.1) is written in terms of a general non-linear model of the type

$$d_i = f_1(\mathbf{x}_i) \quad (6.2)$$

for $i = 1 \dots N$ where N is the number of predicted 'data' points and $\mathbf{x} = (x, y)$. Here, $f_1(x)$ depends upon a series of parameters \mathbf{m} which could consist of log-transmissivities, flux conditions and the like.

In the case where head measurements are taken, the associated noise-corrupted case is

$$d_i^* = f_1(\mathbf{x}_i) + \epsilon_i \quad (6.3)$$

Where the data d_i^* consist of a collection of discrete values of hydraulic heads and ϵ_i is the noise.

The inverse problem consists of trying to reconstruct the parameter vector \mathbf{m} , based on the observed data. As mentioned, the inverse problem is viewed in a Bayesian context; that is the inversion is viewed as a problem of inference. In order to solve the inference problem, we will use a Bayesian framework to 'update' a prior probability based on consideration of measurements. To apply Bayes' Theorem we need to assign a noise probability density which is consistent with the available information about the noise. If one could predict the 'true' data, the difference between d_i and d_i^* is just ϵ_i , the noise. If it is assumed that the noise has a value ϵ given prior information I , and if the second moment of the noise is known, σ_1 , then an application of the maximum entropy principle leads to a Gaussian distribution for ϵ (Bretthorst, 1988; Kapur, 1989):

$$p(\epsilon|\sigma_1, I) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{\epsilon^2}{2\sigma_1^2}\right) \quad (6.4)$$

Here σ_1 is taken as the root mean square (RMS) noise level and (6.4) is the least informative prior probability density for the noise that is consistent with the given second moment. Even if the second-moment of the noise is not known, the central limit theorem leads to the Gaussian form (Jaynes, 1983). We can treat the noise explicitly as an unknown in Bayes' theorem and then proceed to integrate its effects out.

Having a pdf for the noise and adopting the notation that ϵ_i is the noise at distance x_i , one can apply the product rule of probability theory (assuming independence) to derive the pdf that one would obtain a set of noise values $(\epsilon_1, \epsilon_2, \dots, \epsilon_N)$:

$$p(\epsilon_1, \epsilon_2, \dots, \epsilon_N|\sigma_1, I) = \prod_{i=1}^N \left[\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{\epsilon_i^2}{2\sigma_1^2}\right) \right] \quad (6.5)$$

Kapur (1989) shows that (6.5) arises naturally in the multivariate case when entropy is maximized with correlations unknown.

A non-linear least-squares approach would proceed by minimizing the combined sums in the argument in the exponential of (6.5), and the equivalent maximum likelihood procedure finds the

parameter set that maximizes the logarithm of (6.5). Neither approach incorporates prior information about the model parameters. On the other hand, the Bayesian methodology readily lends itself to the problem of updating prior probabilities based on uncertain field measurements. For example, Kitanidis (1986) and Woodbury and Sudicky (1992) outlined the Bayesian approach in which relevant prior information about the model is incorporated. In the recent work Jiang, Painter and Woodbury adopt a similar approach but following the suggestions of Jaynes and others (for example Kitanidis, 1986; Loredo, 1990; Rubin and Dagan, 1992; Woodbury and Rubin, 2000) they treated the noise variance σ_1^2 as a ‘nuisance’ parameters that is “removed” from further consideration by integration over that parameters (marginalization).

6.1 Metropolis Algorithm

This algorithm is of the type called MCMC, or Markov Chain, Monte-Carlo. The idea is to sample the posterior pdf (section 6.1) developed from a Bayesian inference approach. It can be shown that if the posterior is sampled correctly, then typical Monte-Carlo integrations and parameter inference can proceed in a straightforward manner. The following is an algorithm for the sampling of a posterior probability function, based on the Latin Hypercube. This is followed by a Metropolis algorithm for the sampling of the Likelihood function (see Mosegaard, 1998).

