TRAINING COURSE

ON

SOFTWARE FOR GROUNDWATER DATA MANAGEMENT

UNDER

WORLD BANK FUNDED HYDROLOGY PROJECT

LECTURE NOTES
ON

GROUNDWATER MODELLING SOFTWARES
(UNIT-3)

BY

SUDHIR KUMAR

ORGANISED BY

NATIONAL INSTITUTE OF HYDROLOGY
ROORKEE - 247 667
INDIA

APPLICATION OF MODINV

1.0 INTRODUCTION

The MODINV suite of software comprises a number of programs, built to enhance the usefulness of the popular USGS finite difference flow model, MODFLOW. The MODINV is basically a parameter optimisation program for MODFLOW. Using this software, the specific values taken by any parameter type that MODFLOW can read as a 2-dimensional data array can be optimised such that model-generated heads are as well matched as possible to those observed in the field. Steady state and transient, single and multi-layer, confined and unconfined models can all be calibrated in this manner. As well as providing optimised parameter values, MODINV indicates the reliability of these aquifer parameter value estimates, given the observed head data that is used in calibration. Parameter values can be fixed, grouped or transformed to enhance optimisation stability and efficiency.

2.0 MODINV SOFTWARE

Model parameter (or input) estimation is often referred to as the "inverse problem" to distinguish it from the "forward problem". The latter refers simply to the process of mathematical modelling and the means by which this is achieved for different physical systems. For groundwater modelling, MODFLOW is a program that carries out forward modelling using the finite-difference method; aquifer parameters and inputs are calculated. On the other hand, when model outputs are known and an attempt is made to solve for one or a number of the model parameters or inputs, then inverse modelling is being attempted.

Here, both aquifer input (eg. recharge) and physical properties (eg. transmissivity) will together be referred to as model "parameters" for the sake of simplicity of expression. Table-1 provides a list of such parameter types. As discussed latter, on the assumption that some of the model parameters are known, we are attempting to ascertain the values of the other parameters on the basis of a set of water level or head measurements taken at a number of boresites at one or a number of times.

Table -1 Two dimensional real arrays that can serve as parameter types.

MODFLOW Package	Array	
BCF	primary storage capacity	(all layers)
	secondary storage capacity	(all layers)
	transmissivity	(all layers)
	hydraulic conductivity	(all layers)
	layer bottom elevation	(all layers)
	layer top elevation	(all layers)
	vertical hydraulic conductivity/thickness	(all layers but bottom)
RCH	recharge rate	(all stress periods)
EVT	surface elevation	(all stress periods)
	maximum EVT rate	(all stress periods)
	EVT extinction depth	(all stress periods)

Fundamental to the operation of most inverse modelling algorithms is an ability to calculate model outputs using current estimates of model parameters, i.e., to carry out routine solutions of the forward problem. These model outputs are compared with measurements (in the present case through a weighted sum of squared differences criterion) and the parameters are then adjusted to obtain a more favourable comparison. MODINV uses MODFLOW as its forward processor. However as field-observed head data exist at only a discrete number of bores, the two-dimensional head arrays constituting MODFLOW's output are interpolated to yield MODFLOW-predicted heads at these boresites; it is these interpolated head values that are compared with historical or steady-state head data, and it is the weighted squared sum of the differences between these two sets of heads which is minimized. Hence the total forward model can be considered to be MODFLOW plus the two-dimensional head array interpolation procedure; the forward model outputs are then the heads at boresites whose positions and layer numbers are nominated by the user.

2.1 THE MATHEMATICAL MODEL

The purpose of a mathematical model is to predict the behaviour of a system as it responds to changing conditions. If we consider these conditions as inputs to the model, then once we know the parameters of the model, it is a simple matter to obtain model outputs for as many different inputs as we like. For a groundwater model the inputs are normally considered to be the sources (or sinks) of water (for example recharge, EVT rate, well pumping rate, etc) while the parameters are aquifer physical properties such as transmissivity and storage capacity. Boundary conditions such as lateral model inflow rate or constant head levels are often considered to be part of the model itself, being neither an input nor a parameter, the latter term normally referring to something which can be adjusted at the model calibration stage. The model outputs are the heads or water levels in the aquifer

The partial differential equation describing groundwater flow is

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)-W=S_s\frac{\partial h}{\partial t} \qquad (1)$$

where,

K_{xx}, K_{yy}, K_{zz}	: hydraulic conductivities along the major axes
h	: potentiometric head
W	: volumetric flux per unit volume and represents sources and or sinks
S_s	: is the specific storage of the aquifer
t	: time

Considering the above equation together with the given boundary conditions as the model, then the parameters (K_{xx}, K_{yy}, K_{zz} and S_s) are the model, "distributed", parameters.

The fact that they are distributed means that they require for their complete description a knowledge of their value at every point within the three-dimensional space occupied by the aquifer. That is, as functions of position, they should be represented as $K_{xx}(x,y,z)$, $K_{yy}(x,y,z)$, etc. Model inputs are functions of time as well as location, and hence can be written as $W(x,y,z,t)$. Similarly, the model output, h, is both time and space dependent and can likewise be represented as $h(x,y,z,t)$. Hence equation 1 can be represented conceptually by the following equation:

$$M(K_{xx}, K_{yy}, K_{zz}, S_s; W) = \mathbf{h} \qquad \ldots (2)$$

where the semi-colon in the bracketed term above separates the model parameters (time-independent), from the inputs (time-dependent). Again boundary conditions, (though they may be time-dependent) are assumed to be part of the model, and hence the "M" term.

To model a natural system, simplifying assumptions are made to all terms of equation (2). Spatial discretization of the M operator, through which a differential equation is converted to a matrix equation, is fully described in the MODFLOW manual. To construct a parameter estimation algorithm on which to base MODINV, some further simplifications are made. In particular, it is assumed that all distributed parameters are "piecewise constant", ie. they are constant within each of a number of zones which, when put together, cover the area of the model. These zones of constant parameter value do not necessarily coincide for each different parameter type; however each parameter type now requires for its complete description only a few numbers, these representing the values that the parameter takes within each zone of constancy for that parameter.

