GROUND WATER

FIELD REPORTS

AN EXERCISE IN GROUND-WATER MODEL CALIBRATION AND PREDICTION

by David L. Freyberg[a]

Abstract. For a classroom exercise, nine groups of graduate students calibrated a numerical ground-water flow model to a set of perfectly observed hydraulic head data for a hypothetical phreatic aquifer. All groups used exactly the same numerical model and identical sets of observed data. After calibration, the students predicted the hydraulic head distribution in the aquifer resulting from a modification in one boundary condition. A quantitative analysis of the results of this calibration-prediction exercise vividly demonstrates some of the difficulties in parameter identification for ground-water flow models. Group predictions differed significantly. Successful prediction was strongly correlated with successful estimation of conductivity values, and was essentially unrelated to successful estimation of aquifer bottom elevations or with the number of trial-and-error simulations required for calibration. Most importantly, success in prediction was unrelated to success in matching observed heads under premodification conditions. In this sense, good calibration did not lead to good prediction.

Introduction

Estimating the values of aquifer flow parameters is one very important application of ground-water modeling technology. Numerical modeling techniques are often the only practical way of obtaining parameter estimates for use in predicting future aquifer responses, particularly over large spatial scales. This application of models is commonly described as solving a parameter identification, or inverse, problem. The parameter identification problem may be defined simply as the solution of the ground-water flow equation for the values of its parameters using observations of the dependent variable (hydraulic head), aquifer geometry, and boundary conditions. The method of solution can vary from trial-and-error matching of observed head data using sequential simulations (often called calibration) to sophisticated, automated techniques that find parameter sets which optimize a multiobjective criterion characterizing the goodness of fit between simulated and observed head fields.

The nature of the inverse problem for ground-water flow has been studied extensively over the last twenty years. Yeh (1986) provides a useful review of this literature and the state of the art. Several well-known characteristics of the general parameter identification problem have dominated attempts to develop effective solution techniques. Neuman (1973) and Neuman and Yakowitz (1979) give clear descriptions of these characteristics. First, parameter estimates, particularly for hydraulic conductivity (or transmissivity), are highly sensitive to noise or errors in observed head data. Small observation errors can lead to large errors in estimated conductivity. Second, the solution to the inverse problem often, indeed typically, is nonunique. More than one parameter field may yield indistinguishably good fits to observed head data.

These characteristics of the inverse problem arise no matter what general solution approach is used. They lead to several important practical problems: instability; uncertainty in identified parameter values; and uncertainty in predictions made with identified parameter values. A number of approaches have been suggested for dealing with these difficulties. Yeh (1986) provides a thorough summary. Prominent among these include: incorporating all available prior information about parameter values, such as from well logs, pumping tests, or geologic mapping (cf., Neuman, 1980; Cooley, 1982, 1983; Kitanidis and Vomvoris, 1983); reducing the number of parameters to be calibrated by subdividing the field into a small number of homogeneous zones (cf., Yeh, 1986); reducing the number of parameters by introducing interpolation functions (cf., Yeh, 1986); and explicitly accounting for the statistical structure of the spatial variability and uncertainty in the head field caused by the spatial variability of aquifer parameters and observation errors (cf., Kitanidis and Vomvoris, 1983).

[a]Assistant Professor, Department of Civil Engineering, Stanford University, Stanford, California 94305-4020.

Received November 1986, revised December 1987, accepted January 1988.

Discussion open until November 1, 1988.

Several investigators have explored the predictive reliability of calibrated models by a posteriori comparison of model predictions with observed behavior. For example, Konikow and Person (1985) and Person and Konikow (1986) have examined and evaluated solute transport predictions made for an irrigated stream-aquifer system in the Arkansas River Valley. Konikow (1986) carefully assessed the quality of water-table elevation predictions over a 10-year period in the Salt River and lower Santa Cruz River basins in Arizona. Lewis and Goldstein (1982) compare solute transport model predictions with observed data for a 7-year prediction period at the Idaho National Engineering Laboratory. In general, these studies have demonstrated only modest predictive success. While it is difficult to define unambiguously the source of prediction errors, these studies have all pointed to the importance of correctly predicting future system stresses and boundary conditions, as well as parameter values, in assuring adequate predictions of system response.

McLaughlin (1984) and McLaughlin and Johnson (1987) provide another perspective on predictive reliability by comparing three groundwater modeling exercises in the San Juan basin in New Mexico. Each study used the same data base and same numerical flow code and yet provided quite different predictions. While data were not available to test the various predictions, McLaughlin carefully assesses the impact of data interpretation, assumptions, and parameter values on the prediction discrepancies.

The purpose of this paper is to provide a vivid demonstration of some of the difficulties in parameter identification using the results from a relatively simple exercise developed for a graduate-level course in ground-water hydrology. This exercise was tailored to the classroom and was not intended to simulate the full complexity of a complete parameter identification problem. However, by aggregating and then examining the results across all of the students participating in the exercise, a revealing picture of the nature and difficulties of the inverse problem emerges with a perspective somewhat different from that obtained from case studies or theoretical investigations.

Hypothetical Aquifer System

The parameter identification exercise is based on a hypothetical aquifer system and the most common parameter identification technique, trial-and-error calibration of a limited parameter set. The aquifer system was designed in the context of both the goals of the exercise and a particular numerical model, the well-known USGS Modular Three-Dimensional Finite-Difference Ground-Water Flow Model (MacDonald and Harbaugh, 1984). Figure 1 presents a discretized, schematic diagram of the aquifer. For simplicity and clarity, this discretized form for the aquifer is taken as its true