1. define a model space with $i = 1 \dots M$ parameters. For example, these could be transmissivity of a particular zone and so on. So,

$$\mathbf{m} = (m_1, m_2, m_3 \dots m_M)$$

2. for the model space vector \mathbf{m} we define a prior probability $p(\mathbf{m})$. So for each model parameter there is a marginal pdf and together they form the multivariate prior. These could be assumed to be Gaussian (initially) for the transmissivity zones
3. Using the Latin Hypercube, generate $j = 1 \dots MC$ realizations of \mathbf{m} from the prior pdf.
4. Take the first two realizations \mathbf{m}_1 and \mathbf{m}_2 , and we will call these $\mathbf{m}^{\text{current}}$ and \mathbf{m}^{new}
5. Compute

$$S(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^N (d_i^* - g(\mathbf{m}))^2$$

for both the new and current model vectors. Here d_i^* are the observed data points and there are N of these. For example measured heads at observational wells. The $g(\mathbf{m})$ function is the result of the MODFLOW (or some other) calculation at each of the data points.

6. Accept the transition from ‘old’ to ‘new’ with the following transition probability

$$P_{\text{accept}} = 1 \quad \text{if } S(\mathbf{m}_{\text{new}}) \leq S(\mathbf{m}_{\text{current}})$$

or

$$P_{\text{accept}} = \exp(-\Delta S/\sigma^2) \quad \text{where } S(\mathbf{m}_{\text{new}}) > S(\mathbf{m}_{\text{current}})$$

Note that

$$\Delta S = S(\mathbf{m}_{\text{new}}) - S(\mathbf{m}_{\text{current}})$$

and σ^2 is the variance of the observational noise that is assumed to be iid to begin with.

7. If the transition is accepted, $\mathbf{m}_{\text{current}} = \mathbf{m}_{\text{new}}$. If the transition is not accepted $(\mathbf{m}_{\text{current}})_{j+1} = (\mathbf{m}_{\text{current}})_j$
8. For the current parameter set $\mathbf{m}_{\text{current}}$ compute the travel time τ for the particle tracker to move from the repository to a compliance surface.
9. Store the 'current' value of the model parameters, $\mathbf{m}_{\text{current}}$ at each iteration, and τ . Do this even if the point repeats, i.e. does not make the transition from 'current' to 'new'.
10. Pick the next value of \mathbf{m}_j from the previously generated values (see 3). Call this new value \mathbf{m}_{new} . Go to 5.

6.2 Analysis

The set of current states $(\mathbf{m}_{\text{current}})_j$ where $j = 1 \dots MC$ generated by this algorithm with converge towards a set of sample from the posterior pdf $p(\mathbf{m}|d^*, I)$. This is essentially the Bayesian formulation where a prior pdf is modified according to observed data.

Now, the full ensemble of the values of \mathbf{m} and τ that are saved constitute a full independent sample out of the entire space of models that satisfy the data. The values of τ can simply be analyzed with conventional histograms, etc. Values of the say, smallest or least desirable travel time can be associated with a parameter set \mathbf{m} .

7 Stopping Criteria

In the presence of noise, it is important that the data predicted by an inverse does not reproduce the observed data exactly. A model generating a data misfit below its expected value will show structures that are artifacts of the data. The expected value of the data misfit is related to the noise in the data set. This point is discussed in the next paragraph. In the converse situation, if the data misfit is too large then information about the model contained in the data will be lost (Schlax, 1984). In the constrained simplex procedure (CS), the centroid of the L search points moves towards the global minimum. With each successive iteration the entire simplex figure becomes more compact. Ultimately, the centroid would move to the minimum and the entire simplex figure would collapse to the centroid value at that point. As mentioned, this maximum likelihood point may not give the most sensible estimate of the parameters in all cases. As well, meaningful covariances could not be generated due to the clustering effect of the points in parameters space around the objective function minimum. Conversely, it is undesirable to underfit the data and arrive at poor covariance estimates. Therefore, an appropriate stopping criteria must be defined. As noted by Fullagar and Oldenburg (1984), the appropriate level of misfit is a subjective choice. Note that no assumptions are made regarding the underlying statistical distribution of the functional, the parameters, or the data when the parameter space is sampled to compute covariances. However, in order to terminate the optimization procedure, it is necessary that some assumptions be made in this regard. When working with L_2 norm functionals, the assumption is made that the values of the objective function are normally distributed about some mean value. In the constrained simplex algorithm, when the average value of the L objective function values is below the expected value of noise, the procedure is terminated and a sampling of the parameter surface is carried out using all search points.