Model inputs, W, are subdivided by MODFLOW into two types, viz. those that are distributed across the mesh (recharge and EVT) and those that take on "point" or "line" distributions (well recharge, drainage, river, general head boundary). For the first group MODFLOW requires either a two-dimensional matrix of inputs (recharge) or a number of two-dimensional matrices by which the source or sink two-dimensional matrix can be calculated (EVT); for the second group MODFLOW requires either the model input, or the means by which it can be calculated, at each pertinent cell, the latter being nominated specifically by row, column and layer number. MODFLOW assumes that each input is piecewise constant in the time domain, the period of constancy being referred to as a "stress period". For recharge and EVT, let us also assume that the recharge rate and the properties that determine the EVT rate can be represented, like the model parameters, by a limited number of spatial zones within each of which the recharge or a particular EVT characteristic is constant for a given stress period. These zones do not need to coincide for different stress periods or for each of the three different EVT characteristics and recharge.

Let us now introduce a further simplification by acknowledging that, when calibrating the aquifer, we only have field measurements to compare with the model output (head or water level) at a finite, relatively small, number of points, these being the locations of bores within the model area. Furthermore these borehole heads are known at only one or a finite number of times, corresponding to field sampling episodes. Assuming that we have a means by which a MODFLOW head array output can be interpolated onto these same boresites at

specific times corresponding to the measurement times, our model output can be considered as a number of sequences of real numbers, each sequence providing a time series of water levels or heads at a particular bore. Forcing these model heads to coincide as closely as possible with the measured heads at the measurement times is the basis for model calibration. Of course in the steady-state case, there is only one observed head and one model head for each bore.

After making all these assumptions, equation (2) can be rewritten

$$M(p) = \mathbf{h} \qquad \ldots. (3)$$

where **p** is finite-dimensional vector of numbers representing all of

i. the values taken by each of the model parameters within each of their respective constant-parameter zones.

ii. the values taken by recharge rate within each of its constant-recharge zones for each stress period,

iii. the values taken by each of the three parameters determining EVT in each of their respective constant-value zones for each stress period, and

iv. the values taken by the coefficients that determine single cell model inputs, (ie. rivers, drains, wells and general head boundaries) in each of the model cells that are subject to such inputs for each stress period.

h is also a finite dimensional vector. It consists of head (ie. model output) values at specific boresites at times for which a model output has been requested. If **p** contains N elements and **h** contains M elements, then M is a continuous vector function from N-dimensional space to M-dimensional space.

M as written in equation (3) does not have an inverse; i.e., if **h** is known it does not follow that **p** can be determined. This is easily demonstrated by considering a model that is subject to both recharge and EVT. If the former is increased everywhere while the latter is decreased by the same amount, the model's output heads will be unchanged. Hence, given h, it is impossible to determine both recharge and the EVT coefficients. With other model parameters, such as the transmissivity distribution, also having a strong effect on h, it is easy to see that there are many different p's which will produce the same, or almost the same, h. Some of these different p's can be obtained from any given **p** which satisfies (3) by varying some of its elements in such a way that the effect of changing one or a number of these elements is balanced exactly (or almost exactly) by simultaneously changing one or a number of its other elements in a certain manner. This is an example of parameter correlation, of which more will be said later.

So it is obvious that if we are going to use measured aquifer heads as the basis for model calibration, it will be necessary to assume that some elements of **p** are known. We will then be left with the problem of estimating the remaining elements, for which we may or may not be able to obtain a solution, depending on the degree of parameter correlation that remains. If parameter correlation is still too high, a numerical inversion algorithm will not converge to a solution or, at best, will show signs of instability. This is not the fault of the algorithm, for it cannot answer an impossible question. In general, the fewer the parameter types for which you require estimates, and the fewer constant-parameter-value sub-areas for those parameter types, the more likely is the algorithm to perform well. Of course, with fewer parameter sub-areas the degree of fit between model and observed heads may not be as good (see later); however you may not be able to escape the fact that your borehole head data is insufficient for any finer detail of aquifer property determination.

Returning to equation (3), then, we can rewrite it as follows:

$$m(p) = \mathbf{h} \qquad \ldots. (4)$$

where those parameters and inputs of the model which are assumed known are now included in the revised model function, m, and **p** is of reduced dimension. In fact it must be of smaller dimension than **h** because when we derive a set of simultaneous equations to solve for the elements of **p** in the next section, there must be fewer unknowns than equations (the number elements of h) in order for the system of equations to be capable of solution.

In the MODINV algorithm, the elements of **p** can be the parameter values taken by up to three different types of parameter within their respective constant-parameter sub-areas. Parameter types can be anything that MODFLOW can read as a two-dimensional array of real numbers. This includes transmissivity, storage capacity, recharge for any (or up to three) stress periods, etc.; one "parameter type" corresponds to each such two-dimensional array. Though recharge is strictly a model input rather than a parameter, and though quantities such as maximum EVT rate likewise govern another model input, we will refer to anything that MODFLOW can accept as a two-dimensional real array (with the exception of initial heads) as a "parameter type" in the discussion that follows. The values taken by any such "parameter type" within its (unique) zones of piecewise-constancy are thus admissible elements of the vector p.

2.2 THE INVERSE PROBLEM

m in equation (4) is a continuous function of p, mapping N-dimensional space into M- dimensional space where N is the number of elements of **p** (equal to the number of individual parameter values requiring estimation) and M is the number of elements of **h** (equal to the number of observation times multiplied by the number of observation bores). Let J be the Jacobian matrix of m. This is the matrix whose ith row is the derivative of the ith element of h with respect to each of the elements (in order) of p. Hence J has M rows (same as h) and N columns (same as p). Let h_0 and p_0 satisfy equation 4. If we now change each of the elements of p_0 by a small amount to obtain the vector p, the resultant

change to the head vector can be approximately calculated as

$$\Delta h = h - h_0 \approx J(p - p_0) = J\Delta p \qquad \ldots. (5)$$

where the approximation improves as Δp, and hence Δh, are reduced. Let us assume that we presently have an estimate for each of the values of each of our unknown parameter types within their respective constant-value sub-areas; let this vector of estimates be p_0 . Using our model (ie. MODFLOW), h_0 is readily calculated using equation 4. Let h in equation 5 be the vector of heads observed from bores in the aquifer; of course the elements of h must pertain to the same bores at the same measurement times as do the corresponding elements of h_0. Then in equation 5, there is only one unknown, viz. p, the vector the model is calibrated provided, of course, that the constant-parameter sub-area boundaries are well chosen.

