

# Influence of the regularization weighting matrix on parameter estimates

L. R. Bentley

*Department of Geology and Geophysics, University of Calgary, Calgary, Alberta T2N 1N4, Canada*

(Received 13 September 1995; revise received 5 April 1996; accepted 12 June 1996)

Optimization techniques are often used in parameter estimation. An objective function can be formed with a measurement term and regularization term. The measurement term is formed from the weighted sum of squares of the difference between head measurements and their computed values. The regularization term is formed by the weighted sum of squares of departures from the original parameter estimates. Most often, a diagonal regularization weighting matrix is used. A diagonal matrix corresponds to an assumption of no spatial correlation between the errors in the original parameter estimates, but the errors will most often be correlated. Consequently, non-diagonal weighting matrices should be used. Numerical experiments have been performed to test the influence of non-diagonal regularization weighting matrices on parameter estimates. It is demonstrated that using the correct structure for the weighting matrix can improve transmissivity estimates. The transmissivity fields computed from the non-diagonal weighting matrices are closer in statistical structure and more accurate than those computed with diagonal weighting matrices, and the connectivity of the extreme values is more accurately imaged. The use of non-diagonal weighting matrices can improve the computed flow field, and, therefore, the accuracy of transport simulations. The results indicate that using the correct weighting can improve transmissivity estimates more effectively than adding many new observation points. Copyright © 1996 Elsevier Science Ltd

## INTRODUCTION

Groundwater flow simulators are routinely used in water supply, dewatering and contamination studies. The effective use of groundwater flow simulators is limited by the modeler's ability to establish realistic, spatially detailed estimates of the model parameters. Hydrogeologic parameters are inferred from aquifer tests, laboratory tests, or by comparison with values obtained in other areas. The amount and the spatial distribution of test data are always limited, and the test results contain errors. Consequently, the values and the spatial distribution of the parameters are never known with complete certainty (for a more complete discussion see ref. 7). In general, when a numerical model is parameterized with the initial estimates of the parameters, the calculated hydraulic heads will fail to match the field observations, because the initial estimates are not exact.

Consequently, a calibration step is required when modeling a flow system. In the calibration step, the parameters are adjusted until the values of computed

hydraulic heads are sufficiently close to a set of field observations. Most often, calibrations are performed by trial and error. Parameter identification techniques based on statistical methods or optimization techniques have also been used to improve the imprecise original parameter estimates (for a recent review see ref. 11). The fact that these techniques are not routinely used is evidence that further work is needed.

The two dimensional, isotropic, steady state groundwater flow equation is,<sup>14</sup>

$$-\frac{\partial}{\partial x} \left( T \frac{\partial h}{\partial x} \right) - \frac{\partial}{\partial y} \left( T \frac{\partial h}{\partial y} \right) + B(h - h_r) - I_f - \sum_{p=1}^{N_Q} Q_p \delta(\bar{x} - \bar{x}_p) = 0 \quad (1)$$

where  $h$  is hydraulic head,  $T$  is transmissivity,  $B$  and  $h_r$  are leakage parameters,  $I_f$  is infiltration,  $Q_p$  is the pumping rate for well  $p$  (positive in),  $N_Q$  is the number of pumping wells and  $\delta(\bar{x} - \bar{x}_p)$  is the Dirac delta function. In addition, appropriate boundary conditions

must be specified to complete the description of the problem. The application of discretization theory leads to a system of algebraic equations that approximates equation (1). These equations can be represented as a matrix expression,

$$\mathbf{A}\mathbf{H} = \mathbf{f} \quad (2)$$

where  $\mathbf{A}$  is a matrix that is a function of the discretized system parameters such as  $T$ ,  $\mathbf{H}$  is a vector of head values at the nodes and  $\mathbf{f}$  is a forcing vector which contains boundary conditions and system forcings.

One approach to parameter estimation is to solve the following optimization problem,<sup>4</sup>

$$\begin{aligned} & \text{MIN} \\ & \text{w.r.t. } \mathbf{H} : \mathbf{P} \quad (\hat{\mathbf{h}}(\mathbf{H}) - \mathbf{h}^*)^t \mathbf{W}_M (\hat{\mathbf{h}}(\mathbf{H}) - \mathbf{h}^*) \\ & \quad + (\mathbf{P} - \mathbf{P}^*)^t \mathbf{W}_P (\mathbf{P} - \mathbf{P}^*), \\ & \quad \text{Subject To } \mathbf{R}_G = 0 \end{aligned} \quad (3)$$

where  $\hat{\mathbf{h}}(\mathbf{H})$  is a vector of trial function values at the measurement points,  $\mathbf{H}$  is a vector of hydraulic heads at the nodes,  $\mathbf{h}^*$  is a vector of the measurement values,  $\mathbf{W}_M$  and  $\mathbf{W}_P$  are weighting matrices,  $\mathbf{P}$  is a vector of parameters,  $\mathbf{P}^*$  is a vector of original estimates and  $\mathbf{R}_G = \mathbf{A}\mathbf{H} - \mathbf{f}$  is a vector containing the discretized governing equations. The objective is to find a set of heads that is close to the field measured values (first term in equation (3)) and a set of parameters that is close to the original estimates (second term in equation (3)), subject to the condition that two sets honor the discretized governing equations. The parameter set  $\mathbf{P}$  can include any specified value such as infiltration rates or boundary condition values, but in the following, only the natural log of the transmissivity will be estimated.

In this paper, the penalty method<sup>4</sup> is used to solve equation (3), but the discussion which follows is generally applicable to other formulations such as those derived from maximum likelihood estimation,<sup>7</sup> or regression theory.<sup>9</sup> Bentley<sup>4,5</sup> discusses the relationship between penalty method and those of Cooley<sup>9</sup> and Carrera and Neuman.<sup>7</sup>

Calculating parameters from measurements of state variables is an inverse problem. It is well known that solutions of inverse problems may be plagued by problems of identifiability, nonuniqueness and instability. The problems arise because the dimension of the parameter set is, in essence, infinite, and the number of observations is limited. If we restrict ourselves to the estimation of transmissivity, the dimension of the parameter set is the same as the number of elements or blocks in the computational mesh which is still typically many times larger than the number of head observations. In order to further reduce the dimensionality of the system, most reported parameter estimation results rely on zoning parameters into large areas of equal value. Zoning is used to reduce the number of decision variables and to avoid problems associated with

over-parameterization.<sup>11</sup> Zoning is well suited for modeling abrupt geological contacts. However, if the changes in transmissivity are not abrupt or the locations are not well known, its use is problematic. In effect, zoning establishes absolute correlation of the original transmissivity estimate errors for all elements within a zone. By necessity, zoning destroys the description of the smaller scale heterogeneities. Zoning is a fundamental model decision which affects the spatial distribution of the parameters. Generally, zoning takes place before the optimization step. It is difficult to determine whether the spatial extent of the zones is appropriately defined or not.

The second term in equation (3) is known as the regularization or prior information term. The purpose of the term is to penalize departures from our initial estimates of the parameters based on the premise that they contain prior information about the magnitude and distribution of the parameters derived from tests or geological insight. The effect of the regularization term is to reduce the problems associated with indentifiability, nonuniqueness and stability. If every parameter has a regularization residual associated with it, then we can estimate a greater number of parameters than we have measurement residuals. Without the regularization terms, the problem would be under-constrained if an attempt was made to estimate more parameters than number of measurements.

The solution of equation (3) requires the specification of the weighting matrices  $\mathbf{W}_M$  and  $\mathbf{W}_P$ . According to Bard,<sup>3</sup> 'Although we cannot prove optimal properties in the general case (non-normal distributions with non-linear models), it is still reasonable, and approximately optimal, to use weights which are the elements of the inverse of the covariance matrix'. Accepting this philosophy and assuming that the measurement errors are uncorrelated with the original parameter estimate errors, the weighting matrices  $\mathbf{W}_M$  and  $\mathbf{W}_P$  should be the inverse of the covariance matrix for errors in the head measurement and the inverse of the covariance matrix for the errors in the original parameter estimates, respectively. In practice, we never know the covariance of the errors, and we must estimate them. In this paper, the parameter of interest is the natural log-transmissivity ( $\ln -T$ ). Since  $\ln -T$  is generally correlated in space, it is reasonable to assume that errors in the original estimates of  $\ln -T$  will be correlated as well, and, in general, the weighting matrix  $\mathbf{W}_P$  will be non-diagonal. If the original  $\ln -T$  estimates are derived using kriging, then they are definitely correlated.<sup>8</sup> The weighting matrices have two basic effects. Original parameter estimates and head measurements that are reliable will have associated small error variance values. Since inverting a small variance value leads to a large value in the weighting matrix, the more reliable data will be more strongly weighted than the less reliable data. If the weighting matrix is not diagonal, it will also have a

spatial filtering effect. Most reported results have been obtained with diagonal regularization weighting matrices. However, results presented by Bentley<sup>5,6</sup> have demonstrated that the use of non-diagonal weighting matrices can improve  $\ln-T$  estimates. The purpose of the following discussion is to explore the effect of non-diagonal regularization weighting matrices on inverse solutions.

**TEST PROBLEM**

A transmissivity field derived from the simulated Walker Lake transmissivity field<sup>10</sup> is used to investigate the influence of non-diagonal regularization weighting matrices on parameter estimation solutions. The Walker-Lake field has realistic features such as spatial connectivity of extreme values and nested scales of heterogeneity. The original field was subdivided into a  $300 \times 260$  grid with cell dimension of one meter by one meter. The  $\ln-T$  field has a maximum integral range of correlation of 34 m and a minimum integral range of 15 m with the direction of maximum correlation being N14°W.