The appropriate measure of misfit is the χ^2 test when data errors are Gaussian, independent with zero-mean. Here the χ^2 misfit is defined as:

$$\chi^2 = \sum_{i=1}^I (d_i - d_i^*)^2 / \sigma_i^2 \quad (7.1)$$

and d_i is a computed data point. In the controlled random search method I is replaced by M . If $I \geq 5$ the expected value of χ^2 is approximately I and the standard error of χ^2 is $\sigma_{\chi^2} = \sqrt{2I}$. Any model is defined as acceptable when

$$I - \sigma_{\chi^2} \leq \chi^2 \leq I + \sigma_{\chi^2} \quad (7.2)$$

The most likely model is the one corresponding to $\chi^2 = I$. Therefore, we may apply this statistical test to (4.7). In (4.7) it is recognized that the expected value of the functional (noise level) is equal to $\sigma_h^2 M$, because the median of χ^2 is approximately equal to the number of degrees of freedom, and the objective function has been implicitly scaled by a factor of σ_h^2 . If the L_1 norm functionals are minimized then a different strategy is adopted. These functionals can be viewed as following exponential distributions. For the L_1 norm the expected value of χ^1 is $(\frac{2}{\pi})^{\frac{1}{2}} M$ with a variance of $(1 - \frac{2}{\pi})M$ (Parker and McNutt, 1980).

7.1 Computation of Covariances

The constrained simplex algorithm produces L acceptable models, each of which satisfies the data, and the centroid of the global minimum is expected to be more meaningful than the single estimate obtained by most algorithms when the surface of the objective function is non-Gaussian. This approach produces equivalent results to a Monte-Carlo solution to the inverse (see Parker, 1997).

Therefore, the average of all the L search points is taken as the best estimate of the parameters. The parameter space can also be sampled to obtain the parameter covariances, without assuming linear behavior close to the minimum value. An advantage to this strategy is that covariances of model parameters may be established regardless of the norm that is minimized. However, assumptions on the statistical nature of the covariances must be made in order to compute confidence intervals (for instance, Gaussian for L_2 , exponential for L_1), and to choose a stopping criterion. Silva and Hohmann (1983) adopt this strategy. Wiggins (1972) also showed that model covariances can be recovered from repetitive Monte-Carlo samples of parameters. These covariances are estimated:

$$\text{cov } m_{ij}^{est} = \frac{1}{(L-1)} \sum_{k=1}^L [P_{ik}^* - \bar{p}_i^*][P_{jk}^* - \bar{p}_j^*] \quad (7.3)$$

where $\text{cov } m_{ij}^{est}$ is an element of the covariance matrix, P_{ik}^* is the k 'th estimate of the i 'th parameter and, \bar{p}_i^* is the average value of the i 'th parameter such that:

$$\bar{p}_i^* = \frac{1}{L} \sum_{k=1}^L P_{ik}^* \quad (7.4)$$

An advantage of the method is that we may also compute the covariance of hydraulic head estimates at data points, $(\text{cov } h^{est})$ by storing the computed hydraulic head values in a matrix H^*

that correspond to each parameter estimate. For each search point ($k = 1 \dots L$) there will be a vector of length M of hydraulic head values. Then,

$$[cov h^{est}]_{ij} = \frac{1}{(L-1)} \sum_{k=1}^L [H_{ik}^* - \bar{h}_i^*][H_{jk}^* - \bar{h}_j^*] \quad (7.5)$$

and

$$\bar{h}_i^* = \frac{1}{L} \sum_{k=1}^L H_{ik}^* \quad (7.6)$$

In the results section, only the standard deviations in h^{est} will be examined. We do not consider interpretation of the off-diagonal terms.

