

APPLICATION OF GAUSS ALGORITHM AND MONTE CARLO SIMULATION
TO THE IDENTIFICATION OF AQUIFER PARAMETERS

By Timothy J. Durbin

U.S. GEOLOGICAL SURVEY

Open-File Report 81-688



8035-11

Sacramento, California
1983

UNITED STATES DEPARTMENT OF THE INTERIOR

JAMES G. WATT, SECRETARY

GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information
write to:

District Chief
U.S. Geological Survey
Federak Building, Room W-2235
2800 Cottage Way
Sacramento, CA 95825

Copies of this report
can be purchased from:

Open-File Services Section
Western Distribution Branch
U.S. Geological Survey
Box 25424, Federal Center
Denver, CO 80225
Telephone: (303) 234-5888

CONTENTS

	Page
Abstract -----	1
Introduction -----	2
Identification problem -----	3
Ground-water model -----	3
Objective function -----	4
Gauss optimization technique -----	5
Levenberg parameter -----	6
Parameter discretization -----	7
Numerical example -----	9
Effects of data errors -----	11
Summary and conclusions -----	14
References cited -----	15

ILLUSTRATIONS

	Page
Figures 1-5. Maps showing:	
1. Ground-water subbasins, generalized geology, and steady-state water-level contours for the upper Coachella Valley ground-water basin, California -----	17
2. Element layout for the model of the Whitewater River and Garnet Hill subbasins -----	18
3. Geographic distribution of natural recharge to Whitewater River and Garnet Hill subbasins -----	19
4. Subdomains and water-level points used in computing optimum transmissivity from steady- state conditions -----	20
5. Optimum transmissivity obtained by objectively adjusting large-scale variations of transmissivity and by subjectively adjusting small-scale variations of transmissivity -----	21
6. Graph of convergence of the Gauss optimization technique for different search radii, r -----	22

TABLES

	Page
Table 1. Optimum surrogate parameters computed from field data -----	23
2. Summary statistics on random recharge estimates -----	24
3. Optimum surrogate parameter values computed using random recharge estimates -----	25
4. Hydraulic-head responses predicted with model using optimum surrogate parameters computed using random recharge estimates -----	26

CONVERSION FACTORS

In this report measurements are given in International System of Units. Conversion factors from International System of Units to inch-pound units are given below.

<u>Multiply</u>	<u>By</u>	<u>To obtain</u>
hm ³ (cubic hectometers)	810.7	acre-ft (acre-feet)
km (kilometers)	.6214	mi (miles)
m (meters)	0.3048	ft (feet)
m ² /s (meters squared per second)	10.76	ft ² /s (feet squared per second)
m ³ /s (cubic meters per second)	35.31	ft ³ /s (cubic feet per second)

ALTITUDE DATUM

National Geodetic Vertical Datum of 1929 (NGVD of 1929): A geodetic datum derived from a general adjustment of the first-order level nets of both the United States and Canada, formerly called mean sea level.

APPLICATION OF GAUSS ALGORITHM AND MONTE CARLO SIMULATION
TO THE IDENTIFICATION OF AQUIFER PARAMETERS

By Timothy J. Durbin

ABSTRACT

The Gauss optimization technique can be used to identify the parameters of a model of a ground-water system for which the parameter-identification problem is formulated as a least-squares comparison between the response of the prototype and the response of the model. Unavoidable uncertainty in the true stress on the prototype and in the true response of the prototype to that stress will introduce errors into the parameter-identification problem. These errors are subsequently transferred to the prediction problem. The fit of the model to the data that are used to identify parameters is not a good indicator of these errors. A Monte Carlo simulation of the parameter-identification problem and of the prediction problem can be used, however, to evaluate the effects on water-level predictions of errors in the recharge (and pumpage) data used in the parameter-identification problem.

INTRODUCTION

Some important problems encountered in simulating the behavior of ground-water systems can be classified as either prediction problems or parameter-identification problems. The prediction problem is to estimate the response of the prototype to specific inputs by using a mathematical model of the system. The parameter-identification problem is to find an optimum set of parameters for the model from concurrent prototypical input and output observations. In the parameter-identification problem, the mathematical structure of the equation characterizing the prototype is given, but the specific values of the coefficients of the equation, the initial conditions, or the boundary conditions are unknown. These unknowns are the parameters in the parameter-identification problem.

Current methods of parameter identification can be divided into direct or indirect methods (Neuman, 1973). The direct methods treat the model parameters as dependent variables in a formal boundary value problem (Stallman, 1956; Nelson, 1968; Nelson and McCollum, 1969; Kleinecke, 1971; Emsellem and de Marsily, 1971; Sagar, 1973; and Frind and Pinder, 1973). With indirect methods, existing estimates of the parameters are improved iteratively (Kruger, 1961; Jacquard and Jain, 1965; Jahns, 1966; Vemuri and Karplus, 1969; Coats and others, 1970; Slater and Durrer, 1971; Yeh and Tauxe, 1971; Lovell, 1971; Bruch and others, 1974; and McLaughlin, 1974). The indirect methods are generally computationally less efficient than the direct methods. However, the indirect methods are probably more flexible with respect to the different types of models that can be used.