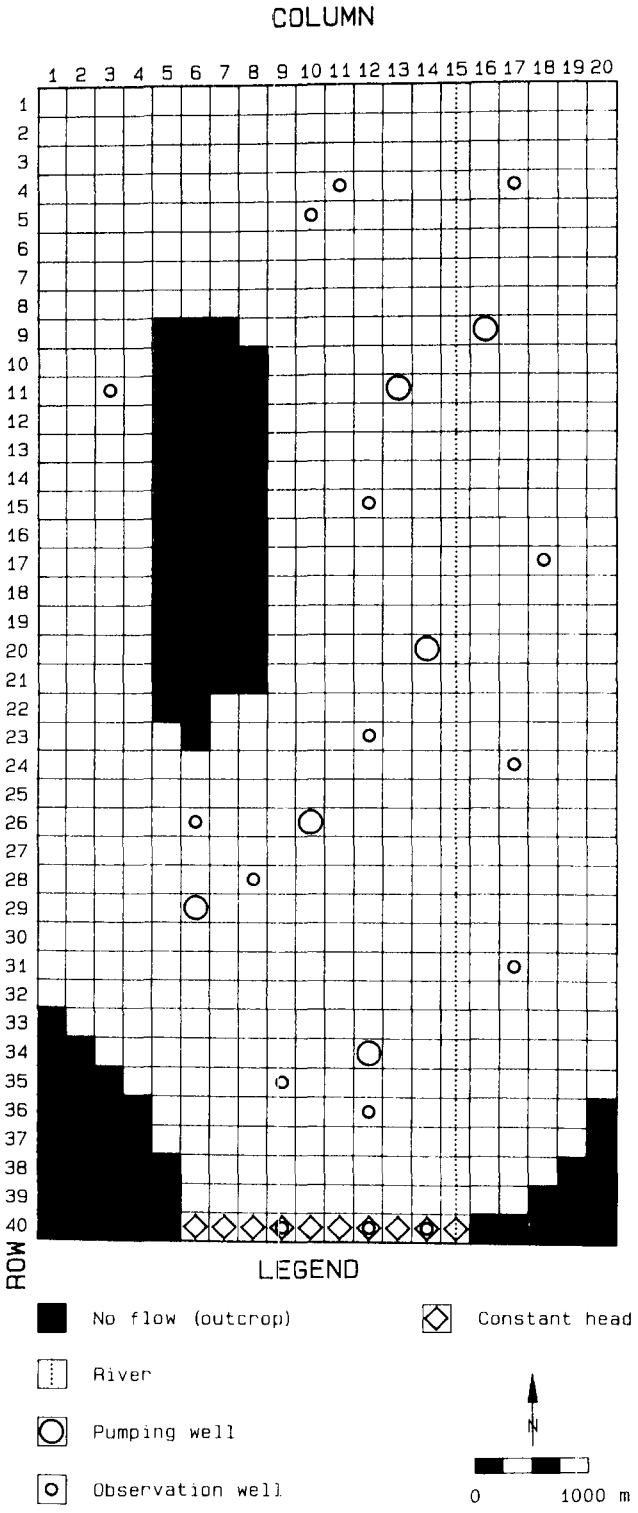

Fig. 1. Discretized schematic diagram of the hypothetical aquifer used in the parameter identification exercise.

Table 1. Magnitudes of Boundary Heads and Aquifer Stresses

	Location (row, column)	*Head (m)*		*Location (row, column)*	*Rate ($\times 10^{-3}$ m^3/s)*
Boundary heads	40,6	16.9	Pumping rates	9,16	8.2
	40,7	16.4		11,13	4.1
	40,8	16.1		20,14	3.9
	40,9	15.6		26,10	0.83
	40,10	15.1		29,6	0.72
	40,11	14.0		34,12	4.3
	40,12	13.0			
	40,13	12.5			
	40,14	12.0			
	40,15	11.4			

	Location (row, column)	*Bottom (m)*	*Surface (m)*	*Location (row, column)*	*Bottom (m)*	*Surface (m)*
River elevations	1,15	20.00	20.10	21,15	15.00	15.56
	2,15	19.75	19.87	22,15	14.75	15.33
	3,15	19.50	19.65	23,15	14.50	15.11
	4,15	19.25	19.42	24,15	14.25	14.88
	5,15	19.00	19.19	25,15	14.00	14.65
	6,15	18.75	18.97	26,15	13.75	14.43
	7,15	18.50	18.74	27,15	13.50	14.20
	8,15	18.25	18.51	28,15	13.25	13.97
	9,15	18.00	18.28	29,15	13.00	13.75
	10,15	17.75	18.06	30,15	12.75	13.52
	11,15	17.50	17.83	31,15	12.50	13.29
	12,15	17.25	17.60	32,15	12.25	13.07
	13,15	17.00	17.38	33,15	12.00	12.84
	14,15	16.75	17.15	34,15	11.75	12.61
	15,15	16.50	16.92	35,15	11.50	12.38
	16,15	16.25	16.70	36,15	11.25	12.16
	17,15	16.00	16.47	37,15	11.00	11.93
	18,15	15.75	16.24	38,15	10.75	11.70
	19,15	15.50	16.02	39,15	10.50	11.48
	20,15	15.25	15.79	40,15	10.25	11.25

Net recharge rate: 1.6×10^{-9} m/s.
River bottom conductance: 0.05 m^2/s.

state. No attempt is made to relate the discretized form to an underlying continuous geometry or to continuous distributions of aquifer parameters.

The aquifer is shallow and phreatic, underlain by an impervious stratum allowing no leakage. The computational grid is square, with nodes 250 m apart. Vertical homogeneity is assumed adequate to allow treatment as a single layer. Aquifer boundaries on the north, east, and west allow no flow, while the southern boundary maintains a constant head. In addition, an outcrop introduces an internal no-flow boundary, as shown on the figure. The aquifer receives spatially uniform areal recharge from infiltration. A river in hydraulic contact with the aquifer flows from north to south along the entire length of the aquifer. Six extraction wells are located in the aquifer, along with 16 nonproducing observation wells. All aquifer stresses and boundary conditions are taken to be constant in time, yielding steady flow conditions in the aquifer. Table 1 gives the magnitudes of these stresses and boundary conditions. Figure 2 presents the discretized bottom topography of the aquifer, while Figure 3 shows the discretized distribution of hydraulic conductivity. The aquifer bottom is relatively horizontal on the east, sloping gently downward to the west and south. The bottom rises steeply to the outcrop in the western half of the aquifer and contains an incised channel extending up the west side of the oucrop. Bottom elevations are greatest in the northwest corner of the aquifer. Hydraulic conductivities are greatest