8 Verification

The basic Price (1977) algorithm was programmed (CRS.FOR) with modifications suggested to use the worst value in the simplex rather than any random value for reflection. This code was tested against 'TEST2.FOR', a modified-constrained simplex method (suggested by Guin) for a variety of examples given in the Price (1977) paper. CRS was able to identify local minima much better than the Simplex method.

The MCMC program, METRO.FOR was developed that solves the inverse with the Metropolis algorithm (see section 6.1). A program was developed to illustrate the various solution methods to the regression of x against y for a linear relationship. This is a very simple but easy example to understand and allows for the comparison of the solutions against classical linear regression. The revised program to do this is called INVMETRO.FOR.

The first example consists of 50 values of x and y . The actual relationship is $y = ax + b$ where $a = 1.0$ and $b = 1.52$. Fifty values are sampled randomly from $x = 300$ to 500 and the corresponding y values were corrupted with Gaussian, additive noise $N[0, 1]$. The Metropolis algorithm was checked and the slope and intercept calculations were acceptable the variance of the intercept was inaccurate. The correct value was 7.922 and metro produced 0.306; the resolution of which is not known and will require more investigation.

The fifty values were input to the CRS code could produce the same, or very similar, regression coefficients for a, b . In the simulation below, the algorithm stopped searching when a value of the Chi squared misfit dropped to a value of N , the number of data. The following table indicates the statistics of a sample (size 100) of the entire surface that was sampled. Note that the actual values computed by the algorithm (parameters and minimum function) were actually the same as the Least squares results. See the attached code CRSX2.FOR and CRS.OT1.

This first example illustrates that the CRSX2 code (CRS algorithm) faithfully reproduces the expected results when compared to regression. Note that regression automatically assumes that there is no prior information, so one can set very non-informative priors. The priors in the above simulations were a low of 0 and high of 10 for both parameters.

The second example consists of (again) the same 50 values of x and y , chosen before. The fifty values were input to the program to check if the CRSX2 code, with a t-likelihood could produce the same results as the Chi squared likelihood. Note the "t" distribution assumes no knowledge of the noise in the data. Results in are in the file CRS.OT2. The priors for the second run are shown in Table 2:

Parameter	CRS	Variance	Least Squares	Variance
a (slope)	0.9965	3.5×10^{-6}	0.9966	3.68×10^{-6}
b (intercept)	1.611	0.7854	1.524	0.79226
σ (noise)	1.0283	-	1.0283	-

Table 1: Output CRS algorithm (Chi misfit) and comparison to linear regression

Parameter	CRS	Variance	Least Squares	Variance
a (slope)	0.9965	3.96×10^{-6}	0.9966	3.68×10^{-6}
b (intercept)	1.5979	0.8020	1.524	0.79226
σ (noise)	1.0283	-	1.0283	-

Table 2: Output CRS algorithm (t misfit) and comparison to linear regression

The priors for the first run are given in Table 3.

9 Summary to Date

Program CRSX2 performs well on LS regression problems, at least with the χ^2 norm is minimized. Gaussian minimization performed not as well and the t distribution misfit is the best although more thought needs to go into this in terms of the appropriate stopping criterion.

With regard to the Metropolis algorithm if one uses a symmetric proposal density (that is one that the probability of choosing x_{new} does not depend on x_{old}) then it is symmetric. The Metropolis algorithm has variants, like the *rejection* method. This may be useful to explore because many realizations are rejected during the random walk. This idea needs further research.