A new grid of  $15 \times 13$  elements with cell dimensions of  $20 \times 20$  m was constructed to represent the computational grid that would be used in test simulations. The block transmissivities of the new grid are the geometric mean of the original grid transmissivities contained within each upscaled block. The new grid has an average  $\ln-T$  of 4.01 and a  $\ln-T$  variance of 0.629. The  $\ln-T$  values are plotted in Fig. 1. Note the high transmissivity

channel that runs southward from the northern boundary and then bends towards the eastern boundary. Areas of lower transmissivity separate the channel from the relatively high transmissivities in the northeast and southwest corners.

A flow simulation was run with constant head boundary conditions of  $H = 3$  meters and  $H = 0$  meters for the northern and the southern boundaries, respectively. The eastern and western boundaries are no-flow. The boundary conditions produce a flow field of approximately  $0.54 \text{ m}^2/\text{day}$ . The head field undulates around the linear decline in head with maximum departures on the order of 0.4 m (Fig. 2). It would be most realistic to use the head field computed from the forward solution of the original one meter by one meter grid to generate synthetic measurement data, but that would introduce issues of upscaling which would confound the study of the effects of the weighting matrices. Consequently, simulated measurements were created by taking the forward solution head values from the upscaled grid and corrupting them with uncorrelated Gaussian noise of standard deviation 0.01. Since the measurement errors are uncorrelated, the weighting matrix  $\mathbf{W}_M = 10^4 \mathbf{I}$  is used for all of the following calculations, where  $\mathbf{I}$  is the identity matrix.

In the following test problems we will use equation (3) to estimate the  $\ln-T$  of all 195 elements in the computational grid. Consequently, the vector  $\mathbf{P}$  in equation (3) consists of one value of  $\ln-T$  for each individual element. The mean of  $\ln-T$  is used for the initial estimate of  $\ln-T$ ,  $Y^* = 4.0$ , for all 195 elements in all of the test problems. A histogram of the errors in  $Y^*$

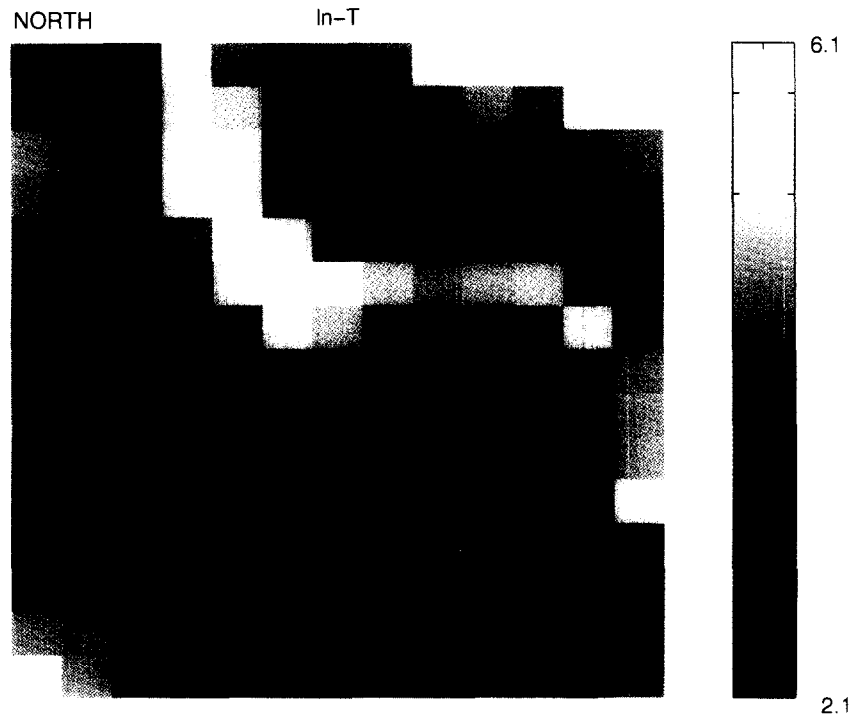


Fig. 1. Upscaled Walker Lake  $\ln-T$  field.

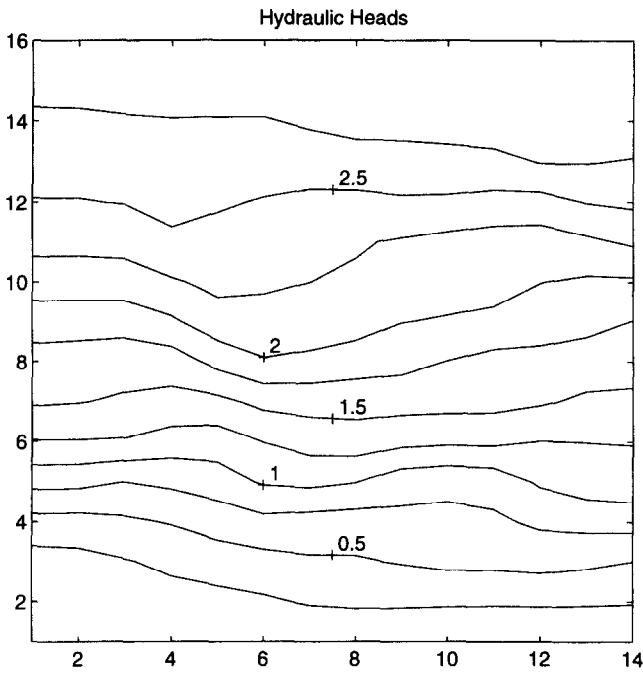


Fig. 2. Heads (m) computed from the  $\ln-T$  field of Fig. 1.

is shown in Fig. 3(a). The mean of the errors is approximately zero so that they are unbiased. The variance in the errors is 0.630. The spatial correlation of the errors as represented by the covariance is shown in Fig. 3(b). The covariance of the errors shows little dependence on direction, so omnidirectional covariance models are used. The spatial correlation of the errors is well represented by a Gaussian covariance model<sup>12</sup> with correlation length  $\lambda_c = 45$  m.

$$C(h) = 0.63 \exp \left[ \frac{-3 * h^2}{\lambda_c^2} \right] \tag{4}$$

where  $h$  is the lag distance between element centers. In the following, we will also be using exponential covariance models<sup>12</sup> to generate weighting matrices,

$$C(h) = 0.63 \exp \left[ \frac{-3 * h}{\lambda_c} \right]. \tag{5}$$

Referring to Fig. 3(b), we see that the exponential model of the error covariance is only a crude approximation to the true structure. In the case of  $\lambda_c = 50$  m, the correlation of the errors is greater than the model's value for the near lags, and for  $\lambda_c = 90$  m, the model's values are greater than the correlation of the errors for the far lags.

The common procedure of using diagonal weighting matrices is equivalent to a *de facto* assumption of no spatial correlation between the errors. If the errors are assumed to have no spatial correlation, then the covariance matrix is,

$$C_p = 0.63I \tag{6}$$

where  $I$  is the identity matrix.

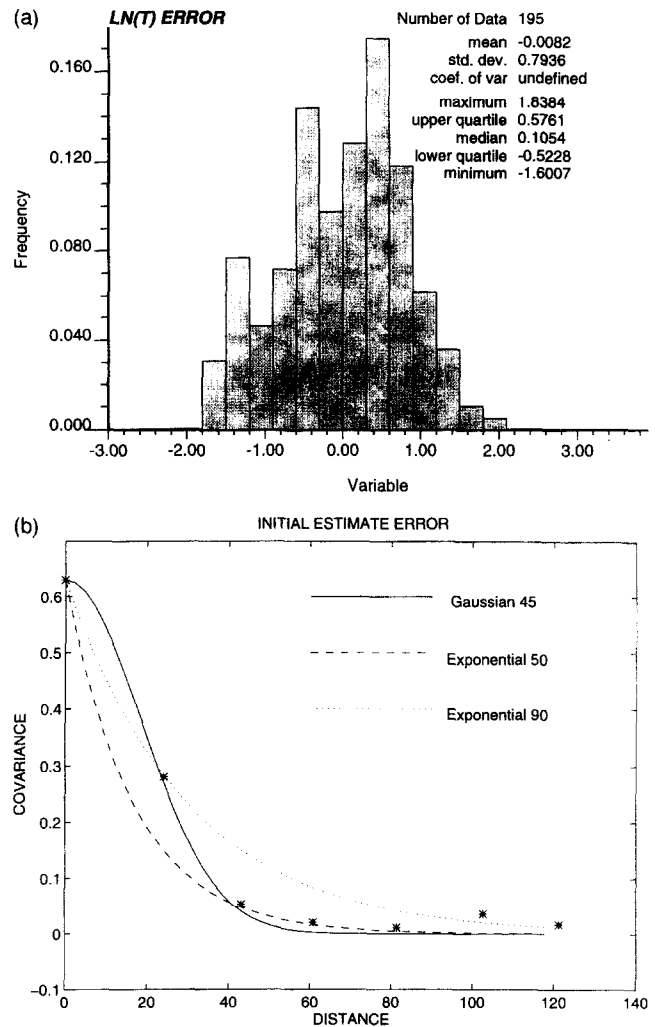


Fig. 3. (a) Histogram of the errors in the original estimates of  $\ln-T$ ; (b) calculated covariance of the original estimate errors (stars), Gaussian covariance model with  $\lambda_c = 45$  m (solid line), exponential covariance model with  $\lambda_c = 50$  m (dashed line) and exponential covariance model with  $\lambda_c = 90$  m (dotted line).

The following procedure is used to generate the weighting matrix  $W_p$ . For each element pair, we compute the lag distance  $h$  by calculating the distance between the element centers. The matrix that approximates the covariance of the errors is then computed using either equation (4), (5) or (6). Once the covariance matrix has been constructed, the weighting matrix is calculated,

$$W_p = C_p^{-1}. \tag{7}$$

In the following, we will examine the effect of the weighting matrix  $W_p$  on the solution by examining solutions to equation (3) calculated using weighting matrices derived from the covariance models of equations (4), (5) and (6). In addition, the number and location of measurement points will be varied.