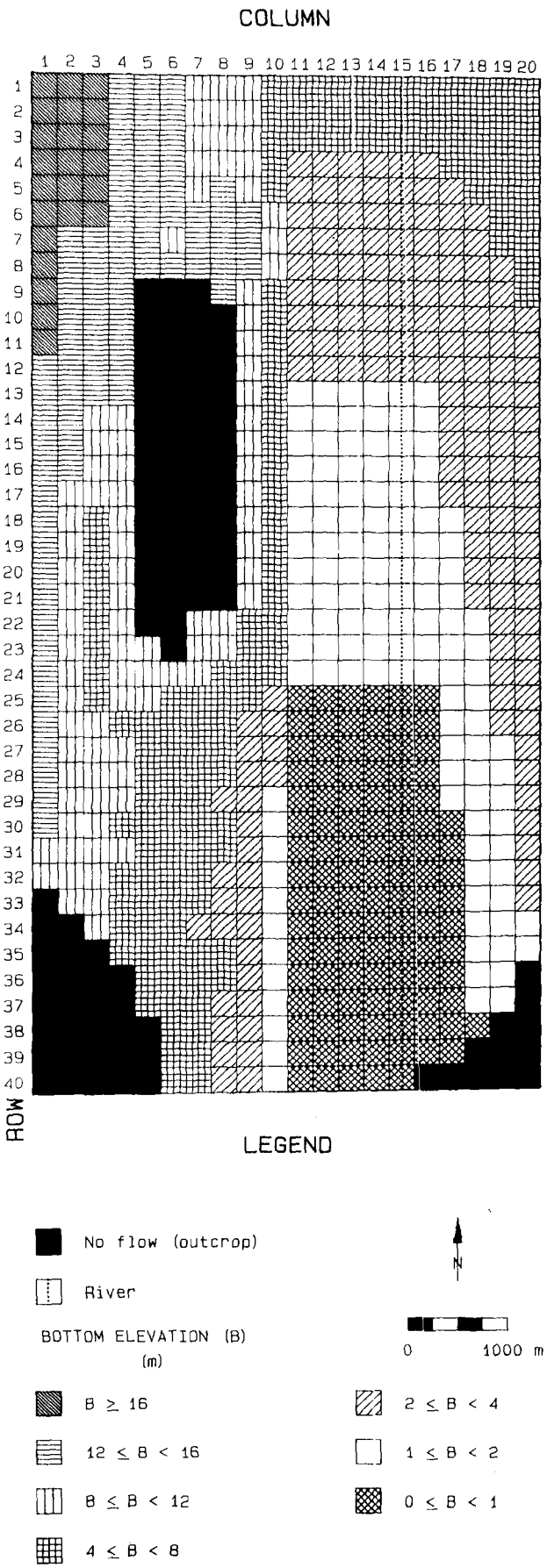

Fig. 2. True discretized distribution of aquifer bottom elevations. Actual resolution is greater than that shown by the shading. Note also that the shading scale is nonlinear.

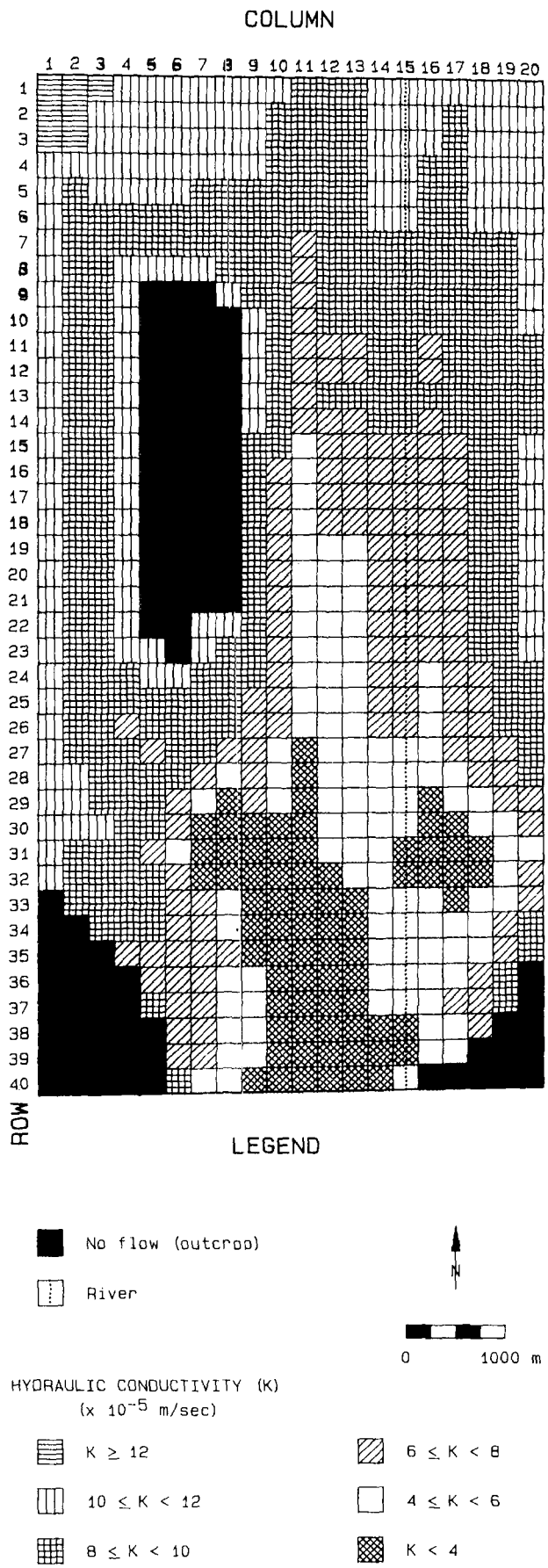

Fig. 3. True discretized distribution of hydraulic conductivity. Actual resolution is greater than that shown by the shading.