10 REFERENCES

- Abriola, L.M., A.H. Demond, K.E. Hayes, 1992. *Summary Report: Second forum on NSF Research activities in Subsurface Systems*, The Univ. of Michigan at Ann Arbor, MI..
- Ali, M.M., Torn, A., and S. Viitanen, 1997. A numerical comparison of some modified controlled random search algorithms, *J. Global Optimization*, **11**, 377-385.
- Bear, J., 1972. *Dynamics Of Fluids In Porous Media*, Elsevier, New York.
- Bear, J. and M.Y. Corapcioglu, 1981. A mathematical model for consolidation in a thermoelastic aquifer due to hot water injection or pumping, *Water Resour. Res.*, **17(3)**, 723-736.
- Bretthorst, G. L., 1988. Bayesian Spectrum Analysis and Parameter Estimation. In *Lecture Notes in Statistics No. 48*, (J. Berger, S. Fienberg, J. Gani, K. Krickeberg, and B. Singer (eds)), 209p.
- Carrera, J., and S.P. Neuman, 1986. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, **22(2)**, 199-210.
- Carrera, J., A. Medina, X. Sanchez-Vila, 1993. *In Quantitative Geology and Geostatistics*, 5, 779-792, Kluwer Publications.
- Copty, N. and Y. Rubin, 1995. A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resour. Res.*, **31(7)**, 1673-1686.
- Ginn, T. R. and J. H. Cushman, 1990. Inverse methods for subsurface flow - A critical review of stochastic techniques, *Stochastic Hydrology and Hydraulics*, **4(1)**, 1-26.
- Hubbard, S.S., Rubin, Y; Majer, E., 1997. Ground-penetrating-radar-assisted saturation and permeability estimation in bimodal systems, *Water Resour. Res.*, **33(5)**, 971-990.
- Hyndman, D. W., Harris, J. M. and S. M. Gorelick, 1994. Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, **30(7)**, 1965-1977.
- Jaynes, E. T., 1983. *Papers on Probability, Statistics and Statistical Physics*, D. Redel, Dordrecht-Holland.
- Jowitt, P.W., 1979. Bayesian estimates of material properties from limited test data, *Eng. Struct.*, Vol. 1, 170-178.
- Kapur, J.N., 1989. *Maximum Entropy Models In Science and Engineering*, Wiley, N.Y., 635p.

- Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, **22(4)**, 499-507.
- Kitanidis, P.K., 1997. A reassessment of the groundwater inverse problem - Comment., *Water Resour. Res.*, **33(9)**, 2199-2202.
- Lahkim, M.B., Garcia, L.A. and J.R. Nuckols, 1999. Stochastic modeling of exposure and risk in a contaminated heterogeneous aquifer. 2: Application of Latin Hypercube sampling, *Env. Eng. Sc.*, **16(5)**, 329-343.
- Loredo, T.J., 1990. From Laplace to supernova SN 1987A: Bayesian Inference. In *Maximum Entropy And Bayesian Methods*, (P. F. Fougere), Kluwer Academic Publishers.
- McKay, M.D., Beckman, R.J., and W.J. Conover, 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, **21(2)**, 239-245.
- McLaughlin, D. and L.R. Townley, 1996. A reassessment of the groundwater inverse problem, *Water Resour. Res.*, **32(5)**, 1131-1161.
- McLaughlin, D and L.R. Townley, 1997. A reassessment of the groundwater inverse problem - Reply., *Water Resour. Res.*, **33(9)**, 2203-2204.
- Mohammad-Djafari, A. , 1996. A full Bayesian approach for inverse problems. In *Maximum Entropy And Bayesian Methods*, (K. M. Hanson and R. N. Silver), Kluwer Academic.135-144
- Mosegaard, K., 1998. Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse Problems*, **14**, 405-426.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., and W.T. Vetterling, 1992. *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press.
- Press, S. J., 1989. *Bayesian Statistics: Principles, Models and Applications*, Wiley, New York, 237p..
- Rubin, Y. and G. Dagan, 1992. Conditional estimation of solute travel time in heterogeneous formations - Impact of transmissivity measurements, *Water Resour. Res.*, **28(4)**, 1033-1040.
- Rubin, Y., Mavko, G. and J. Harris, 1992. Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, *Water Resour. Res.*, **28(7)**, 1809-1816.
- Schreider, Yu. A., 1966. *The Monte Carlo Method*, Pergamon Press, 381pp.
- Smith, L. and D.S. Chapman, 1983. On the thermal effects of groundwater flow, 1 Regional scale systems, *J. Geophys. Res.*, **88(B1)**, 593-608.
- Smith, D. and J. F. Ferguson, 2000. Constrained inversion of seismic refraction data using the controlled random search, *Geophysics*, **65(5)**, 1622-1630.
- Stallman, R.W., 1960. Notes on the use of temperature data for computing groundwater velocity. In *6th Assembly on Hydraulics*, (Soc. Hydrotech. France), Nancy, June 28-30.
- Tarantola, A., 1987. *Inverse Problem Theory*, Elsevier, 613p.
- Ulrych, T.J., Velis, D.R., Sacchi, M.D. and A.D. Woodbury, 2000. *A Bayes tour of inversion: a tutorial*, Accepted, Geop. J..
- Wang, K., Shen, P-Y., and A. Beck, 1989. A solution to the inverse problem of coupled hydrological and thermal regimes. In *Hydrogeologic Regimes And Their Subsurface Thermal Effects*, (Editors A. E. Beck, G. Garven and L. Stegena), AGU Monograph Series Vol 47/IUGG 7.p 7-21