However, there is a problem. If M, the number of heads, is less than N, the number of unknowns, there are less equations than unknowns and we cannot solve for p. If M is greater than N, then we could select a set of N of the M equations represented by (5) and obtain a solution for p; however, selecting another set of N equation may give us a different solution and we are left with the question for which solution is best. If N and M are equal we obtain a unique solution for p, but the lack of redundancy in our observations gives us no protection against the effects of head measurement errors or of an inappropriate parameter sub-area zonation scheme. Without such redundancy, parameter value estimates may be erroneous, and the calibrated model may thus provide a poor basis for predicting future aquifer behaviour.

To solve the inverse problem, then, we must formulate it slightly differently. Again, let us assume that we have a current set of parameter estimates, p_0, and hence a corresponding set of model-generated heads, h_0 ,calculated on the basis of p_0 for a number of bores at a number of times. Corresponding to the elements of h_o we have a vector, h_m, of head measurements. We wish to improve our current estimate, p_o, to a new parameter vector p_1, generating a head vector h_1 through equation (4) that is "closer" to h_m than h_o . As it is foolish to expect that we can choose our parameters such that all the corresponding elements of h_1 and h_m are exactly equal, and as we wish to make use of all the measured heads in establishing p_1, we choose as our criterion for determination of p_1 that

$$\Phi = (h_1 - h_m)^t \; W(h_1 - h_m) = \text{minimum} \qquad \ldots.(6a)$$

where the superscript "t" refers to the transpose of the vector. Φ is often referred to as the "objective function". In this equation, W is a "weighting matrix" which, in MODINV, is assumed to contain diagonal elements only. Hence equation (6a) can also be written as

$$\Phi = \sum_{i=1}^{M} (h_{1i} - h_{mi})^2 \, w_i = \text{minimum} \qquad \ldots. (6b)$$

where w_i is the i'th diagonal element of W. In other words, the weighted sum of the squares of the differences between model and observed heads must be a minimum. The use of weights allows us to give measured head values which we "trust" a greater say in the determination of parameter values than those which we do not. Alternatively it provides a means by which we can enforce a condition that heads calculated in a particular model sub-area, or at a particular time, be better matched to reality than those elsewhere or at other times, if they cannot all be simultaneously matched as well as we would like. The diagonal elements of the matrix W, then, can also be considered as a weighting vector of dimension M, each element of which determines the importance of the corresponding element of h_m in governing the estimation process; it is good practice to select these weights from the interval [0, 1]. If measurements are missing from some bores at certain times, you can use "dummy" measurements for the corresponding elements of h_m , and set the corresponding measurement weights to zero; in this way, such elements have no effect on the estimation process.

Defining

$$\Delta h = h_1 - h_o ; \qquad \Delta p = p_1 - p_o \qquad \text{.... (7)}$$

where h_1, P_1 and h_o , p_o jointly satisfy (4), then Δh and Δp approximately satisfy (5), with the approximation improving with proximity of h_o to h_1 and p_o to p_1 , ie.

$$\Delta h = h_1 - h_o \approx J(p_1 - p_o) = J\Delta p \qquad \text{.... (8)}$$

Substituting (8) into (6)

$$(\Delta h + h_o - h_m)^t W(\Delta h + h_o - h_m) = \text{minimum}$$

which is equivalent to:

$$\Delta h^t W \Delta h + \Delta h^t W(h_o - h_m) + (h_o - h_m)^t W \Delta h + (h_o - h_m)^t W(h_o - h_m) = \text{minimum}$$

ie.

$$\Delta h^t W \Delta h + 2\Delta h^t W(h_o - h_m) = \text{minimum} \qquad \text{.... (9)}$$

where constant terms have been ignored because they cannot be minimized and we have made use of the fact that W is a symmetric matrix. Substituting (8) into (9):

$$(J\Delta p)^t W(J\Delta p) + 2(J\Delta p)^t W(h_o - h_m) = \text{minimum}$$

ie.

$$\Delta p^t (J^t W J)\Delta p + 2\Delta p^t J^t W(h_o - h_m) = \text{minimum} \qquad \text{.... (10)}$$

Now if both terms on the left of (10) are differentiated with respect to each element of Δp and the right hand side is equated to zero in each case (because of the minimum), we obtain

$$(J^tWJ)\Delta p = -J^tW(h_o - h_m) \quad \text{.... (11)}$$

J^tWJ is a N x N matrix (often referred to as the "normal" matrix); hence equation (11) represents N equations in N unknowns which can be solved for the elements of Δp provided J^tWJ is not singular. A singular matrix implies that, even though there may have been more borehole head observations than there are unknown parameter values, there is still insufficient information for unique parameter value determination. For example if, in a steady-state model, you ask that both recharge and transmissivity be determined everywhere in the model, you will obtain a singular normal matrix because the values taken by one parameter type (eg. transmissivity) for a particular head distribution, depend on the values taken by the other parameter type (recharge). However if you assume that recharge is known everywhere you can then estimate the transmissivity distribution, and vice versa.

Problems can also arise if the normal matrix is nearly singular; if this occurs MODINV may have trouble minimizing Φ of equation (6). Considering the steady-state problem again, this can occur if you attempt to estimate the transmissivity distribution with too great a spatial precision in an area where there are too few borehole head observations. If there are many model sub-areas in a zone of sparse measurement, it will be possible to simultaneously vary the transmissivities of these zones in such a way as to maintain the model heads in the observation bores relatively unchanged. This means that the observed borehole heads do not have the power to tell you what the individual transmissivities are; this is the phenomenon of high parameter value correlation again. The higher the degree of such correlation, the closer will the normal matrix approach to singularity, and the greater will be the possibility of numerical instability.