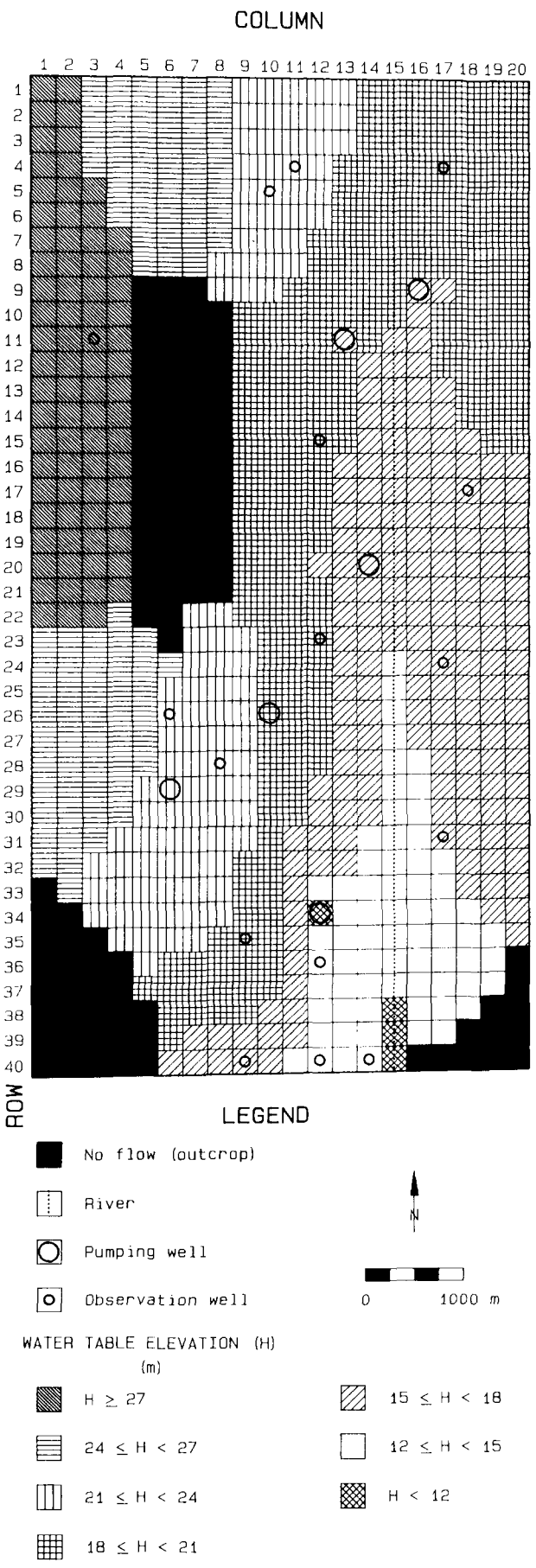

Fig. 4. True water-table elevations under premodification conditions. Actual resolution is greater than that shown by the shading.

along the north, east, and west boundaries, and adjacent to the outcrop. Values trend from larger to smaller moving away from boundaries and the outcrop, leading to a low-conductivity channel to the west of the outcrop and a broad low-conductivity zone underlying and paralleling the river. Conductivities are least in the southern portion of the aquifer.

Figure 4 presents the discrete, steady-state, hydraulic head distribution computed by the MacDonald-Harbaugh code for this aquifer. Note that the shading scheme of Figure 4 has a resolution of 3 m, while computed heads were rounded to the nearest 0.1 m. Under the conditions of the exercise this discrete head distribution is taken as the true state of the aquifer. The true flow field is not particularly complex, although it is far from uniform. Gradients are greatest in the southern portion of the aquifer near the constant-head boundary and to the west of the river. Note that the river forms a reasonably effective boundary. A ground-water divide occurs to the west of the impervious outcrop. Each of the production wells produces a drawdown cone which perturbs the regional flow field. Well 34,12 produces the largest perturbation, while Well 29,6 has relatively little impact. Areal recharge from net infiltration provides essentially all the inflow to the aquifer. The constant-head boundary is an outflow boundary along its entire length, while the river is effluent for all reaches but one. Well 9,16 induces significant recharge from the river and is responsible for the one influent reach of river. The overall flux balance for the aquifer is shown in Table 2.

Problem Description

The goal of parameter identification is rarely the parameter estimates. Rather, the ultimate goal is nearly always defined by a prediction problem. In this case, the problem was to predict the impact on water-table elevations of reducing the conductance of the river bottom by two orders of magnitude (as a result of lining the channel, for example).

Table 2. Flux Balance for True Parameter Set

Flux	*Inflow* ($\times 10^{-3}$ m^3/s)	*Outflow* ($\times 10^{-3}$ m^3/s)
Recharge	69.5	—
River leakage	1.53	44.0
Wells	—	22.05
Constant-head boundary	—	4.95
Net flux	71.0	71.0

More specifically, the participants were asked to predict quantitatively the impact of this modification to the river on the water-table elevations in the six grid blocks in which the production wells are located. This prediction was to be made using a parameter set found by trial-and-error calibration of the MacDonald-Harbaugh code. A subset of the information characterizing the true discrete aquifer system under premodification conditions formed the available data base. While not especially realistic, such a prediction problem is useful in illustrating some of the key issues of the parameter identification problem.

The information available to the participants was, in fact, quite extensive. The participants knew: the exact horizontal geometry of the aquifer, including discretization; the nature of the boundary conditions; all information characterizing interaction of the river with the aquifer; and all production well pumping rates. They received censored information on the elevation of the aquifer bottom elevation and on premodification, steady-state water-table elevations. In particular, aquifer bottom elevations were available only for grid blocks containing wells, both observation and production. Water-table observations were available only for the 16 observation well grid blocks. (In the following discussion, this information is often referred to as "the observed data.") The participants also did not know the recharge rate from infiltration, but were informed that the recharge rate is spatially uniform. All information was provided error-free; observation errors and noise were not simulated. Table 3 summarizes the calibration-prediction problem.

The participants were told that the measure of success for the exercise would have three equally weighted components:

1. A component measuring the prediction error, consisting of the root mean square error (RMSE) of the water-table elevations in the six production well grid blocks under the modified conditions,

$$\mathrm{RMSE_P} = \sqrt{\frac{1}{6}\sum_{i=1}^{6}[\hat{h}^i_{mod} - h^i_{mod}]^2} \qquad (1)$$

where $\hat{h}^i_{mod}$ is the predicted water-table elevation at the i^{th} production well grid block under modified conditions, and h^i_{mod} is the true water-table elevation at the i^{th} production well grid block under modified conditions. This is essentially a least-squares measure of the goodness of prediction fit.

2. A component measuring the error in the calibrated conductivity values over the entire aquifer, consisting of the RMSE of the calibrated conductivity grid:

$$\mathrm{RMSE_K} = \sqrt{\frac{1}{705}\sum_{i=1}^{705}[\hat{K}_i - K_i]^2} \qquad (2)$$

where $\hat{K}_i$ is the estimated hydraulic conductivity in the i^{th} grid block, and K_i is the true hydraulic conductivity in the i^{th} grid block.

3. A component measuring the efficiency of use of computer resources: N, the number of calibration trials using the numerical model.