## PERFORMANCE MEASURES

In order to compare the solutions which follow, several performance measures and statistical parameters are introduced. In the following  $Y_A$  are the correct or true  $\ln -T$  values,  $Y^*$  are the original estimates of  $\ln -T$  (in all cases 4:0), and  $\hat{Y}$  are the values of  $\ln -T$  computed with the optimization procedure. Similarly,  $H_A$  are the simulated 'true' heads,  $H^*$  are the values of the simulated head measurements ( $H_A$  plus noise), and  $\hat{H}$  are the values of the heads computed using the optimization solution values,  $\hat{Y}$ .

The objective function in equation (3) contains two terms. The first term is the measurement residual term, denoted  $J_M$ , and the second term is the regularization term, denoted  $J_Y$ . Defining  $\hat{J} = J_M + J_Y$  as the value of the objective function at the solution, the variance of the weighted errors can be approximated by (equation (25) of ref. 4),

$$s^2 = \frac{\hat{J}}{NM} \quad (8)$$

where  $NM$  is the number of head measurements. If all of the assumptions included in the statistical model assumed in equation (3) are satisfied, the value of  $s^2$  should be close to 1.0.

The variance in the solution values of  $\ln -T$ ,  $\hat{Y}$ , is computed,

$$\sigma_{\hat{Y}}^2 = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 \quad (9)$$

where  $\bar{Y}$  is the average value of the 195 values of the  $\ln -T$  solution,  $\hat{Y}$ . The variance of the correct  $\ln -T$  field,  $Y_A$  is 0.63 and the variance of the original estimate of  $\ln -T$ ,  $Y^*$  is zero.

The average sum of the squared differences of two variables,  $A$  and  $B$ , is,

$$SS(A - B) = \frac{1}{N} \sum_{i=1}^N (A - B)^2 \quad (10)$$

where  $N$  is the number of terms.

$SS(\hat{Y} - Y_A)$  is a measure of the accuracy of the  $\ln -T$  solution. In this case  $N = 195$ , the number of elements. The original estimates have  $SS(Y^* - Y_A) = 0.63$  and any value of  $SS(\hat{Y} - Y_A)$  which is less than 0.63 represents an improvement on the original estimate.

With each solution, a set of head values,  $\hat{H}$ , are produced. A measure of the accuracy of the solution is  $SS(\hat{H} - H_A)$  and  $N = 224$ , the number of nodes. In addition we can compute the sum of squares of the head differences only at the measurement point locations,  $SS(\hat{H} - H_A)_M$  where  $N = NM$ , the number of measurement points. The values of  $SS(\hat{H} - H_A)_A$  and  $SS(\hat{H} - H_A)_M$  approach zero as the solution accuracy improves. The measurement values were computed by adding random noise of standard deviation 0.01 to the

value of  $H_A$  at the measurement points and, consequently, the value of  $SS(H^* - H_A)$  is approximately  $10^{-4}$ .

Finally, the number of iterations for convergence is tracked as an indicator of the solution difficulty. In the following test problems, calculations were terminated if 200 iterations were completed, and tests with 200 iterations have not necessarily converged to the global minimum, although they may be close.

## RESULTS

The test problem was solved using a series of different regularization weighting matrices,  $W_p$ . As shown in Fig. 3(b), the best fit to the spatial correlation of the errors in the original parameter estimates is a Gaussian model of correlation length  $\lambda_c = 45$  m. In practice the actual structure of the covariance of the errors is never known with certainty, so it is important to test the robustness of the solutions to errors in the estimate of the covariance structure. Consequently, one set of weighting matrices was derived using the Gaussian model, equation (4) with correlation lengths  $\lambda_c = 20, 30, 40, 45$  and 50 m. It is known that covariance matrices constructed using Gaussian models are poorly conditioned,<sup>1</sup> and difficulties arose during attempts to invert matrices with correlation lengths longer than 50 meters. A second set of weighting matrices was derived using the exponential model, equation (5) with correlation lengths  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m. Figure 3(b) contains a comparison of exponential covariance models of correlation lengths 50 and 90 m with the actual values. For scale, the center-center distance between adjacent elements is 20 m, the width of the grid is 260 m and the length is 300 m. The assumption of no spatial correlation between the errors in  $Y^*$  leads to a diagonal weighting matrix of the kind usually used in parameter estimation, and consequently, the problems were also solved with a diagonal weighting matrix derived from equation (6).

The number and location of the measurement points exerts strong control over the solution. In order to assure that the significance of the results is not specific to a single set of measurement point locations, the tests were duplicated for different measurement point locations. Figure 4 shows the locations of the measurement points for four different tests. In test (a), 98 measurement points were regularly spaced. Test (b) results were computed using a set of 49 regularly spaced measurement points, and tests (c) and (d) were computed using two different sets of 49 randomly spaced measurement points.

The results of the tests are presented in Figs 5-11. In all cases, the horizontal scale is logarithmic and ranges from 20 to 200 m, circles are the Gaussian model results, crosses the exponential model results and stars are the diagonal weighting matrix results. Figure 5 presents the

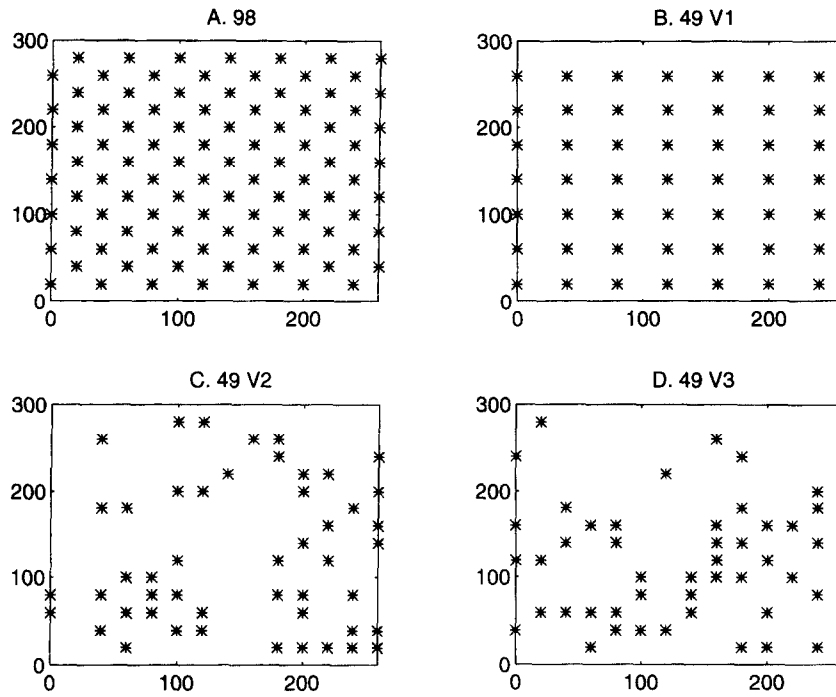


Fig. 4. Measurement point locations. (a) 98 regularly placed points; (b) 49 (version 1) regularly placed measurement points; (c) 49 (version 2) measurement points randomly placed; (d) 49 (version 3) measurement points randomly placed.

variation in the number of iterations required to converge. It is evident that the problem is more easily solved after the addition of off diagonal weighting matrix terms. Most of the benefit has been derived by the time the correlation length has reached 30 m.

Figure 6 presents the variation in  $SS(\hat{Y} - Y_A)$  with weighting matrices.  $SS(\hat{Y} - Y_A)$  is a primary measure of the accuracy of the parameter estimate. In cases (a) and (b), where the measurement points are regularly spaced, the Gaussian model with correlation length 40–45 m