From equation (11)

$$\Delta p = - (J^tWJ)^{-1}J^tW(h_o - h_m) \quad \text{.... (12)}$$

while from (7)

$$p_1 = \Delta p + p_o \quad \text{.... (13)}$$

and an improved set of parameter values has been obtained. Because (8) is only approximately correct (especially if p_0 is a poor estimate of the aquifer parameters so that Δp needs to be large), the process outlined above needs to be repeated to obtain another estimate p_2, then another, p_3 etc. until further improvement is impossible, or until Φ) of equation (6) is low enough to indicate an acceptably good fit between model and field data.

MODINV does not use equation (12) for parameter improvement; rather, it uses a slight modification of it. Defining (Φ_0 as the value of the objective function (weighted sum

of squared head differences between model and observed heads) with model heads calculated on the basis of the parameter set p_0, it can be shown that

$$\nabla\Phi_0 = 2J^tW(h_o - h_m) \quad \ldots\ (14)$$

where $\nabla\Phi_0$ is the gradient of Φ_0 with respect to the elements of p_0. Substituting (14) into (12) and using (13) we obtain:

$$p_1 = P_0 - \frac{(J^tWJ)^{-1}\nabla\Phi_0}{2}$$

Generalizing this to the i+1th iteration:

$$p_{i+1} = P_i - \frac{(J^tWJ)^{-1}\nabla\Phi_i}{2} \quad \ldots\ (15)$$

The hardest part of using equation (15) to improve parameter value estimates is calculating the Jacobian matrix, J. In MODINV, finite differences are used. For p_i, the parameter set at the beginning of the i+1th iteration, MODFLOW is run to obtain the corresponding h_i vector. Then a single element of p_i is increased by a small amount and MODFLOW is run again to determine a new set of heads, h_i^j, where the superscript "j" indicates that the jth parameter value, ie. the jth element of p_i, was varied. The jth column of the Jacobian is then calculated as the vector

$$\frac{h_i^j - h_i}{\delta p_i^j}$$

where δp_i^j is the change to the jth element of the parameter value vector p_i. When convergence has nearly been obtained, MODINV uses central differences for greater accuracy in derivative calculations. In this case each p_i^j is first increased, and then decreased, by δp_i^j to obtain, respectively, the set of heads $h_{i(1)}^j$ and $h_{i(2)}^j$. The jth column of the Jacobian is then approximated as the vector

$$\frac{h_{i(1)}^j - h_{i(2)}^j}{2\delta p_i^j}$$

The gradient vector is calculated in similar fashion.

Calculation of head derivatives by finite differences is very time consuming. In fact for most MODINV runs, this accounts for over 70% of the computing time. There are more efficient methods of derivatives calculation that can be used under confined aquifer conditions (ie. when equation (1) is linear); there are also other optimization methods available that do

not require the calculation of explicit head derivatives with respect to individual parameter values for their implementation. However the calculation of derivatives of head with respect to parameter values by finite differences is perfectly general, being useable for any distributed parameter type that MODFLOW can read, under both confined and unconfined conditions. Also the Gauss-Marquardt method (see next section) which makes use of these derivatives for optimization, converges to a solution in fewer steps by far than most other methods. Note that the reason why it does not completely converge in a single step is that the relationship between Δh and Δp upon which all of the above theory is based (equation 8) is only approximately correct.

MODINV provides you with the choice of optimizing either parameter values themselves or the logarithms or "logistic transformations" of parameter values, the last being defined by the relationship

$$P_t = \log [p/(p-1)] \quad \ldots\ (16)$$

where **p** is the parameter value and p_t is its logistic transformation. In the latter two cases, everything said so far about parameter estimation and derivatives calculation applies just as well, provided the transformed parameter value, rather than the parameter value itself, is considered as the parameter value to be estimated. There are two advantages that accompany log or logistic parameter value transformation in certain cases:

1. There is strong evidence that the probability distribution satisfied by some aquifer parameters (transmissivity, hydraulic conductivity) is log-normal, rather than directly normal. For estimating such parameter types it is better to optimist the logarithms of the parameter values than the parameter values themselves because parameter stochastic property inferences drawn from a least squares inversion (discussed later), assume that parameter values possess a multidimensional normal probability distribution. Also, in such cases, optimization convergence appears to be faster and more stable.

2. Some parameters must take values within a certain range for them to have any meaning. For example, transmissivity must never be negative and storage capacity must be between 0 and 1. If a parameter is left at the mercy of an iterative adjustment procedure that pays no attention to whether it is given sensible values or not, errors could result. By optimizing the log of the parameter, the parameter itself can never become negative. Similarly, no matter what value is given to the logistic transformation of a parameter, the parameter itself will never be outside the interval (0,1).

2.3 THE GAUSS-MARQUARDT-LEVENBERG METHOD

Equation 15 describes the Gauss method of solution of the inverse problem. Defining

$$N = J^t W J \quad \ldots\ (17)$$

$$f_i = \nabla\Phi_i / 2 \quad \text{.... (18)}$$

(15) can be rewritten:

$$\Delta p_i = -N^{-1} f_i \quad \text{.... (19)}$$

While the method often converges rapidly (the more rapidly it converges, the fewer optimization iterations are required), its performance is not perfect, especially in cases where parameter correlations are high. To make the method more robust and reliable, it is usually modified in a manner similar to that originally outlined by Levenberg in 1944 and then by Marquardt in 1963, though the method is normally named after the latter author.