Note that this measure includes no component that measures the quality of the calibrated fit to the available observed head data under premodification conditions, such as the RMSE of the calibrated heads in the observation well grid blocks. The participants were given no information on the magnitude of conductivity and no head data under modified conditions. They were thus able to evaluate and control directly only one component

Table 3. The Calibration-Prediction Problem

Problem	*Unknowns*	*Available information*
Calibration	Spatially uniform recharge rate. Hydraulic conductivity for all 705 grid blocks. Bottom elevations for all grid blocks without wells.	Aquifer geometry. Boundary conditions. River characteristics. Pumping rates at six production wells. Bottom elevations in 16 grid blocks containing observation wells and six containing production wells. Water-table elevations in 16 grid blocks containing observation wells.
Prediction	Modified water-table elevations at six grid blocks containing production wells.	All of the above information. Calibrated recharge rate. Calibrated bottom elevations at remaining 683 grid blocks. Calibrated hydraulic conductivity at all grid blocks.

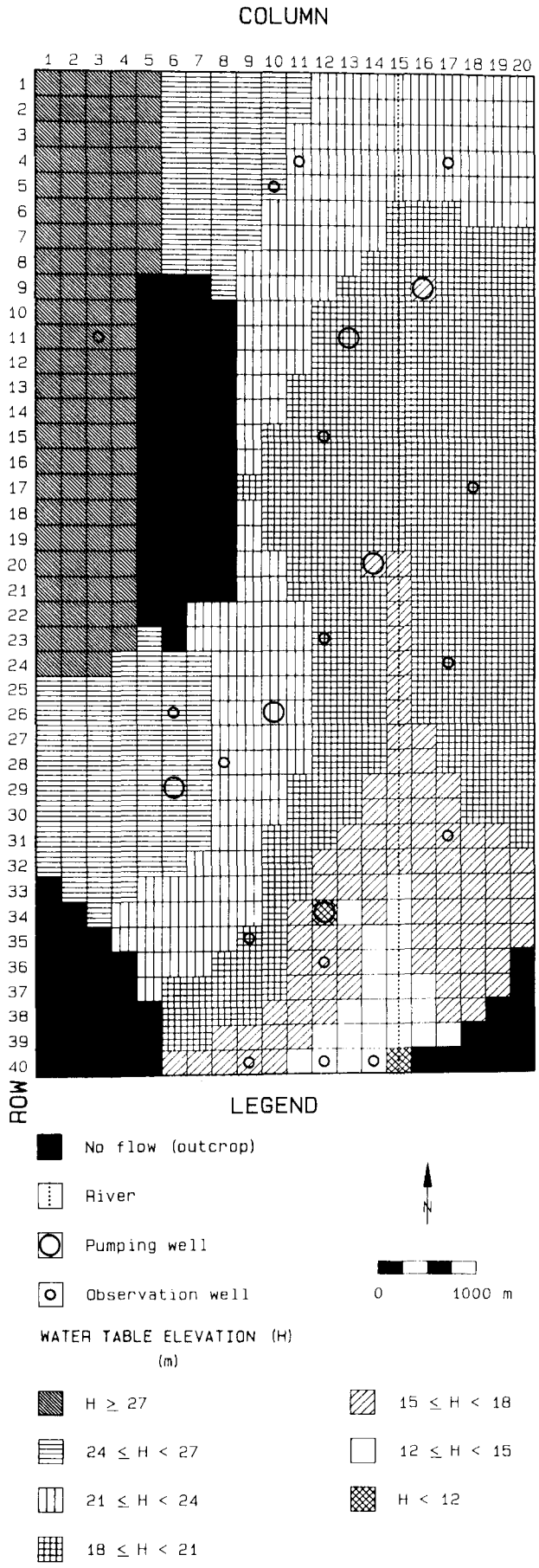

Fig. 5. True water-table elevations after modification. Actual resolution is greater than that shown by the shading.

of the measure of success, the number of calibration trials. They had to develop their own measure of the quality of the calibrated fit which could be evaluated directly from the observed data. Their clear challenge was to develop a calibration measure for premodification conditions that would lead to a good estimation of the conductivity field and to a good prediction of aquifer water-table elevations under modified conditions.

Participants were organized into groups of two or three. Each group received essentially unlimited access to an IBM PC/AT microcomputer within the two-week time limit for the exercise. Communication between groups was encouraged. Immediately prior to this exercise, the students had completed an assignment which used the MacDonald-Harbaugh model to explore thoroughly the sensitivity of steady flow in a geometrically simple phreatic aquifer to changes in parameter values and external stresses.

Results

The true discrete hydraulic head distribution resulting from the two order of magnitude reduction in the conductance of the river bottom is shown in Figure 5. As in Figure 4, the resolution of the shading is 3 m. Somewhat unrealistically, water-surface elevations in the river were assumed to be unchanged after modification. The river/aquifer interaction algorithm used in the MacDonald-Harbaugh code is based on the simple application of Darcy's Law across a layer of sediments assumed to line the channel bottom (MacDonald and Harbaugh, 1984, pp. 209-217). Therefore, because the unmodified river is effluent, the reduction in river bottom conductance leads to a general increase in water-table elevations throughout the aquifer. As a result, all river reaches after modification are effluent. As shown in the flux balance of Table 4, relatively more discharge now occurs across the southern constant-head boundary, and less occurs across the river bottom. Gradients

Table 4. Flux Balance for True Parameter Set after Modification

Flux	*Inflow* ($\times 10^{-3}$ m^3/s)	*Outflow* ($\times 10^{-3}$ m^3/s)
Recharge	69.5	–
River leakage	–	40.5
Wells	–	22.05
Constant-head boundary	–	6.91
Net flux	69.5	69.5