- Woodbury, A.D. and L. Smith, 1985. On the thermal effects of three dimensional groundwater flow, *J. Geophys. Res.*, **90(B1)**, 759-767.
- Woodbury, A.D., J.L. Smith, and W.S. Dunbar , 1987. Simultaneous inversion of temperature and hydraulic head data, 1. Theory and application using hydraulic head data , *Water Resour. Res.*, **23(8)**, 1586-1606.
- Woodbury, A.D. and J.L. Smith, 1988. Simultaneous inversion of temperature and hydraulic head data, 2. Application with thermal data , *Water Resources Research* , **24(3)**, 356-372.
- Woodbury, A. D., Narod, B., Chandra, B., and B. Bennist , 1991. Temperature measurements in geotechnical studies using low-noise high-resolution digital techniques, *Canadian Geotechnical Journal*, **28(5)**, 639-649.
- Woodbury, A.D., and E. A. Sudicky, 1992. Inversion of the Borden tracer experiment data: Investigation of stochastic moment models, *Water Resour. Res.*, **28(9)**, 2387-2398.
- Woodbury, A.D., and T. J. Ulrych, 1993. Minimum relative entropy: Forward probabilistic modeling , *Water Resour. Res.*, **29(8)**, 2847-2860.
- Woodbury, A. D., 1997. A probabilistic fracture transport model: Application to contaminant transport in a fractured clay deposit, *Can. Geotech. J.*, **34**, 784-798.
- Woodbury, A.D., 1999. *Simultaneous inversion of thermal and hydrologic data: Where do we go from here*, Eos Trans. AGU 80(46), F337 .
- Woodbury, A.D., F.W. Render and T.J. Ulrych, 1995. Practical probabilistic groundwater modeling , *Groundwater*, **33(4)**, 532-538.
- Woodbury, A.D., and T.J. Ulrych , 1998. Minimum relative entropy and probabilistic inversion in groundwater hydrology, *Stoch. Hydrol & Hydr.*, **12**, 317-358.
- Woodbury, A.D., and Y. Rubin, 2000. A Full-Bayesian approach to parameter inference from tracer travel time moments, *Water Resources Research*, **36(1)**, 159-171.
- Woodbury, A.D. and T.J. Ulrych, 2000. A Full-Bayesian approach to inference problem for steady-state flow, *Water Resour. Res.*, **36(8)**, 2081-2093.