To implement the method, N. the normal matrix, is modified by increasing all of its diagonal elements. In the MODINV algorithm N is modified to N_m by adding a fixed amount, λ, to all diagonal elements, ie.

$$N_m = (N + \lambda I) \quad \text{.... (20)}$$

where λ is a positive constant and I is the N x N identity matrix. If N_m now replaces N in (19) a Δp is obtained which has been found to be more reliable in many cases than that obtained solely using the Gauss method (ie. a λ of zero). (It should be noted that when λ is high, the resulting Δp is the same as that obtained using the so-called "gradient" method of parameter value adjustment.) However we are still left with two problems when using the Marquardt enhancement of the Gauss method. These are

(i) how should k be determined at each optimization iteration, and

(ii) once a Δp is obtained using N_m in place of N in (19), what fraction of this Δp should actually be added to p_i to determine the p_{i+1} which yields the minimum weighted sum of squared differences between model and observed heads.

The first problem, the value of λ, is not formally solved; instead, experience dictates the best choice. In the initial stages of an estimation process λ should normally be high (otherwise the solution may not converge), especially if the initial parameter value estimates (ie. the elements of p_0) are poor. As the process progresses and the upgraded estimates become better and better, λ is normally reduced because the Gauss method has a superior performance over the gradient method for parameter values which are close to optimum.

PREINV, the MODINV preprocessor, asks you for an initial λ. This is the λ value used in the very first attempt at reduction of Φ, the objective function, of equation (6). For each MODINV optimization iteration, one or a number of λ's are tried. For the first optimization iteration the initial λ is used first; for later optimization iterations a λ is first tried which is reduced by a certain factor (supplied by you to PREINV) below that which worked best for the previous iteration. Unless the objective function is drastically reduced with this first λ, a second λ, reduced from the first by this same factor, is tested. If the use

of the second λ achieves a **p** vector that lowers Φ by a significantly greater amount than that achieved through the use of the first one, then λ is lowered again and the process is repeated. When it is judged that Φ cannot be further significantly reduced by lowering λ, or a maximum of five it's have been tested, the best **p** is accepted as the updated parameter set. Sometimes, however, λ must be increased to obtain an improved parameter set. While experimenting with different values of λ in this fashion is a little cumbersome, and certainly consumes computer processing time, it is worth the effort because it is important, for each optimization iteration, to achieve a good parameter improvement. With each new optimization iteration the Jacobian matrix must be recalculated, and this is the most time-consumptive part of the whole inversion exercise; hence the best or nearly the best, parameter improvement possible must be achieved for each optimization iteration in order that fewer overall iterations are required in the whole inversion process. If a few values of λ must be tested to maximize the improvement realised for each iteration, then it is worth the effort. Fortunately it has been found that, if the initial λ and its adjustment factor are well chosen, most iterations require the use of only one or two λ's so that little time needs to be spent in this kind of experimentation.

The second problem, that of the step size, is solved in the following manner. Once a λ has been chosen and an N_m calculated and substituted into (19), the latter equation is solved for Δp. However because we have used N_m rather than N. Δp now indicates only the direction of parameter change, the actual size of the change being $\beta\Delta p$, where ,β is a factor which must be calculated such that $\beta\Delta p$ provides the maximum possible reduction in o of any parameter changes that take place in the direction of Δp. It has been shown by Carrerra and Neuman (1986b) that ,β can be calculated as

$$\beta_i = \frac{(hm - h_i)^t W \gamma_i}{\gamma_i^t W \gamma_i} \qquad \ldots. (21)$$

for the i+1th iteration, where

$$\gamma_i = \partial h_{i+1} / \partial \beta_i \qquad \ldots. (22)$$

In MODINV, the γ vector is calculated using finite (forward or central, as appropriate) differences.

2.4 MEASURED HEAD STANDARD DEVIATIONS AND THE REFERENCE VARIANCE

In the above discussion, no assumptions were made concerning the probability distributions of head observations or (transformed) parameter values for model sub-areas. Our sole criterion for deciding on a set of model parameter values was that the sum of the weighted squared differences between model and measured heads be minimized using this set. Hence we arrived at a set of parameter values for which the fit between model and observed heads is optimum in the weighted least squares sense.

If we make the assumption that both the head measurements and the individual parameter values (those values taken by parameter types within their constant-parameter zones) are normally distributed, then we can say something quantitative about the level of uncertainty pertaining to these estimated parameters values. It is to this topic that we now turn. Note that, in the discussion that follows, if a parameter value has been mathematically transformed so that its log or logistic transformation is estimated, then the following discussion is applicable to the transformed parameter value rather than the parameter itself; in what follows, "parameter value" will refer to whichever of these is being estimated.

At first it may seem that the idea of observed heads being subject to a probability distribution is fallacious because they can normally be determined to the nearest centimetre at least. This is certainly correct, but when you come to calibrate your model it is likely that you will have to accept discrepancies between model and observed heads that are much greater than this. These discrepancies are attributable to the fact that head levels, as actually measured, are subject to small-scale random spatial variations superimposed on regional head variations because of the presence of aquifer spatial inhomogeneity. The actual aquifer recharge, transmissivity etc. distribution is far more complex than our model has the power to replicate, and probably far more complex than we have the ability, or inclination, to measure. Our model seeks to reproduce the first-order or major determinators of groundwater flow as they are expressed in the definition of constant parameter value sub-areas within the model. Second-order earth physical property variations which are superimposed on these major earth property subdivisions are not modelled; rather we are content to acknowledge their existence by noting that every head measurement is subject to both a deterministic effect (which we attempt to predict using the model), and a random effect (attributable to the fact that the model is a simplification of reality). of course if it becomes apparent, through running MODINV with a given aquifer physical property zonation scheme, that the random head variations are excessively large, we may be inclined to add additional sub-areas to our model, thus needing to estimate a greater number of parameter values. Or we may adjust sub-area boundaries. But we acknowledge that we will never remove these random head variations entirely, because we will never have a perfect fit between model and reality.