Table 5. Summary of Performance Measures for the Calibration-Prediction Problem

	— *Announced Measures* —		— — — — — — — — — — *Additional Measures* — — — — — — — — — —					
Group	*N*	$RMSE_K$ *(m/s)*	$RMSE_P$ *(m)*	$RMSE_B$ *(m)*	$RMSE_C$ *(m)*	$RMSE_{CA}$ *(m)*	$RMSE_{PA}$ *(m)*	ΔR *(%)*	ΔK *(%)*
A	15	6.2E-05	2.84	1.82	0.37	0.72	1.34	43.8	38.5
B	25	3.7E-05	2.27	1.41	0.18	0.32	0.94	−37.5	−44.6
C	21	4.7E-05	1.87	1.99	0.18	0.72	0.93	12.5	21.5
D	19	1.8E-04	7.64	2.31	0.29	0.53	5.18	237.	238.
E	25	2.4E-04	7.96	2.02	0.06	0.44	5.59	256.	285.
F	21	3.5E-05	2.31	1.67	0.12	0.27	0.95	−37.5	−43.1
G	28	3.7E-05	2.26	1.41	0.15	0.32	0.94	−37.5	−44.6
H	27	3.0E-05	1.26	3.16	0.16	0.46	0.56	12.5	16.9
I	38	3.9E-05	3.60	2.90	0.14	0.52	1.65	−43.1	−47.7

Definitions:

N = Number of calibration trials.
$RMSE_K$ = Root mean square error in calibrated conductivity field.
$RMSE_P$ = Root mean square error in predicted head at production wells.
$RMSE_B$ = Root mean square error in calibrated bottom elevations.
$RMSE_C$ = Root mean square error in calibrated head at observation wells.
$RMSE_{CA}$ = Root mean square error in entire calibrated head field.
$RMSE_{PA}$ = Root mean square error in entire predicted head field.
ΔR = Error in calibrated net recharge rate, percent of true value.
ΔK = Error in mean of calibrated conductivity distribution, percent of true value.

in the southern portion of the aquifer become steeper in response to the need to increase the flux across the southern boundary. (The head at this boundary was also assumed to be unchanged by the river modification.) Of the six grid blocks containing production wells, the water-table elevation in the block containing Well 34,12 is most sensitive to the modification, while the water-table elevation in the block containing Well 29,6 is least sensitive.

Table 5 summarizes the performance of each of the nine student groups participating in the exercise. The first four columns of the table give the group identifier and the three components of the measure of success the participants knew was to be used to evaluate their performance: the number of trials, N; the RMSE of the calibrated conductivity field, $RMSE_K$; and the RMSE of the predicted heads at the six production well grid blocks, $RMSE_P$. The overall measure of success is shown graphically as a stacked bar chart in Figure 6. Each element of a bar represents the group's performance on one of the components, normalized by the range of values found by all the groups and weighted so that the maximum value of the overall measure is 1.0. For example, for the component representing the number of calibration trials:

$$\text{Bar height} = 0.33 \left[\frac{N_{max} - N}{N_{max} - N_{min}} \right] \qquad (3)$$

where N is the number of calibration runs; N_{max} is the maximum number of runs used by any group; and N_{min} is the minimum number of runs used by any group.

Figure 6 indicates that six of the nine groups performed roughly comparably. Group A achieved the highest score on the weighted scale, largely on the strength of the relatively few trials they used for their calibration. Group H was clearly most successful in terms of both prediction RMSE and calibrated conductivity field RMSE. Their prediction RMSE of 1.26 m is more than 0.6 m less than that of the second-best prediction.

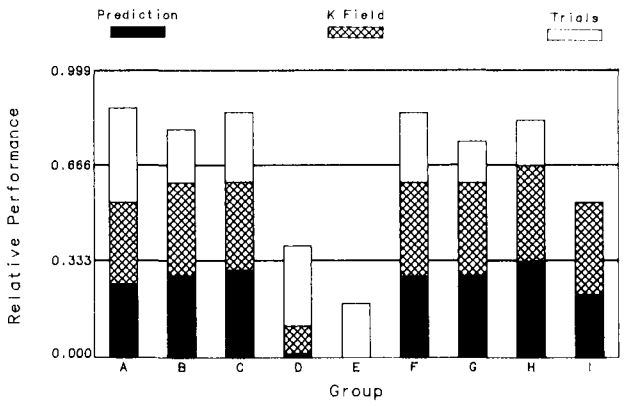

Fig. 6. Relative performance of student groups on the three-component measure of success.

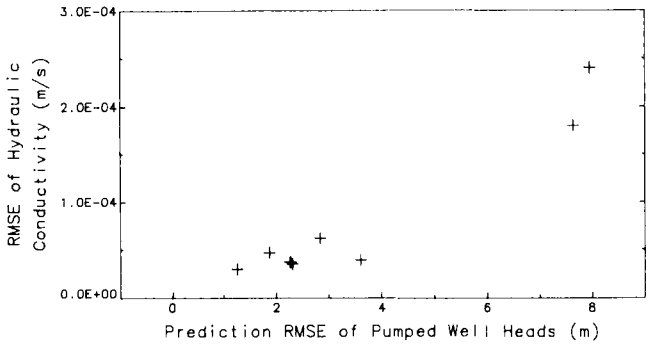

Fig. 7. RMSE of calibrated hydraulic conductivity field, $RMSE_K$, vs. RMSE of the predicted heads at the six pumping well grid blocks, $RMSE_P$.

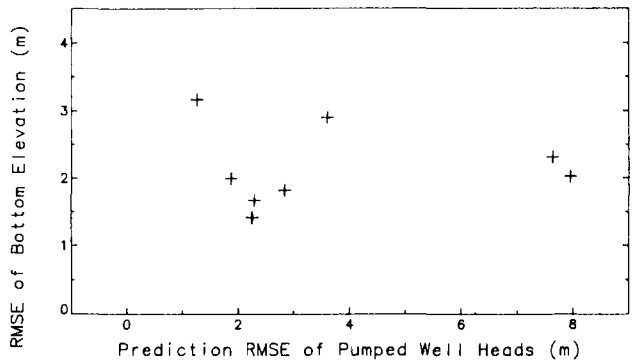

Fig. 8. RMSE of calibrated aquifer bottom elevations, $RMSE_B$, vs. RMSE of the predicted heads at the six pumping well grid blocks, $RMSE_P$.

Not surprisingly, there is a fairly strong correlation between prediction error and conductivity error, as Figure 7 demonstrates graphically. For this problem, successful prediction is predicated on a good estimate of the conductivity field. Note, however, that the data in Table 5 show essentially no correlation between prediction RMSE and the number of trials. Given the relative inexperience of the participants, this too is probably not surprising.