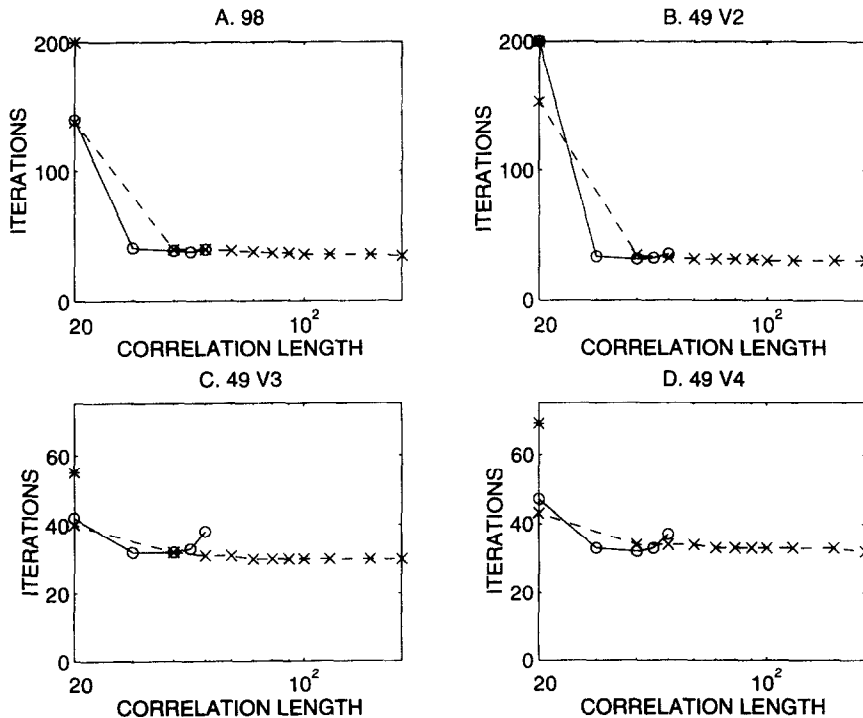
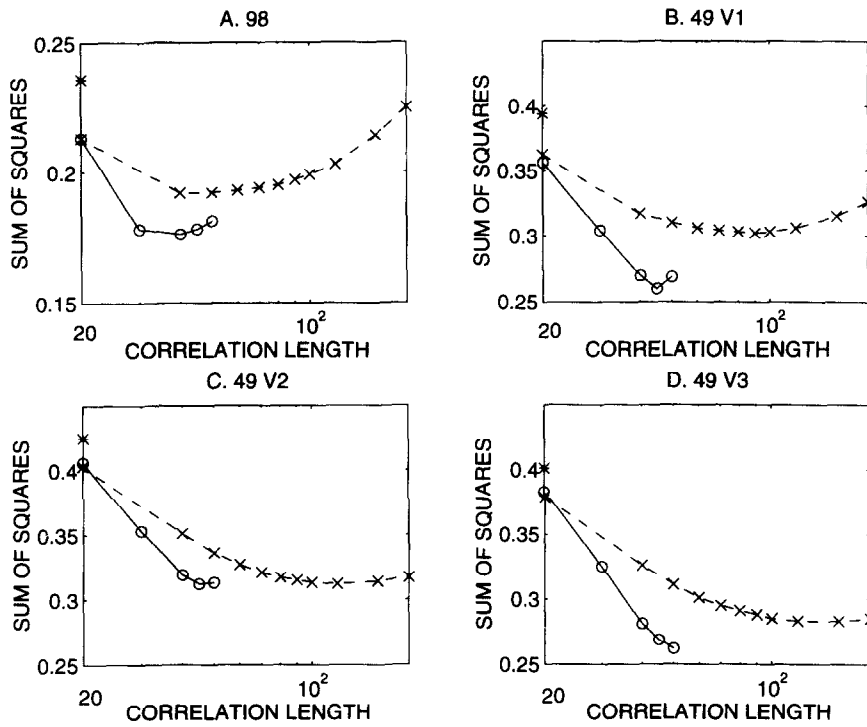
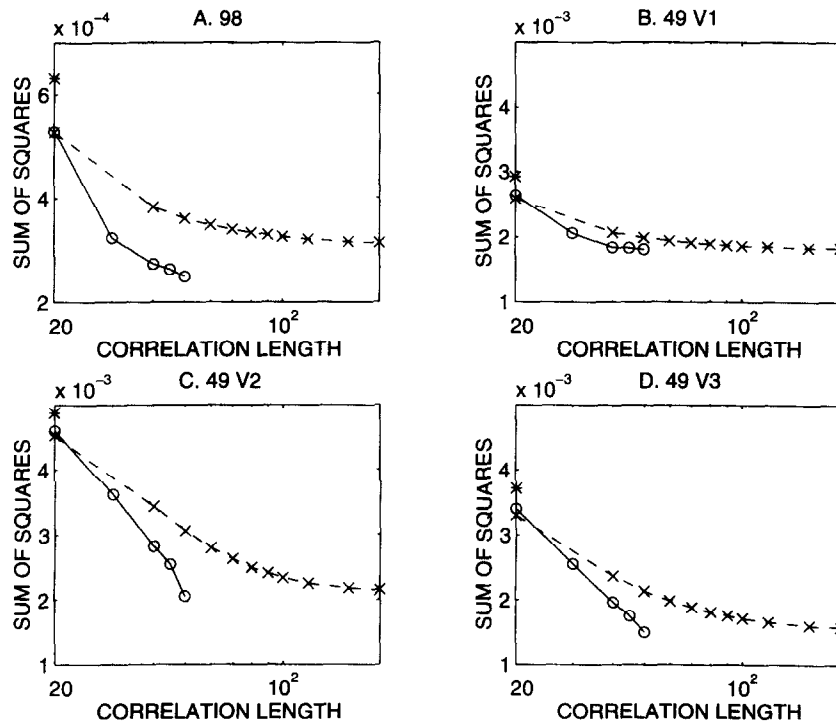


Fig. 5. Number of iterations versus weighting matrix. (a)–(d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.



**Fig. 6.**  $SS(\hat{Y} - Y_A)$  vs weighting matrix. (a), (b), (c) and (d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.



**Fig. 7.**  $SS(\hat{H} - H_A)_A$  vs weighting matrix. (a), (b), (c) and (d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.

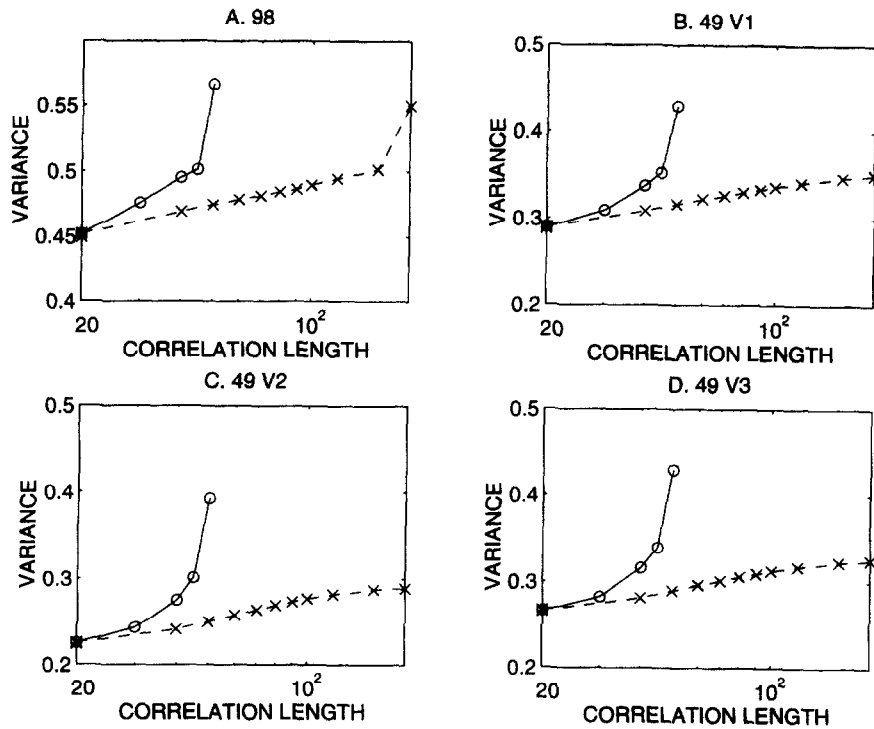


Fig. 8.  $\sigma_y^2$  vs weighting matrix. (a), (b), (c) and (d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.

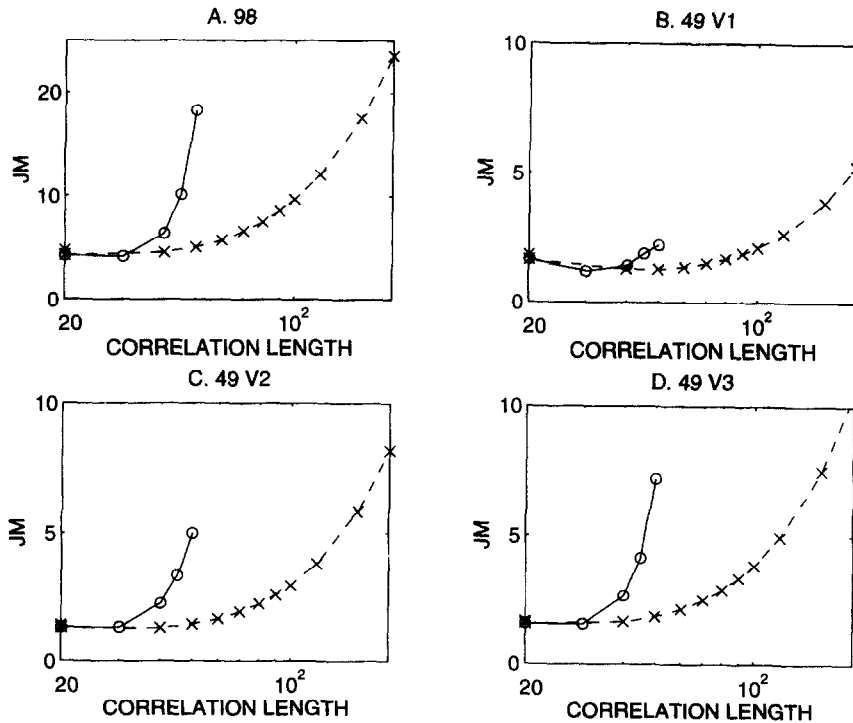
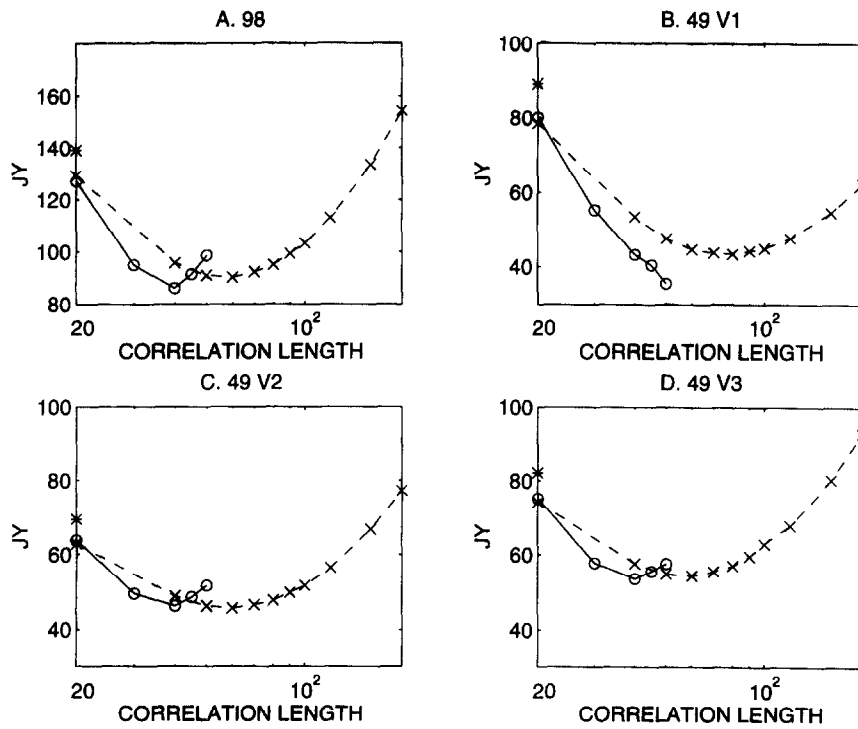
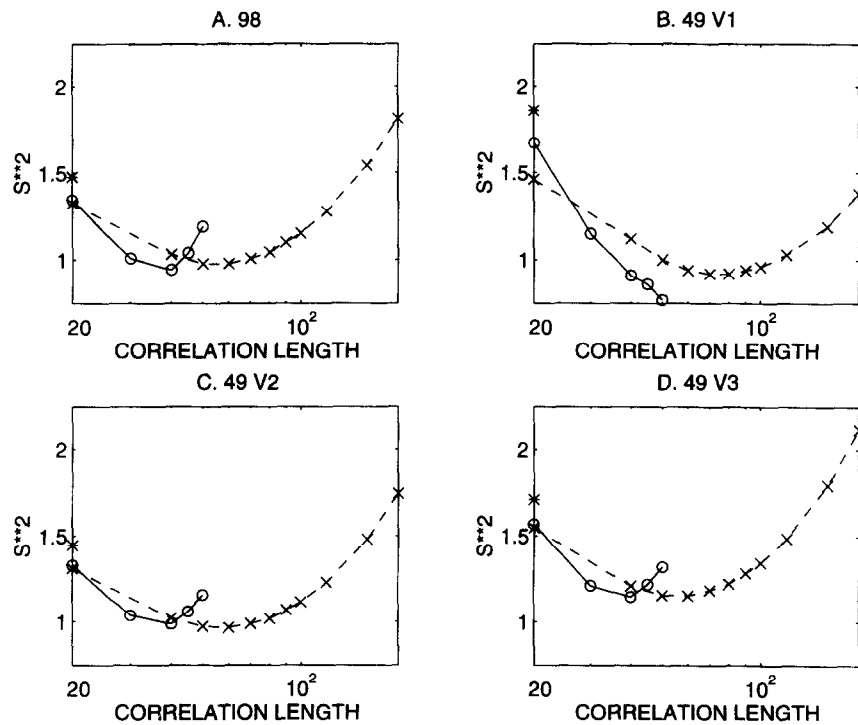