10.1 Additional Readings and Bibliography

- Box, M.J., 1965. A new method of constrained optimization and a comparison with other methods, *Computer Journal*, **8**, 42-52.
- Brebbia, C.A., 1978. *The Boundary Element Method For Engineers*, Pentech Press.
- Birtles, A.B. and E.H. Morel, 1979. Calculation of aquifer parameters from sparse data, *Water Resour. Res.*, **15(4)**, 832-844.
- Carrera, J. and S.P. Neuman, 1986a. Estimation of aquifer parameters under transient and steady-state conditions, 1, Maximum likelihood method incorporating prior information, *Water Resour. Res.*, **22(2)**, 199-210.
- Carrera, J. and S.P. Neuman, 1986b. Estimation of aquifer parameters under transient and steady-state conditions, 2, Uniqueness, stability and solution algorithms, *Water Resour. Res.*, **22(2)**, 211-227.
- Carrera, J. and S.P. Neuman, 1986c. Estimation of aquifer parameters under transient and steady-state conditions, 3, Application to synthetic and field data, *Water Resour. Res.*, **22(2)**, 228-242.

- Clifton, P.M. and S.P. Neuman, 1982. Effects of kriging and inverse modeling on conditional simulation of the Avra Valley Aquifer in Southern Arizona, *Water Resour. Res.*, **18(4)**, 1215-1234.
- Cooley, R.L., 1977. A method of estimating parameters and assessing reliability of models of steady state groundwater flow, 1. Theory and numerical properties, *Water Resour. Res.*, **13(2)**, 318-324.
- Cooley, R.L., 1979. A method of estimating parameters and assessing reliability of models of steady state groundwater flow, 2. Application of statistical analysis, *Water Resour. Res.*, **15(3)**, 603-617.
- Cooley, R.L., 1982. Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 1. Theory, *Water Resour. Res.*, **18(4)**, 965-976.
- Cooley, R.L., 1983. Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 2. Applications, *Water Resour. Res.*, **19(3)**, 662-676.
- Cooley, R.L. and R.L. Naff, 1985. *Regression modeling of groundwater flow*, USGS Open File Report 85-180.
- Dagan, G., 1985. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem, *Water Resour. Res.*, **21(1)**, 65-72.
- Draper, N.R. and H. Smith, 1981. *Applied Regression Analysis*, Wiley & Sons, New York.
- Freeze, R.A. and J.A. Cherry, 1979. *Groundwater*, Prentice-Hall.
- Fullagar, P.K. and D.W. Oldenburg, 1984. Inversion of horizontal loop electromagnetic frequency soundings, *Geophysics*, **49(2)**, 150-164.
- Gavalas, G.R., P.C. Shah and J.H. Seinfeld, 1976. Reservoir history matching by Bayesian estimation, *Soc. Pet. Eng. J.*, **16**, 337-350.
- Hoeksema, R.J. and P.K. Kitanidis, 1985. Comparison of Gaussian conditional mean and kriging estimation in the geostatistical solution to the inverse problem, *Water Resour. Res.*, **21(6)**, 825-836.
- Journel, A.G. and C.J. Huijbregts, 1978. *Mining Geostatistics*, Academic Press, New York.
- Kitanidis, P.K. and E.G. Vomvoris, 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state and one dimensional simulations), *Water Resour. Res.*, **19(3)**, 677-690.
- de Marsily, G., G. Lavedan, M. Boucher, and G. Fasanino, 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In *Geostatistics For Natural Resource Characterization, Part 2*, (G. Verly et al., eds.), Reidel.
- de Marsily, G., 1986. *Quantitative Hydrogeology*, Academic Press.
- Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, New York.
- Neuman, S.P., 1980. A statistical approach to the inverse problem in aquifer hydrology, 3. Improved solution technique and added perspective, *Water Resour. Res.*, **16(2)**, 331-346.
- Neuman, S.P., 1982. Statistical characterization of aquifer heterogeneities: An overview. In *Recent Trends in Hydrogeology*, (T.N. Narasimhan, ed.), G.S.A. Special Paper 189.