Table 5 also presents six additional performance measures. These measures provide further insight into several important aspects of the calibration-prediction problem. Included are:

1. $RMSE_B$, the RMSE of the calibrated bottom elevations;
2. $RMSE_C$, the RMSE of the calibrated head at the 16 grid blocks containing observation wells;
3. $RMSE_{CA}$, the RMSE of the calibrated head at all grid blocks;
4. $RMSE_{PA}$, the RMSE of the predicted head at all grid blocks;
5. ΔR, the percent error in the calibrated net recharge rate; and
6. ΔK, the percent error in the mean of the calibrated conductivity field.

Comparison of the measures associated with the calibratable parameters, i.e., hydraulic conductivity, net recharge rate, and bottom elevation, is revealing. First, note that the errors in net recharge rate, ΔR, and in the mean hydraulic conductivity, ΔK, are highly positively correlated. Because recharge is the only significant inflow to the aquifer, an overestimation of the magnitude of the recharge rate compensates in a general way for an overall overestimation of the magnitude of the hydraulic conductivity (and vice versa).

Figure 8, plotting the error in calibrated bottom elevations, $RMSE_B$, against the prediction error in the six pumping well grid block heads, $RMSE_P$, demonstrates graphically that prediction success was unrelated to success in identifying aquifer bottom elevations. This result is consistent with the typically observed lack of sensitivity of water-table elevations to small changes in bottom elevation.

Figure 9 demonstrates that the prediction error for the hydraulic head at all 705 grid blocks, $RMSE_{PA}$, is highly correlated with the prediction error for the hydraulic head at the six grid blocks containing pumping wells, $RMSE_P$. For this example, a good prediction at the six pumping well grid blocks guaranteed a good overall prediction. Included on Figure 9 is the line of equal RMSE. It is interesting to note that the RMSE, an average prediction error, is in fact consistently smaller for the overall prediction than for the prediction at the six pumping well grid blocks. Such behavior reflects the fact that the grid blocks with pumping stresses are relatively close to the river, and sensitivity to errors in conductivity and bottom elevation when river characteristics are changed decreases with distance away from the river.

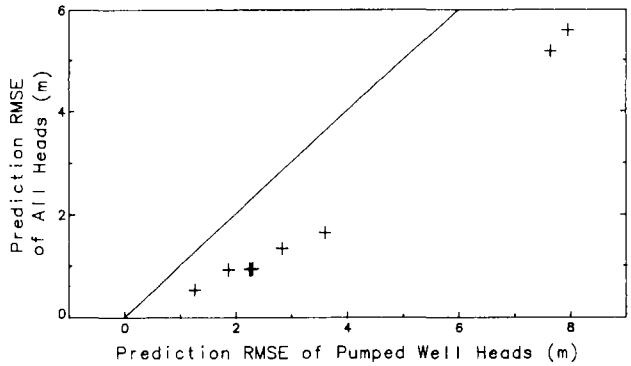

Fig. 9. RMSE of the predicted heads at all grid blocks, $RMSE_{PA}$, vs. RMSE of the predicted heads at the six pumping well grid blocks, $RMSE_P$.

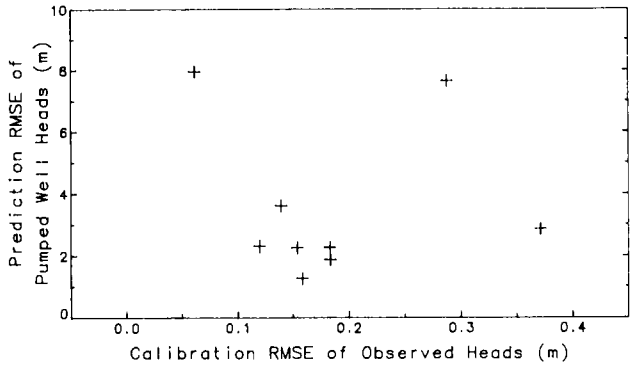

Fig. 10. RMSE of the predicted heads at the six pumping well grid blocks, $RMSE_P$, vs. RMSE of the calibrated heads at the 16 observation well grid blocks, $RMSE_C$.

The most revealing demonstration of the nature of the calibration-prediction problem comes from a comparison of the prediction and calibration errors given in Table 5. Figure 10 is a plot of the error in the *predicted* heads for the six pumping well grid blocks, $RMSE_P$, against the error in the *calibrated* heads at the 16 observation well grid blocks, $RMSE_C$. Most obvious in the plot is the lack of any significant correlation between the two error measures. The ability of a calibrated parameter set to reproduce the observed data was clearly not an indicator of the ability of that parameter set to predict the system response under modified conditions. In fact, in this case the "best" calibration ($RMSE_C$ = 0.06 m, by Group E) yielded the worst prediction ($RMSE_P$ = 7.96 m), while the best prediction (by Group H) resulted from an "average" calibration. Also note that in general, prediction RMSE's are an order of magnitude larger than calibration RMSE's.

The relationship between the error in the calibrated conductivity field and the *calibration* error of the observed heads is shown on Figure 11. Given the discussion of the previous paragraph and the strong correlation between conductivity error and prediction error noted earlier, the lack of correlation between conductivity error and calibrated head error is not surprising. A "good" calibration was achieved just as easily using a set of poorly estimated parameter values as it was using a set of well-estimated parameters. This result emphasizes that the quality of the parameter set cannot be measured by the quality of the calibration on observed heads.

A final perspective is provided in Figure 12. Shown there is the error in the entire calibrated head field, $RMSE_{CA}$, plotted against the error in the calibrated head at the 16 observation well grid blocks, $RMSE_C$. Note that the point corresponding to the group achieving the best prediction is found near the center of the array of points. Recall that the prediction error for the entire head field was less than the prediction error at the six pumping well grid blocks (see Figure 9). Here, in contrast, the calibration error for the entire head field is greater than the calibration error at the 16 observation well grid blocks. (The line of equal RMSE is included on Figure 12.) The explanation for this behavior lies in the localized sensitivity of head to conductivity. The tendency among the calibrators was to improve the calibration at the observation well grid blocks by adjusting conductivity values close to the observation wells. Such local "improvements," of course, have relatively little impact away from the observation wells; therefore, larger overall calibration errors typically occur. Note, however, that the overall calibration error for each group is always smaller than the overall prediction error.