Fig. 9.  $J_M$  vs weighting matrix. (a), (b), (c) and (d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.



**Fig. 10.**  $J_Y$  vs weighting matrix. (a), (b), (c) and (d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.



**Fig. 11.**  $S^2$  vs weighting matrix. (a), (b), (c) and (d) correspond to measurement point locations found in Fig. 4. Circles connected by the solid lines are results for Gaussian covariance models and crosses connected by dashed lines are results from the exponential model. Identity matrix results (stars) are plotted at correlation length 20 m. The Gaussian model has calculations made for  $\lambda_c = 20, 30, 40, 45$  and 50 m and the exponential model at  $\lambda_c = 20, 40, 50, 60, 70, 80, 90, 100, 120, 160$  and 200 m.

provides the best solution, and it is better than the exponential model. The exponential model provides a better solution than the diagonal weighting matrix with the best results being for  $\lambda_c = 40$  and 50 m for case (a) and  $\lambda_c = 90$  m for case (b). Both covariance models contain minimums in  $SS(\hat{Y} - Y_A)$ , but the exponential solutions have similar accuracy over broad ranges in correlation length scales. Not surprisingly, the Gaussian minima are sharper and the results are more sensitive to changes in the correlation length scales. In cases (c) and (d), where the measurement points are irregularly spaced, longer correlation length scales provide the better solutions, with the optimal lengths for the exponential model being between 120 and 160 m. In all cases, the non-diagonal weighting matrices provide better solutions than the diagonal weighting matrices. All of the solutions of case (a) with 98 measurement points are better than any solution of cases (b)–(d) with 49 measurement points, however, the best solution of case (b) is approaching the accuracy of the diagonal weighting matrix solution of case (a).

Figure 7 presents the variations of the  $SS(\hat{H} - H_A)_A$  with weighting matrices. In contrast to  $SS(\hat{Y}_A)$ , the head solutions continually improve as the correlation length increases. Again, the results computed with the Gaussian model are more sensitive to changes in correlation length than those computed with the exponential model. The heads are more accurate when non-diagonal weighting matrices are used than when the diagonal weighting matrix is used. This is especially true when the measurement points are irregularly spaced.

At the measurement points,  $SS(\hat{H} - H^*)$  is on the

order of  $10^{-6}$  whereas the  $SS(\hat{H} - H_A)_M$  is close to  $1 \times 10^{-4}$  in all cases. The solutions at the measurement points are quite close to the measured values at the measurement points because the solution is under-constrained in the sense that there are many more parameters to estimate than measurement points. In a case which is over-constrained,  $SS(\hat{H} - H^*)$  would be expected to be near  $10^{-4}$  given the present weighting.<sup>4</sup>

In Fig. 8 the variance of the ln- $T$  solution,  $\sigma_{\hat{Y}}^2$ , is plotted as a function of the weighting matrices. Adding the non-diagonal weighting terms increases the variance of the final estimate, and the longer the correlation length scale the more the increase. All of the solutions have less variance than the true ln- $T$  field (0.63). However, not all of the increased variance is due to an improvement in the estimate, since  $SS(\hat{Y} - Y_A)$  has distinct minima at correlation length scales which are shorter than the maximum (Fig. 6).

Figure 9 presents the change in the measurement residual portion of the functional with changes in the weighting matrices. The plots clearly demonstrate that as the correlation length increases, the measurement functional increases in magnitude. Consequently, the results in Fig. 9 imply that increasing the correlation length scale in the covariance model causes the regularization terms to exert more influence on the solution. In contrast, Fig. 10 shows that the regularization portion of the functional has distinct minima. The estimate of the solution variance (Fig. 11) also contains distinct minima. The minima are found at correlation lengths which are close to the correct values found in the data (Fig. 3(b)). However, the correct values do not

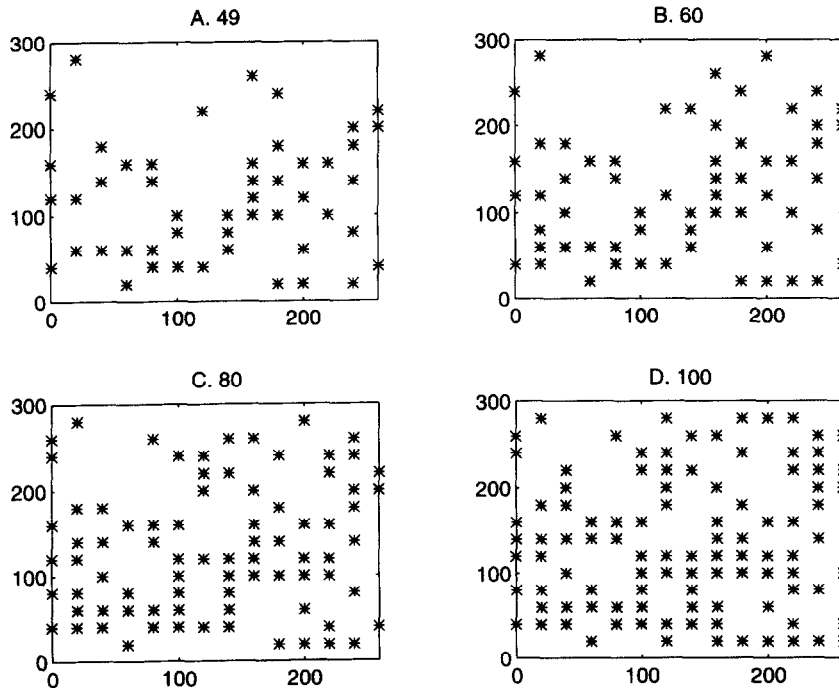


Fig. 12. Test 2 measurement point locations for (a) 49 points; (b) 60 points; (c) 80 points; (d) 100 points. Locations in (a) are the same as Fig. 4(d).

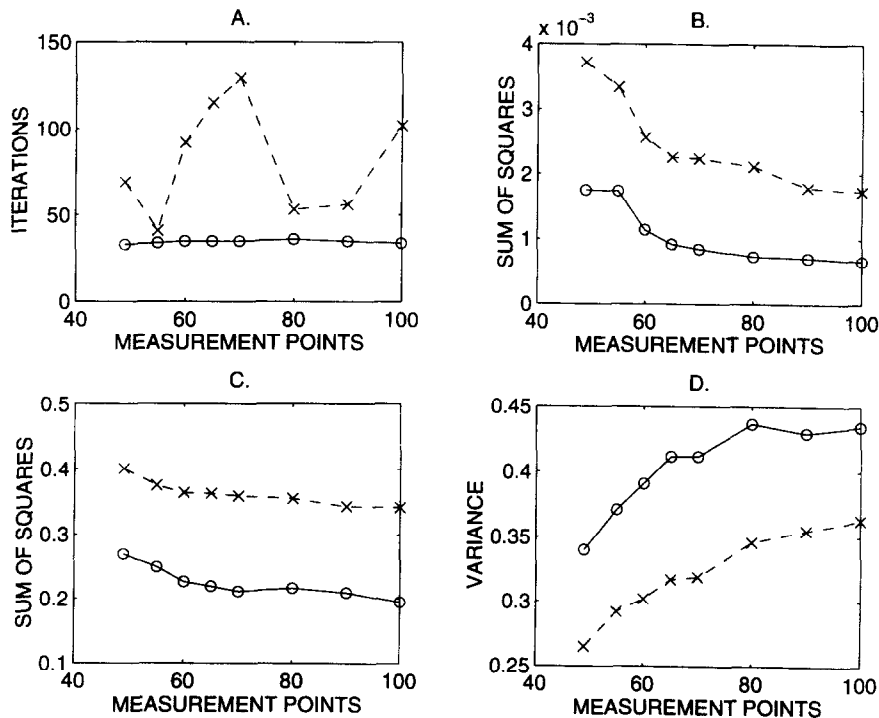
necessarily give the best solutions as evidenced by Figs 6 and 7. The shape of the solution variance curves (Fig. 11) are similar to those of the regularization portion of the functional (Fig. 10), and Figs 9, 10 and 11 indicate that the regularization term dominates the regression.