Because it is thus a stochastic variable, a complete representation of each head measurement must include both the measured value itself, and a quantitative description of the probability function from which this measurement was taken. This description is simplified if we assume a normal probability distribution for the heads. In this case, our vector of observed head values, h_m, can be considered as a collection of sample values of random variables whose mean is estimated at each sample point and observation time as the best-fit-model head at that point and time. The weight matrix is proportional to the inverse of the measured heads covariance matrix, ie.

$$W = \sigma_o^2 \, V_h^{-1} \qquad \ldots . \quad (23)$$

where V_h is the measured head covariance matrix and σ_o^2 is a proportionality constant, referred to as the reference variance. When head measurements are made in the field the latter's value is unknown. However it is determined as part of the least squares estimation process; see below.

Because W is a diagonal matrix, so too is V_h. This means that we assume that the head measurements are uncorrelated both in space and time, ie. head measurement uncertainties at any one bore at any one time have nothing to do with those at another bore at another time. This assumption is not as obviously true as it first sounds, given the origin of measured head random variations as discussed above. If an unaccounted-for transmissivity inhomogeneity, for example, places a lower limit on the squared model minus measured head sum, then its effects may be apparent on any bores that are in, or close to, the heterogeneity; head "errors" are thus correlated for all bores affected by the some inhomogeneity. Also, if we are carrying out a parameter estimation procedure using head measurements taken at a number of times, then it is likely that heads at a particular bore will be over- or under-estimated by the model not in a completely random fashion, but for a number of sampling times in a row (see Carrerra and Neuman, 1986a). However the MODINV algorithm does not try to incorporate such spatial or temporal measured head correlation into the inversion process, both for reasons of simplicity and because the degree of spatial and temporal correlation is difficult to estimate and depends on the final model, which it is the estimation algorithm's purpose to determine. The weight matrix, then, being proportional to the inverse of the measured head covariance matrix for a set of uncorrelated heads, is a diagonal matrix. A diagonal element is large if the uncertainty level pertaining to the corresponding head measurement is considered to be small and vice versa. If, for an evenly distributed set of measurement bores, you assign smaller weights to head measurements in certain parts of your model, you are asking MODINV to give greater importance to the matching of model and observed heads over other parts of the model, presumably because the heads in the area with low weights are subject to greater uncertainty, probably because of greater aquifer heterogeneity there.

Any diagonal term of a covariance matrix expresses the variance of the pertinent parameter value; the variance is the square of the standard deviation. By allocating relative weights to your measured heads, you are, in reality, allocating relative variances and hence relative standard deviations. The absolute variances for these heads depends on how good a fit you end up achieving between model and measured heads. On the assumption that your model (including its constant-parameter sub-areas as defined by you) is correct (an assumption which you should always treat with suspicion), a good overall fit between modelled and measured heads indicates that the head measurement standard derivations must be small. It can be shown (eg. Mikhail (1976, p288) that an unbiased estimate for σ_o^2 is given by

$$\sigma_o^2 = [(h - h_m)^t W(h - h_m)] /r = \Phi / r \qquad \ldots. (24)$$

where **h** is the vector of optimized model heads, h_m is the vector of measured heads and r is the redundancy. The latter is defined as the number of observations minus the number of parameter values for which estimates are required, ie.

$$r = M - N \quad \ldots. (25)$$

with M and N defined earlier in Section 3.

Once parameter values have been optimized, given your set of parameter zonation sub-area boundaries, σ_o^2 can be calculated from (24); you can then calculate the standard deviation of individual head measurements by multiplying the inverse of each head measurement weight by the newly-determined reference variance (equation 23) and taking the square root. If you arrive at a figure that you consider too high, you can change the model by, for example, maintaining the same number of parameters and shifting 9 parameter zonation boundaries, or by adding some extra parameter constant-value zones (so that "random" variations responsible for the unsatisfactorily high head measurement variances now become incorporated into the model). However you should beware of trying to use a model with too many parameters as computing times for MODINV rise linearly with the number of parameter values for which an estimate is required. Also, the more parameter values there are, the more likely are some combinations of values to be highly correlated. This means that you may not end up with a model that is any better (in terms of its ability to predict water levels over the model area) than one parameterized with fewer variables because this high degree of parameter value correlation will be reflected in high parameter value variances (see next section). Also, convergence problems and numerical instability may raise their ugly heads. Carrerra and Neuman (1986c) provide a good discussion on complexity in aquifer parameterization, to which you are referred for more details; in general, simpler is better.

In the MODINV algorithm, the reference variance is calculated after each parameter upgrade. In PREINV you are asked to provide a reference variance which, if achieved, will cause optimization to be terminated. If you indicate, using this reference variance, an overall measured minus model head discrepancy that you can tolerate, further optimization can be forestalled once (and if) this tolerable discrepancy has been achieved.

2.5 THE PARAMETER VALUE COVARIANCE MATRIX

It can be shown that the parameter value covariance matrix can be estimated by

$$V_p = \sigma_o^2 (J^t W J)^{-1} \quad \ldots. (26)$$

The diagonal elements of this matrix are the variances (id the squared standard deviations) of the individual parameter values while the off-diagonals elements are the covariances between respective parameter value pairs; these latter are indicative of how highly correlated two different parameter values are. It is important to note that the derivation of (26) relies on two assumptions, neither of which are strictly correct in the groundwater modelling context.

The first assumption is that heads and parameter values are normally distributed. While this assumption may be more closely adhered to if parameters are transformed, it will never be completely correct. As with measured heads, the concept of parameter values as

random variables relates to the effects of aquifer inhomogeneities superimposed on the simplifications inherent in the process of model construction.

The second assumption is that model head and parameter value variations are linearly related in the manner described by equation (8). As previously discussed, this linear relationship is only approximately correct, with the approximation worsening for larger parameter and head variations about specific head and parameter values, the latter being related to each other through the nonlinear relationship of equation (4). Hence if the covariance matrix indicates that a parameter standard deviation is large, which is the same as saying that the parameter value could vary widely and still be used in the calibrated model, then the exact value of that standard deviation, as provided by the covariance matrix, cannot be correct because such a wide parameter variation will put it outside the range of the linearity assumption.