Discussion

The results of this calibration-prediction exercise, summarized quantitatively in Table 5,

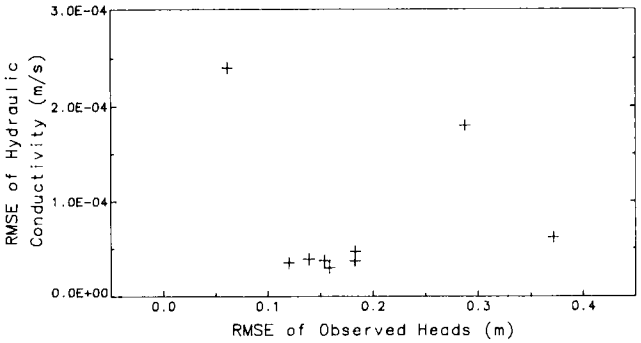

Fig. 11. RMSE of the calibrated hydraulic conductivity field, $RMSE_K$, vs. RMSE of the calibrated heads at the 16 observation well grid blocks, $RMSE_C$.

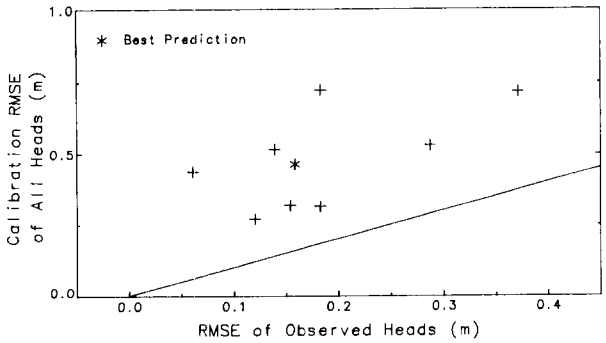

Fig. 12. RMSE of the calibrated heads at all grid blocks, $RMSE_{CA}$, vs. RMSE of the calibrated heads at the 16 observation well grid blocks, $RMSE_C$.

clearly demonstrate the nonuniqueness of solutions to an inverse problem. However, it is important to recognize the sources of this nonuniqueness. First of all, for this problem, differences in calibrated parameter fields and predicted heads under modified conditions are *not* due to: (1) a lack of knowledge of aquifer type, system geometry, or boundary conditions (model errors); (2) numerical discretization errors; or (3) noisy observed data. While all of these factors are extremely important in actual parameter identification problems, in this case calibrators had perfect knowledge of the model form, used identical numerical approximations with the same discretization, and received true observed head data. Nonuniqueness in this case arises solely from: (1) the attempt to identify as many as 1389 parameters using 22 observed data (see Yeh, 1986); (2) the use of different measures of the goodness of the calibration fit; and (3) the use of different strategies to combat anticipated nonuniqueness. For example, some groups attempted to equalize calibration errors among all observation wells, while others focused on fitting at selected wells. In a more telling example, the group achieving the best prediction chose to zone the conductivity field into a relatively few homogenous regions, while the group producing the worst prediction chose to "tweak" the conductivity field grid block by grid block to achieve a good (in fact, the best) local fit to the observed data.

Perhaps the most important observation from this exercise is that the significance of the differences in the calibrated parameter fields for the different groups is not apparent in the results from the calibration alone. Simple measures of the goodness of a calibrated fit to head data are inadequate to evaluate the true worth of a calibrated parameter set. Good calibration, in this sense, does not equal good prediction.

Acknowledgments

Preparation of this manuscript was supported by the National Science Foundation through grant ECE-8451565. Thomas C. Black designed, coded, and implemented the software to evaluate the numerical performance measures.

References

Cooley, R. L. 1982. Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 1, Theory. Water Resources Research. v. 18, no. 4, pp. 965-976.

Cooley, R. L. 1983. Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 2, Applications. Water Resources Research. v. 19, no. 3, pp. 662-676.

Kitanidis, P. K. and E. G. Vomvoris. 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Water Resources Research. v. 19, no. 3, pp. 677-690.

Konikow, L. F. 1986. Predictive accuracy of a groundwater model—lessons from a postaudit. Ground Water. v. 24, no. 2, pp. 173-184.

Konikow, L. F. and M. Person. 1985. Assessment of long-term salinity changes in an irrigated stream-aquifer system. Water Resources Research. v. 21, no. 11, pp. 1611-1624.

Lewis, B. D. and F. J. Goldstein. 1982. Evaluation of a predictive ground-water solute-transport model at the Idaho National Engineering Laboratory, Idaho. U.S. Geological Survey Water Resources Investigations Rept. 82-25.

MacDonald, M. G. and A. W. Harbaugh. 1984. A modular three-dimensional finite-difference ground-water flow model. U.S. Geol. Survey, Nat'l. Center, Reston, VA.

McLaughlin, D. B. 1984. A comparative analysis of groundwater model formulation: the San Andreas-Glorieta case study. Hydrologic Engineering Center, U.S. Army Corps of Engineers, Davis, CA.

McLaughlin, D. B. and W. K. Johnson. 1987. Comparison of three groundwater modeling studies. Journal of Water Resources Planning and Management. v. 113, no. 3, pp. 405-421.

Neuman, S. P. 1973. Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty. Water Resources Research. v. 9, no. 4, pp. 1006-1021.

Neuman, S. P. and S. Yakowitz. 1979. A statistical approach to the inverse problem of aquifer hydrology, 1, Theory. Water Resources Research. v. 15, no. 4, pp. 845-860.

Neuman, S. P. 1980. A statistical approach to the inverse problem of aquifer hydrology, 3, Improved solution method and added perspective. Water Resources Research. v. 16, no. 2, pp. 331-346.

Person, M. and L. F. Konikow. 1986. Recalibration and predictive reliability of a solute-transport model of an irrigated stream-aquifer system. Journal of Hydrology. v. 87, pp. 145-165.

Yeh, W. W-G. 1986. Review of parameter identification procedures in groundwater hydrology: The inverse problem. Water Resources Research. v. 22, no. 2, pp. 95-108.

* * * * *

David L. Freyberg is an Assistant Professor of Civil Engineering at Stanford University. After receiving A.B. and B.E. degrees from Dartmouth College in 1972, he spent three years in consulting practice in Boston, Massachusetts. In 1977 he received an M.S. degree from Stanford University. He completed his Ph.D. at Stanford in 1981 and has been on the faculty there since then, where he is responsible for the graduate teaching program in hydrology. His research interests center on contaminant transport in ground water, with a particular emphasis on understanding and modeling the effects of spatial variability and parameter uncertainty in transport prediction and model validation. He is a recipient of a 1985 Presidential Young Investigator Award.