A second set of calculations was conducted to compare the benefit of using non-diagonal regularization weighting matrices to the addition of measurement points. Starting with the pattern in Fig. 12(a), measurement points were added and the problem resolved. Patterns of 49, 55, 60, 65, 70, 80, 90 and 100 points were tested. Figure 12 shows a selection of the patterns. Although theory exists to optimize the selection of the next measurement point (e.g. ref. 13), additional points were simply selected randomly for this simple comparison. The test problems were solved with a diagonal weighting matrix and a weighting matrix derived from the Gaussian covariance model, equation (4), with correlation length of 45 m.

The results are presented in Fig. 13. The diagonal weighting matrix results are presented as crosses connected by dashed lines and the Gaussian model results are presented as circles connected by solid lines. Figure 13(a) shows that solutions using the Gaussian model converged with less iterations than those computed with diagonal weighting matrices, implying that the inverse problem is more easily solved even with less measurement points. Figure 13(b) shows the variations in  $SS(\hat{H} - H_A)_A$ . With the Gaussian model, the head solutions are always better than the results computed

with the diagonal matrices. Only when the number of measurement points reaches 90 do the diagonal weighting matrix results reach the accuracy of the Gaussian model results computed with only 49 measurement points. Figure 13(c) shows the variations in  $SS(\hat{Y} - Y_A)$ . Again, the results computed with the Gaussian model are superior. The  $\ln-T$  solutions computed with a diagonal matrix and 100 measurement points are less accurate than the the results computed with the Gaussian model and only 49 measurement points. Finally, Fig. 13(d) shows the increase in  $\ln-T$  solution variance with the number of measurement points. The Gaussian model results contain more variance than the diagonal matrix results. These results indicate that developing a good covariance model for the regularization weighting matrix can have as much beneficial effect on the solution as adding several measurement points.

Both the weighting matrix and the measurement point locations have an effect on the spatial correlation of the solutions  $\ln-T$  field. Semivariograms from the solutions derived from the measurement points shown in Fig. 4(b) using different weighting matrices are shown in Fig. 14(a). All of the solutions have longer ranges and lower sills than the true  $\ln-T$  field. The solution calculated with a diagonal weighting matrix has the longest range and the lowest sill. The best solution, calculated with a Gaussian model and correlation length of 45 m, has a higher sill and shorter range than the best exponential solution which was calculated with a correlation length

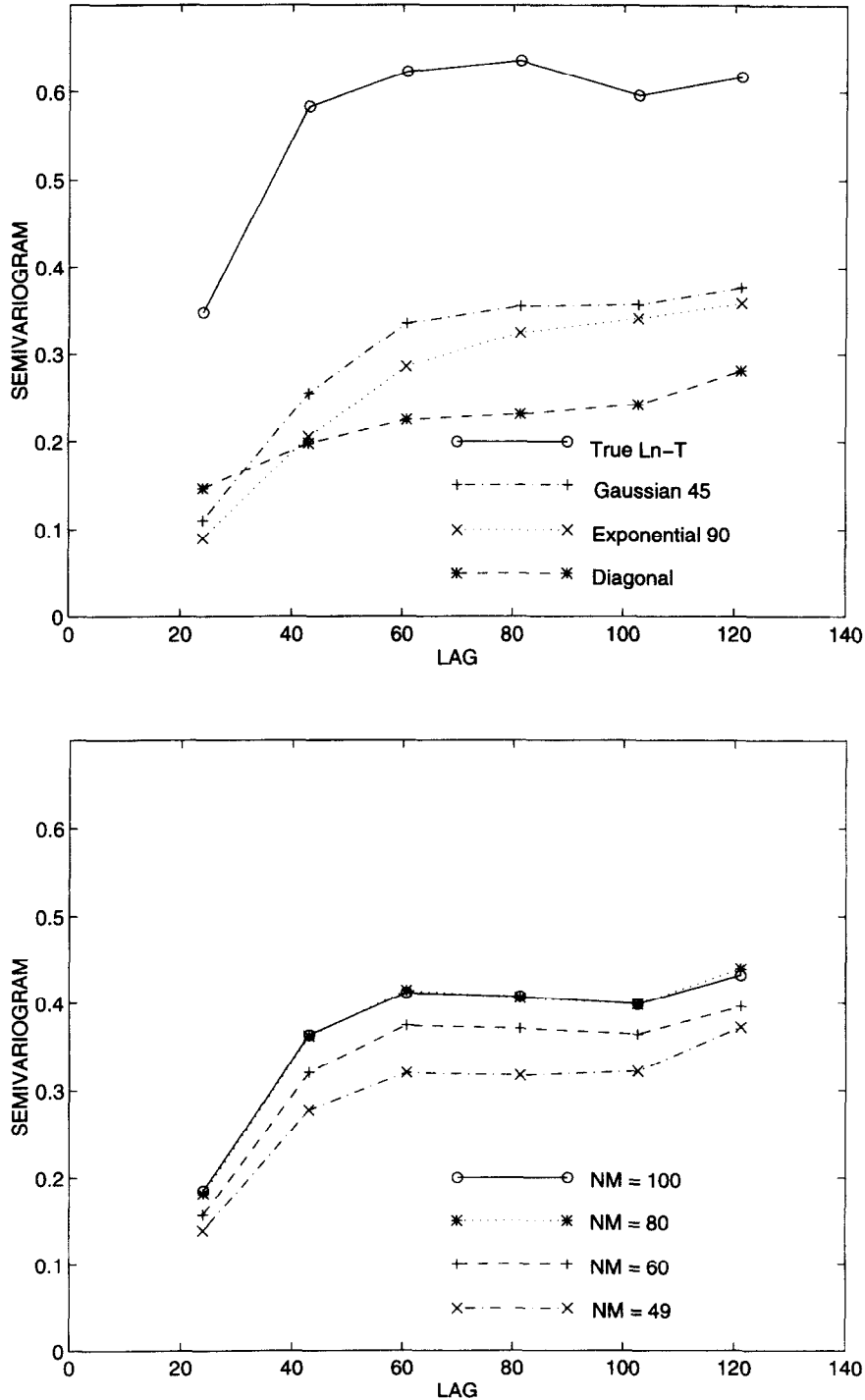


**Fig. 13.** Performance measures versus the number of measurement points. The measurement point locations are from the sequence illustrated in Fig. 12. Circles connected by solid lines are results computed with a Gaussian model with  $\lambda_c = 45$  m and crosses connected by dashed lines are computed with a diagonal weighting matrix (a). Number of iterations vs measurement points; (b)  $SS(\hat{H} - H_A)_A$  vs measurement points; (c)  $SS(\hat{Y} - Y_A)$  vs measurement points; (d)  $\sigma_{\hat{Y}}^2$  vs measurement points.

of 90 m. These results indicate that using a weighting matrix derived from a good model of the covariance of the errors in the original  $\ln-T$  estimates leads to solutions that better match the spatial correlation structure of the true field.

The effect of adding measurement points on the

solution semivariogram is shown in Fig. 14(b). The solutions were computed using a Gaussian model with correlation length 45 m and the measurement locations shown in Fig. 12. As the number of measurement points is increased from 49 to 60 to 80 m, the sill increases reflecting the increase in the variance of the solution.



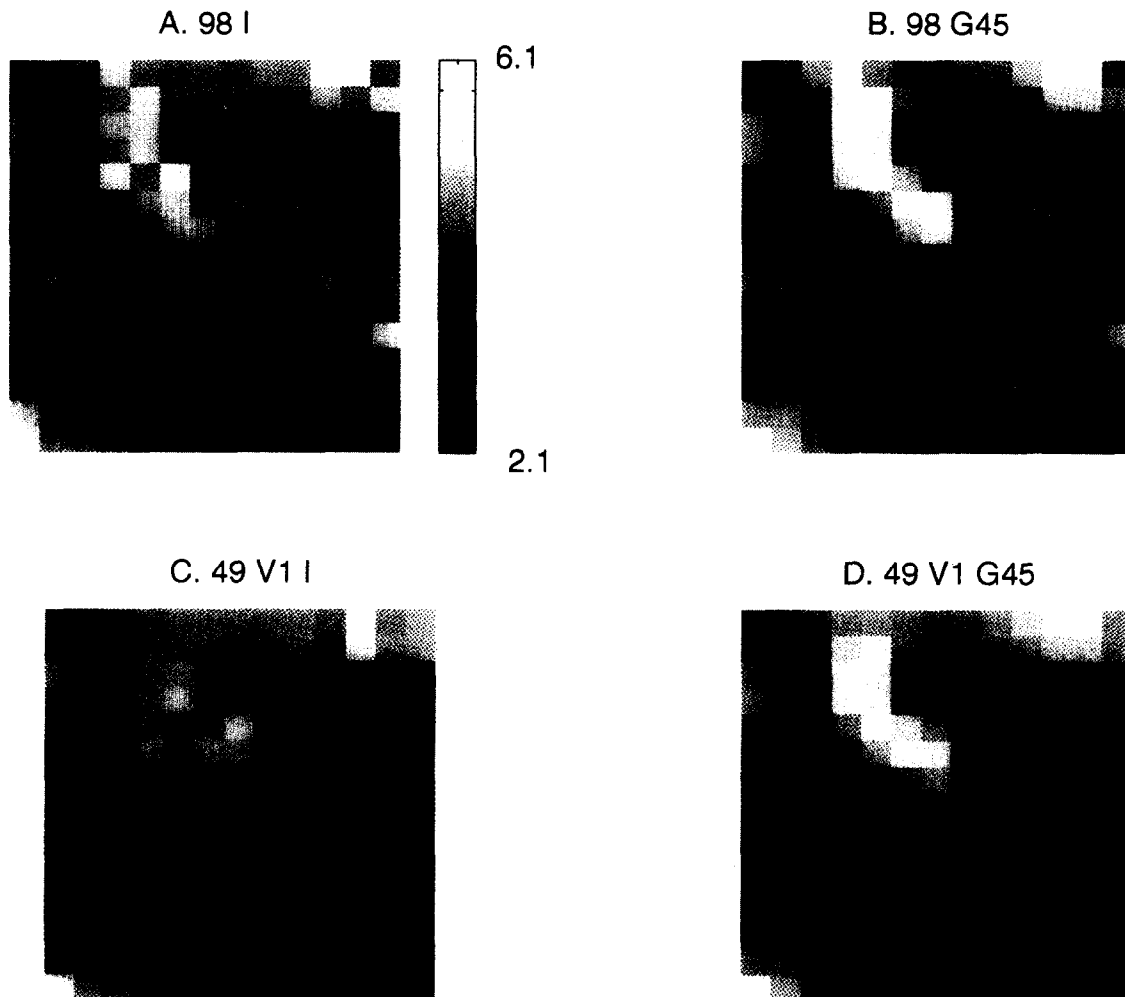
**Fig. 14.** Comparison of semivariograms of  $\ln-T$  computed using different weighting matrices. Semivariograms of the true field, and fields computed with a Gaussian model with  $\lambda_c = 45$  m, an exponential model with  $\lambda_c = 90$  m and a diagonal matrix. The measurement point locations of Fig. 4(b) were used. (b) Semivariograms of  $\ln-T$  vs the number of measurement points. Fields computed using a Gaussian model with  $\lambda_c = 45$  m and measurement point locations from Fig. 11(a) ( $NM = 49$ ), (b) ( $NM = 60$ ), (c) ( $NM = 80$ ) and (d) ( $NM = 100$ ).