Nevertheless, the parameter covariance matrix is one of the most useful pieces of information to come out of the inversion process. Its principle role is that of an indicator of how well your borehole head measurements are able to define aquifer properties (including recharge or EVT if you are estimating either). For while your model heads may be well matched to the measured heads, (the reference variance may be satisfactory), some parameter value standard deviations may still be large. This indicates that, as mentioned above, these parameter values can be made to vary by large amounts with little effect on the model heads at the boresites. If this applies to a single parameter value, it will have a high variance and will be uncorrelated with other parameter values. If, however, two or more parameter values can be simultaneously varied in a certain relationship to each other while causing minimal change to the model heads at the observation bores over time, then these parameter values will each have a high standard deviation and the covariance between pairs of such parameters, as indicated by the pertinent off-diagonal elements of the covariance matrix, will also be large. This indicates high parameter value correlation or, to put it another way, a high degree of stochastic dependence between the pertinent parameter values. If you were to run MODINV again while holding one (or more) of a set of highly- correlated parameter values fixed, the standard deviations of the other members of the set may then be small because the definition of the model now includes the first member(s) of the set as fixed. As the first parameter now has no opportunity to vary in harmony with the others, for minimal resultant head variation at the observation boresites, the standard deviations of the others cannot be as large.

Thus the parameter covariance matrix tells you something about your model that the goodness of fit between model and observed heads cannot tell you. For example, if the density of observation bores is low or zero over a certain part of the aquifer, parameter values estimated in that area may not be well defined, and this will be indicated in the covariance matrix. While your model may appear to be well calibrated because the model replicates observed heads at the existing observation bores with a good degree of accuracy, its capacity to predict water levels over other parts of the aquifer may be highly suspect if the calculation of these latter heads relies on parameter values which are, locally, ill defined. By varying highly correlated parameters in directions defined by the parameter covariance

matrix eigen vectors, you can test the effect of simultaneous parameter variation on modelled heads at both observation bores and elsewhere. If the head change at the observation bores is small, but is great at other places that may be of interest to you, then you may not have enough information to parameterize your aquifer if one of the model's tasks is to predict water levels in this other area with any accuracy.

Correlation between pairs of parameters can be displayed as a correlation coefficient matrix. If σ_{ij} is an element of the parameter value covariance matrix, then the corresponding element of the correlation coefficient matrix is given by

$$\rho_{ij} = \sigma_{ij} / [\sigma_i^2 \sigma_j^2]^{1/2} \quad \ldots. (27)$$

where σ_i^2 and σ_j^2 are the ith and jth diagonal elements of the covariance matrix; obviously the correlation coefficient matrix has diagonal elements of unity. The correlation between different pairs of parameter values is then readily apparent from the pertinent off-diagonal elements, a high degree of correlation between parameter value pairs being indicated by a correlation coefficient close to 1 or -1 (a correlation coefficient cannot be higher than 1 or less than -1).

Another method of displaying the wealth of information that is available in the parameter value covariance matrix, is to display its eigenvalues and eigenvectors. The latter define the directions of the axes of the parameter confidence "ellipse" (actually, it is only an ellipse in two dimensions, ie. if only two parameter values are estimated), whereas the square roots of the eigenvalues are the magnitudes of the semiaxes of the parameter confidence ellipse. If all eigenvectors have only one component, then the axes of the confidence ellipse will lie along the parameter value axes, and parameter values are thus all uncorrelated. In the more usual case, the degree of correlation between different parameter value estimates can be obtained by examining the components of the eigenvectors. If, for example, the i'th, j'th and k'th components of a particular normalized eigenvector are much larger than the other components of that vector, and the eigenvalue corresponding to that eigenvector is larger than most of the eigenvalues corresponding to the other covariance matrix eigenvectors, then this indicates that the *linear combination* of the i'th, j'th and k'th parameter values is better determined than are the individual values; the coefficients of this linear combination correspond to the respective eigenvector elements. See Carrera and Neuman (1986c) for a further discussion of how the covariance matrix eigenvectors and eigenvalues can be used to understand the power and limitations that your measurement set possesses in parameterizing your model.

2.6 FINAL POINTS

As described above, MODINV employs the Gauss-Marquardt method to minimize the sum of the weighted squared head differences between measured and model heads. This "objective function" is the same function that is minimized in the Maximum Likelihood method of parameter estimation. In fact, in the present case, the only difference between the two methods is in the estimated value of the reference variance; in the Maximum Likelihood method, the denominator in equation (24) is M, the number of observations, rather than

M-N, the redundancy. When M is much higher than N the two estimates are close; however σ_o^2 as provided by (24) has the advantage that it is an unbiased estimate.

Some worrying questions are (i) whether Φ, as defined by (6) has a single minimum, and (ii) if so, whether MODINV will always find it. Unfortunately there is no single answer to both these questions that applies to all modelling situations and all parameter types and combinations of parameter types for which estimated values are being sought. Experience in using MODINV has demonstrated that Φ can converge to a local minimum in some cases, for which parameter values are far from optimum. In other cases it will not converge at all. However both of these phenomena are more likely to occur when many parameter values of different types are being simultaneously estimated; in such cases you can often dramatically improve MODINV's performance simply by holding a few key parameter values fixed, or by using fewer parameter values in a less complex areal distribution. Failure to converge to a global minimum is often a signal that parameter value correlations are high and that you are consequently asking too much of your data in trying to resolve individual values. Hence, not only will a simpler model improve MODINV's performance, but it may be a more realistic representation of the true information content of your measurements. As such, predictions made with the calibrated model will tend to be "conservative" in that the possibility of predicting spurious local head variations, resulting from the presence of local, poorly defined parameter values, will be reduced.

3.0 MODINV PROCESSING STEPS

Fig.1 shows a simplified flow chart of the MODINV algorithm; see Table 2 for a list of symbols used in Fig.1. You can tell what part of its algorithm MODINV is executing at any time by inspecting its continually-updated run record which is written to file MODINV.PRN. If you are running MODINV from the terminal and have requested a screen display of computation progress, then additions to file MODINV.PRN are also sent to the screen, allowing you to monitor the progress of the optimization process. If you are running MODINV as a batch job, screen display is not available. However some systems will allow you to read MODINV.PRN, even though it is concurrently held open by MODINV; other systems allow you to read a batch job log file (while the batch job is executing) containing information that would have been sent to the screen if the job were run from the terminal. If either case, periodic inspection of the MODINV output allows you to monitor MODINV run progress.