- Neuman, S.P. and S. Yakowitz, 1979. A statistical approach to the inverse problem in aquifer hydrology, 1. Theory, *Water Resour. Res.*, **15(4)**, 845-860.
- Neuman, S.P., G.E. Fogg and E.A. Jacobson, 1980. A statistical approach to the inverse problem in aquifer hydrology, 2. Case study, *Water Resour. Res.*, **16(1)**, 33-58.
- Neuman, S.P. and E.A. Jacobson, 1984. Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels, *Math. Geol.*, **16(5)**, 499-521.
- Oldenburg, D.W., 1984. An introduction to linear inverse theory, *IEEE Trans. Geosc. and Remote Sens.*, **GE-22(6)**, 665-674.
- Parker, R.L., 1977. *Understanding inverse theory*, Ann. Rev. Earth Planet. Sci. 35-64
- Parker, R.L. and M.K. McNutt, 1980. Statistics for the one-norm misfit measure, *J. Geophys. Res.*, **85(B8)**, 4429-4430.
- Price, W. L., 1965. A controlled random search procedure for global optimization, *Computer Jour.*, **8**, 42-52.
- Reklaitis, G.V., A. Ravindran and K.M. Ragsdell, 1983. *Engineering Optimization, Methods and Applications*, Wiley, New York.
- Schlax, M.G., 1984. *A practical solution for linear inference computations*, M.Sc. Thesis, Department of Geophysics and Astronomy, University of British Columbia.
- Schweifel, H.-P., 1981. *Numerical Optimization of Computer Models*, Wiley.
- Schweppe, F.C., 1973. *Uncertain Dynamic Systems*, Prentice-Hall.
- Silva, J.B.C. and G.W. Hohmann, 1983. Nonlinear magnetic inversion using a random search method, *Geophysics*, **48(12)**, 1645-1658.
- Tarantola, A., and B. Valette, 1982. Generalized nonlinear inverse problems using the least square criterion, *Ann. Rev. Geophys. Space Phys.*, **20(2)**, 219-232.
- Townley, L.R. and J.L. Wilson, 1983. Conditional second moment analysis of groundwater flow: The cumulative effects of transmissivity and head measurements. In *Proc. Int. Conf. Groundwater and Man*, (Sydney), Australia.
- Wiggins, R.A., 1972. The general linear inverse problem: Implication of surface wave and free oscillations for earth structure, *Rev. Geophys. and Space Phys.*, **10(1)**, 251-285.
- Yeh, W.W-G., 1986. Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, **22(2)**, 95-108.
- Yeh, W.W-G., Y.S. Yoon, and K.S. Lee, 1983. Aquifer parameter identification with kriging and optimum parameterization, *Water Resour. Res.*, **19(1)**, 225-233.3.

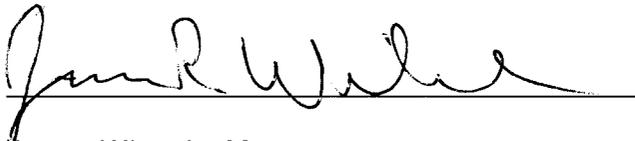
Parameter	Lower	Expected	Upper	PDF Type
a (slope)	0.	0.9965	2.0	TE
b (intercept)	-300.	0.3939	300.	TE
σ (noise)	-	1.0	-	none

Table 3: Parameters and ranges adopted for first verification example. TE refers to truncated exponential distribution.

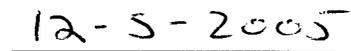
No new entries to Scientific Notebook #664E since September 30, 2004. All entries to pages 1-25 have been made by Allan D. Woodbury. Page 26 added by James Winterle on December 5, 2005 to add closing statement. This scientific notebook is being closed.

Closing Statement

I have reviewed this scientific notebook and find it in compliance with QAP-001. There is sufficient information regarding methods developed that another qualified individual could repeat the activity. There are not data collections, data analyses, or test activities documented in this notebook.

A handwritten signature in black ink, appearing to read "James Winterle", written over a horizontal line.

James Winterle, Manager

A handwritten date "12-5-2005" written in black ink, positioned to the right of the signature and underlined.