The range does not change significantly. In this case going from correlation length of 80–100 m has little effect on the semivariogram and the curves are almost identical.

A comparison between four solutions is shown in Fig. 15. The computed  $\ln-T$  fields can be compared to the true field shown in Fig. 1. The results presented in Fig. 15(a), (b) are calculated using the 98 measurement points shown in Fig. 4(a) with a diagonal weighting matrix and a Gaussian derived weighting matrix of correlation length 45 m, respectively. The results presented in Fig. 15(c), (d) are calculated using the 49 measurement points shown in Fig. 4(b) with a diagonal weighting matrix and a Gaussian derived weighting matrix of correlation length 45 m, respectively. Comparison of Fig. 15(a), (b) with (c), (d) shows that the 98 measurement point solutions contain more detail than the 49 measurement point solutions. Obviously, more measurement points gives a more detailed picture of the distribution of head gradients, allowing for a more

detailed resolution of the  $\ln-T$  field. The Gaussian derived solutions, Fig. 15(b) and (d), are smoother and have better defined areas of high and low  $\ln-T$  than the diagonal weighting matrix solutions. They have done a better job at reproducing the connectivity of the extreme values which is extremely important in transport studies. The lack of variability in the  $\ln-T$  values is evident in the 49 measurement point, diagonal weighting matrix solution (Fig. 15(c)). As stated previously, All the solutions shown in Fig. 15 produced heads that accurately matched the measurements due to the under-determination of the measurement residual functional. In addition, all the solutions have, to some degree, captured the general trend of the true  $\ln-T$  field with the region of high  $\ln-T$  values in the north-central region and the lows separating that high from the northeast and southwest highs. The improvement in the solution due to the use of the Gaussian derived weighting matrix is obvious.

Finally, a series of different  $\ln-T$  fields were used to



**Fig. 15.** Computed  $\ln-T$  fields. (a) and (b) are calculated with the measurement points shown in Fig. 5(a), (c) and (d) are calculated with the measurement points shown in Fig. 5(b). (a) and (c) are calculated with a diagonal weighting matrix. (b) and (d) are calculated with a Gaussian model with  $\lambda_c = 45$  m. The true image is found in Fig. 1.

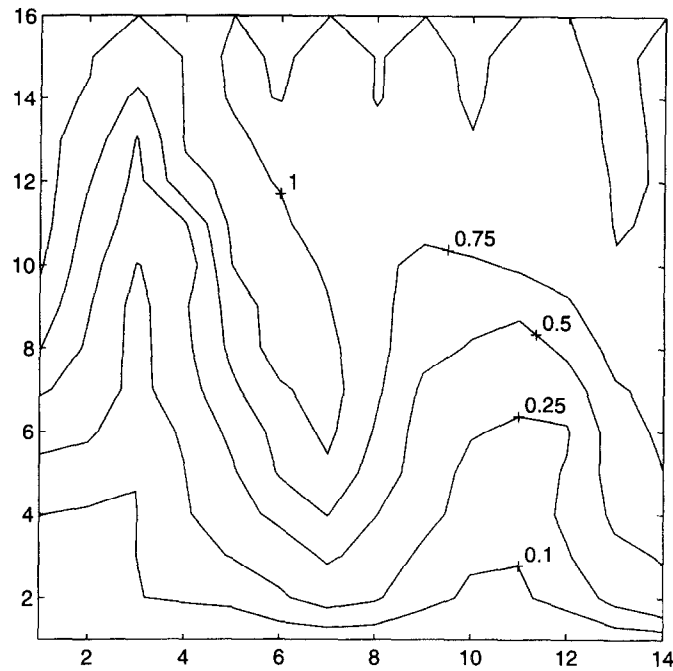


Fig. 16. Transport calculations. Normalized concentration are contoured at a 0.25 interval. Flow fields are derived from the head and transmissivity fields derived using the true  $\ln-T$  field shown in Fig. 1.

perform transport calculations. The fields were used to generate a steady state head field and transport calculations were computed with an upstream weighted finite element model.<sup>2</sup> The aquifer was assumed to have a thickness of one meter, a longitudinal dispersivity of 1 m, a transverse dispersivity of 0.1 m and an effective porosity of one. The initial concentrations are zero throughout the domain. The upper boundary was modeled as a constant concentration of one and as the simulation progresses the simulated solute will migrate from the top to the bottom of the domain. One hundred transport time steps of one day were simulated.

The final concentration distribution which is generated using the true  $\ln-T$  (Fig. 1) is shown in Fig. 16. The concentration field shows a solute finger that starts at the top near column 4 and migrates down towards an exit point near column 7. Smaller solute fingers have developed along the boundaries. The transport results for the four calculated  $\ln-T$  fields shown in Fig. 15 are presented in Fig. 17. The best result is produced by the  $\ln-T$  field which was calculated using 98 measurement points and the Gaussian covariance structure with a correlation length of 45 m (Fig. 17(b)). The results which were computed from 49 measurement points and a diagonal weighting matrix show little channeling and have missed the character of the true transport (Fig. 17(c)). Interestingly, the transport results from the 49 measurement, Gaussian covariance  $\ln-T$  field are comparable to the results from the 98 measurement, diagonal weighting  $\ln-T$  field (Fig. 17(a)). None of the computed fields develop the boundary solute fingers. Overall, the  $\ln-T$  fields computed with the fully populated weighting matrices

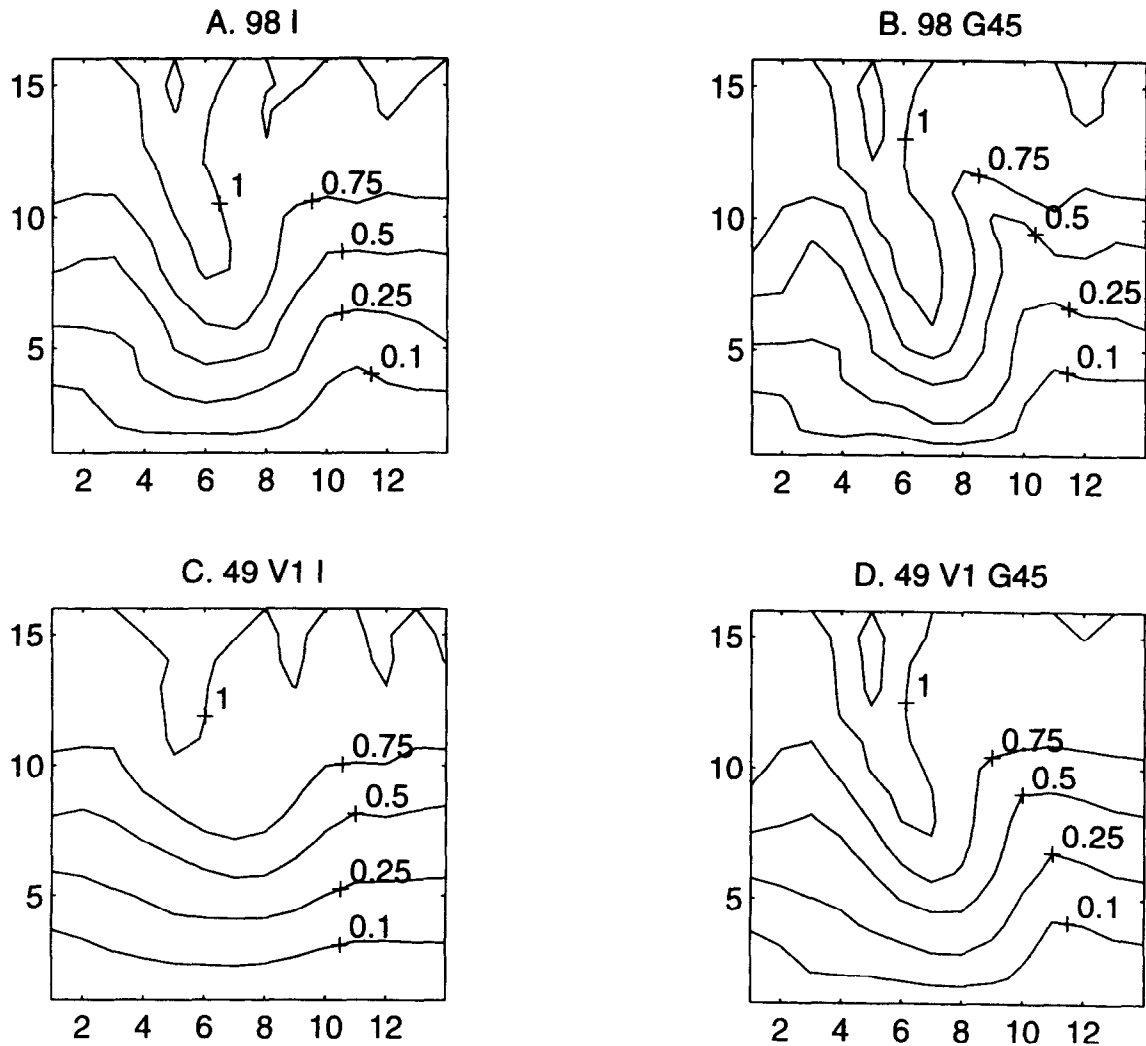
have produced concentration fields that are closer to the true concentration distribution than the diagonal weighting matrix fields.