Table-2 Symbols used in Fig.1

Φ	objective function
$\nabla\Phi$	gradient of the objective function
J	Jacobian matrix
N	normal matrix
I	identity matrix
Σ	covariance matrix
λ	Marquardt lambda
p_i	parameter value estimates for i'th iteration
Δp_i	parameter optimization direction vector for i'th iteration
β	fraction of Δp by which to obtain p_{i+1} from p_i
γ	derivative vector of model heads w.r.t. β
m	number of parameter values requiring optimization
i	optimization iteration number

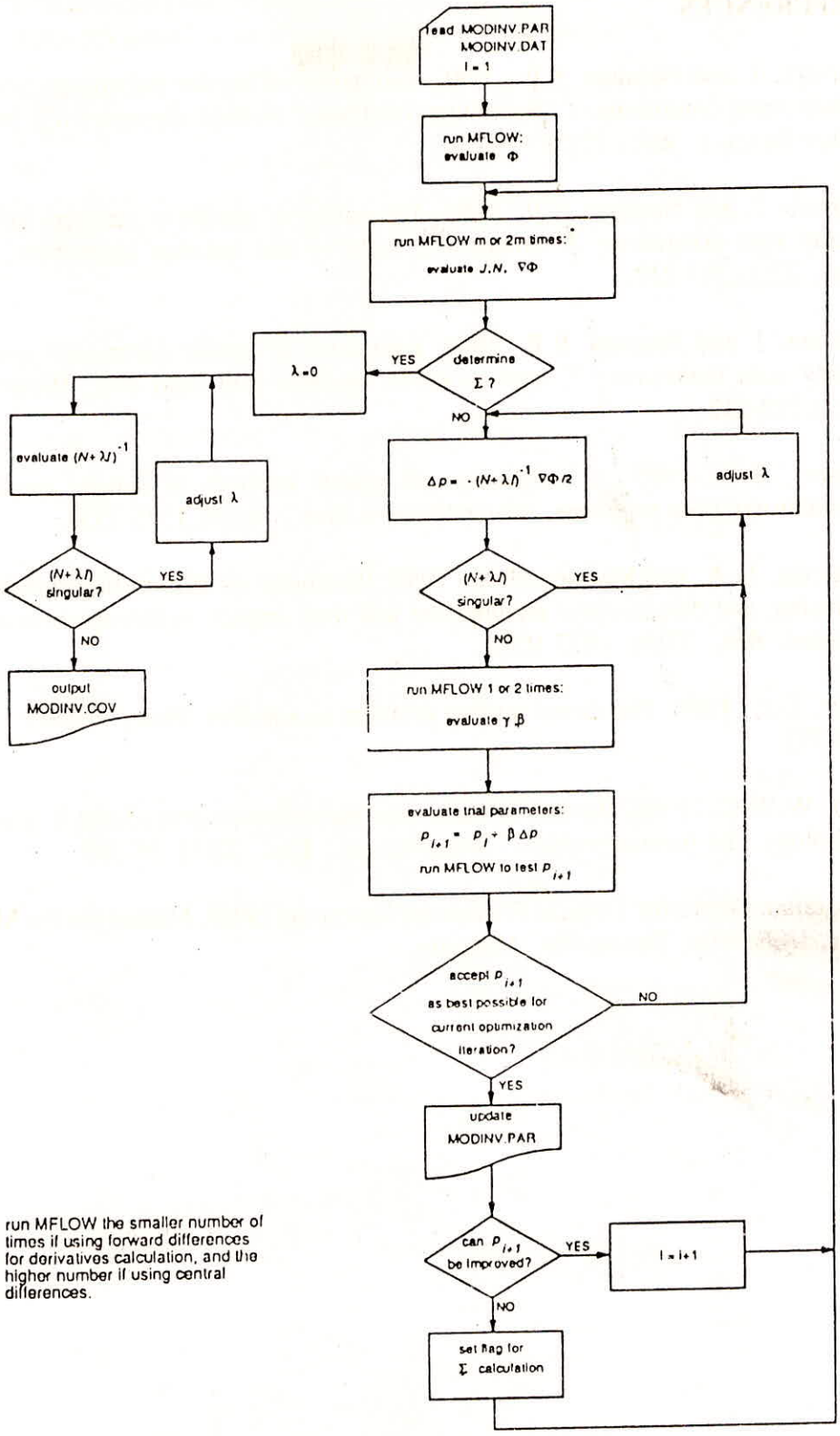

Fig.1 Simplified flow chart of the MODINV algorithm.

REFERENCES

Carrera, J. and Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state conditions: 1. maximum likelihood method incorporating prior information. Water Resourc. Res., 22(2),199-210.

Carrera, J. and Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state conditions; 2. uniqueness, stability and solution algorithms. Water Resourc. Res., 22(2),211-227.

Carrera, J. and Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state conditions: 3. application to synthetic and field data. Water Resourc. Res., 22(2),228-242.

Cooley, R.L., 1985. A comparison of several methods of solving nonlinear regression groundwater flow problems. Water Resourc. Res., 21(10),1525-1538.

Loaiciga, H.A. and Marino, M.A., 1987. Parameter estimation in groundwater: classical, Bayesian, and deterministic assumptions and their impact on management policies. Waler Resourc. Res., 23(6), 1027-1035.

Weir, G.J., 1989. The direct inverse problem in aquifers. Water Resourc. Res., 25(4), 749-753.

Yeh, W.W-G., 1986. Review of parameter identification procedures in groundwater hydrology: the inverse problem. WaterResourc. Res., 22(2), 95-108.

Ausralian Centre for Tropical Freshwater Research (1990), Manual for the MODINV, James Cook University, Townsville, Australia.