## DISCUSSION

The numerical study that was presented is a simplified version of a field case. The examples correspond to the case where the prior information consists only of an estimate of the mean  $\ln-T$  and the covariance structure of the errors in the original  $\ln-T$  estimates. In a real case, we would have some idea of the variations in  $\ln-T$  due to the large number of monitoring wells. Further, we have neglected the confounding effects of small scale variations in  $\ln-T$  and errors in the boundary conditions, recharge rates, geometry, etc.

Even with the simplifications, several important observations and generalizations can be made about the effect of the regularization residual weighting matrix on parameter estimates. Solutions are improved and the number of iterations required to converge decreases with the addition of the non-diagonal terms in the regularization weighting matrix.

The variances of the estimated  $\ln-T$  fields are less than the true  $\ln-T$  field. Adding measurement points improves the accuracy and increases the variance of the estimated  $\ln-T$  fields. The changes are expected, because the additional measurements add detailed information about the spatial distribution of the head gradients and consequently push the solution of the  $\ln-T$  field away from the original estimate. Adding the non-diagonal weighting matrix also adds variability to the estimated



**Fig. 17.** Transport calculations. Normalized concentration are contoured at a 0.25 interval. Flow fields are derived from the head and transmissivity fields derived using the  $\ln-T$  fields shown in Fig. 5. (a) 98 measurement points and a diagonal weighting matrix (Fig. 15(a)); (b) 98 measurement points with a Gaussian model with  $\lambda_c = 45$  m (Fig. 15(b)); (c) 49 measurement points and a diagonal weighting matrix (Fig. 15(c)); (d) 49 measurement points with a Gaussian model with  $\lambda_c = 45$  m (Fig. 15(d)).

$\ln-T$  field. As evidenced by the increase in  $J_M$ , the non-diagonal weighting matrices increase the influence of the regularization terms.

All of the solutions have accurately matched the measurement values at the measurement points. The good match is to be expected since the number of measurement points is substantially less than the number of  $\ln-T$  values that must be estimated. In addition, all of the solutions, to some degree, have captured the gross features of the true  $\ln-T$  field. However, the addition of non-diagonal terms to the weighting matrix decreases the range of the estimated  $\ln-T$  semivariogram to a value more consistent with the true field, produces smoother  $\ln-T$  fields, and enhances the definition of the continuity of the extreme values. The last point is especially critical in transport modeling. The importance to transport modeling is demonstrated by the superior match to the true concentration field by the concentrations calculated with the correct non-

diagonal weighting matrices. By comparing Fig. 17(a) and (d), it is seen that the correct non-diagonal weighting can potentially improve the accuracy of solutions as much as adding many measurement points.

The solution results were sensitive to the choice of the covariance model used to develop the regularization weighting matrix. For this example, the best fit covariance model was a Gaussian model with correlation length 45 m. Models with correlation lengths 40–50 m gave similar results. The Gaussian model was sensitive to changes in the correlation length, and the inversion of the covariance matrix became unstable at correlation lengths greater than 50 m. The sensitivity of the Gaussian model is to be expected since the correlation length is squared in the model (equation (4)). The exponential model did not fit the actual covariance of the errors as well as the Gaussian model, but it produced better conditioned covariance matrices.

The exponential model results were generally less accurate than the Gaussian model results, but they were much less sensitive to variations in the correlation lengths.

Adding measurement points should improve inverse solutions, and this is illustrated in Figs 13, 14(b) and 15. However, Fig. 13 and a comparison of Fig. 15(a) and (d) demonstrate that using a regularization weighting matrix derived from a covariance model that captures the essence of the errors in the original estimates of  $\ln-T$  can improve the solution as much as adding many measurement points. These results imply that a significant amount of model analysis should be devoted to developing a reasonable model for the covariance structure of the original  $\ln-T$  estimates if optimization techniques are going to be used to calibrate groundwater models.

## CONCLUSIONS

The use of optimization techniques instead of the trial and error methods has the advantages that the solutions are not *ad hoc*, and they can be more accurate. One of the main difficulties is that the dimension of the parameter space is generally much greater than the number of measurements. The normal approach has generally been to group the data into zones and solve a lower dimension problem. Zoning is appropriate for well located abrupt geological contacts, but its use may otherwise be problematic. In addition, the zoning is generally done before the optimization, so that all small scale variation is lost. It is often difficult to know if the zoning is appropriate or correct. For example, what would be the correct zones to apply to Fig. 1? How would you define them from a limited number of well tests?

Another approach is to approximate the  $\ln-T$  values of all of the elements. In this approach, zoning decisions are not required before the calibration step. The regularization residuals can be used to alleviate the problems associated with under determination.

The use of regularization terms with diagonal weighting corresponds to the assumption that the errors in the original estimates of  $\ln-T$  are uncorrelated. In general the errors will be correlated, and the use of regularization terms with diagonal weighting matrices leads to solutions that are not optimal. Due to the lack of measurement points, these solutions tend to be somewhat oscillatory (Fig. 15(a), (c)) and unsatisfactory.

Adding an assumed covariance structure to the original estimate errors leads to matrices that are non-diagonal. The affect of the non-diagonal terms is to add spatial correlation to the final results and to reduce oscillations in the final  $\ln-T$  field. The departures from the original estimates are increased, and, in the cases presented here, the variance and spatial correlation of

the  $\ln-T$  estimates are closer to the true field values. The connectivity of the extreme values is better defined. Although results are sensitive to errors in the covariance model, it has been demonstrated that even a relatively poor estimate provides better results than the diagonal weighting matrices. The use of non-diagonal weighting matrices leads to  $\ln-T$  fields which generate more accurate transport model results than the diagonal weighting matrix generated  $\ln-T$  fields.

If optimization is to be used to calibrate models before zoning, non-diagonal regularization weighting matrices should be used. Appropriate non-diagonal weighting can improve a solution more than adding several measurement points to a diagonally weighted solution. A significant amount of effort should be devoted to estimating the appropriate covariance model of the original  $\ln-T$  estimate errors. These errors will be related to the spatial correlation of the  $\ln-T$  fields, and methods for estimating the spatial correlation of  $\ln-T$  from a limited amount of field data need to be developed. Possibly, covariance models can be developed from the geological models, depositional environment models or geophysics.

## ACKNOWLEDGEMENT

Support for this work was provided by the Natural Sciences and Engineering Research Council of Canada under Operating Grant OP0122023.

## REFERENCES

1. Ababou, R., Bagtzoglou, A. C. & Wood, E. F. On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. *Math. Geol.* **26** (1994) 99–133.
2. Babu, D. K. & Pinder G. F. A finite element-finite difference alternating direction algorithm for three dimensional groundwater transport, *Adv. Wat. Res.* **7** (1992) 116–119.
3. Bard, Y., *Nonlinear Parameter Estimation*. Academic Press, New York, 1974.
4. Bentley, L. R., Least squares solution and calibration of steady state groundwater flow systems. *Adv. Wat. Res.* **16** (1993) 137–48.
5. Bentley, L. R., Solving and calibrating groundwater flow systems with the penalty method. In *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Vol. 2, ed. K. W. Hipel. Kluwer Academic, The Netherlands, 1994, pp. 55–67.
6. Bentley, L. R., Groundwater parameter estimation with non-diagonal weighting matrices. In *Computational Methods in Water Resources X*, eds. A. Peters, et al. Kluwer Academic, The Netherlands, 1994, pp. 711–18.
7. Carrera, J. & Neuman, S. P., Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Wat. Resour. Res.* **22** (1986) 199–210.

8. Clifton, P. M. & Neuman, S. P., Effects of kriging and inverse modeling on conditional simulation of the Avra Valley Aquifer in Southern Arizona. *Wat. Resour. Res.* **18** (1982) 1215–34.
9. Cooley, R. L., Incorporation of prior information on parameters into non-linear regression groundwater flow models. 1. Theory. *Wat. Resour. Res.* **18** (1982) 965–76.
10. Desbarats, A. J. & Srivastava, R. M., Geostatistical characterization of groundwater flow parameters in a simulated aquifer. *Wat. Resour. Res.* **27** (1991) 687–98.
11. Ginn, T. R. & Cushman, J. H., Inverse methods for subsurface flow: a critical review of stochastic techniques. *Stochastic Hydrol. Hydraul.* **4** (1990) 1–26.
12. Isaaks, E. H. & Srivastava, R. M., *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford, 1989.
13. McKinney, D. C. & Louks, D. P., Network design for predicting groundwater contamination. *Wat. Res. Resour.* **28** (1992) 133–47.
14. Pinder, G. F. & Gray, W. G., *Finite Element Simulation in Surface and Subsurface Hydrology*. Academic Press, New York, 1977.