The Gauss optimization technique (Gauss, cited in Wilde and Beightler, 1967, p. 299) is an indirect method of parameter identification. The application of this technique to the identification of the parameters of a ground-water model is demonstrated in this paper. In addition, the effects on the modeling process of errors in the input observations are analyzed. The modeling process includes both the identification of the model parameters and the use of the model to predict the response of the prototype to specific inputs. Unavoidable uncertainty in the true stress on the prototype and in the true response of the prototype to that stress will introduce errors into the parameter-identification problem. These errors are subsequently transferred to the prediction problem.

The Gauss technique has been used by others to identify the parameters for models of oil reservoirs (Jahns, 1966) and ground-water systems (McLaughlin, 1974). The originality of the work in this paper is chiefly in the method of parameter discretization, the application of the Gauss technique to a field case, and the analyses of the effects of data errors.

In particular applications, the mathematical equations of ground-water flow only approximately characterize the properties of the prototype. This lack of exact equivalency between the model and the prototype often leads to bias in the parameter estimates. This problem, however, is not addressed in this paper.

IDENTIFICATION PROBLEM

Ground-Water Model

The first step in the identification of parameters is to specify the model to be used and to isolate the unknown parameters. A partial differential equation that describes the steady flow of ground water in a two-dimensional isotropic aquifer is

$$\frac{\partial}{\partial x} \left(T \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(T \frac{\partial h}{\partial y} \right) + Q = 0 \quad (1)$$

where Q is the strength of a source or sink,

$$Q = \sum_i Q_w(x_i, y_i) \delta(x - x_i) \delta(y - y_i) \quad (2)$$

and h is the hydraulic head, Q_w is the rate of recharge or withdrawal at point (x_i, y_i) , T is the transmissivity, and δ is the Dirac delta function.

Equation 1 is to be solved in a domain Ω , which is enclosed by the boundary $\Gamma = \Gamma_1 \cup \Gamma_2$. The boundary conditions are the specified head boundary condition

$$h = \zeta_1 \text{ on } \Gamma_1 \quad (3)$$

and the specified flux boundary condition

$$\frac{\partial h}{\partial n} = \zeta_2 \text{ on } \Gamma_2 \quad (4)$$

where $\partial h / \partial n$ is the outward-pointing normal derivative on Γ_2 , and ζ_1 and ζ_2 are prescribed functions of location on Γ_1 and Γ_2 .

The Galerkin-finite element method is used to solve equation 1. The solution scheme follows Pinder and Frind (1972), except that linear shape functions are defined on triangular elements.

Objective Function

The second step in the identification of parameters is to select an objective function that compares the model response and the prototype response. Let $\theta \Xi (\theta_1, \theta_2, \dots, \theta_k)^T$ be a vector of the parameters to be identified. A least-squares objective function is given by

$$P = \sum_{i=1}^M [h_i - \hat{h}_i(\theta)]^2$$

where h_i is the measured hydraulic head at the data point (x_i, y_i) , \hat{h}_i is the computed hydraulic head at the data point, and M is the number of data points. Let the vector $\theta^* \Xi (\theta_1^*, \theta_2^*, \dots, \theta_k^*)^T$ be a particular point in an unbounded feasible region such that the value of the objective function is smaller than at any other point in the feasible region (McLaughlin, 1974, describes the application of the Gauss algorithm for the case of a closed feasible region). The members of θ^* are called the unconstrained optimum parameters.

A necessary condition at θ^* is that the gradient of P vanish there. The resulting normal equations are

$$\frac{\partial P}{\partial \theta_i} = 0 \quad i = 1, 2, \dots, K \quad (6)$$

or

$$J^T [h - \hat{h}(\theta)] = 0 \quad (7)$$

where

$$J \Xi \begin{bmatrix} \frac{\partial \hat{h}_1}{\partial \theta_1} & \frac{\partial \hat{h}_1}{\partial \theta_2} & \dots & \frac{\partial \hat{h}_1}{\partial \theta_K} \\ \frac{\partial \hat{h}_2}{\partial \theta_1} & \frac{\partial \hat{h}_2}{\partial \theta_2} & \dots & \frac{\partial \hat{h}_2}{\partial \theta_K} \\ \frac{\partial \hat{h}_M}{\partial \theta_1} & \frac{\partial \hat{h}_M}{\partial \theta_2} & \dots & \frac{\partial \hat{h}_M}{\partial \theta_K} \end{bmatrix}$$

is the Jacobian matrix, $h \Xi (h_1, h_2, \dots, h_M)^T$ is a vector of measured hydraulic heads and $\hat{h} \Xi (\hat{h}_1, \hat{h}_2, \dots, \hat{h}_M)^T$ is a vector of computed hydraulic heads.

Gauss Optimization Technique

The third step in the identification of parameters is to select an algorithm for adjusting the parameters so that the objective function is minimized. Equation 7 represents a system of simultaneous equations that are nonlinear in the parameters. This system can be solved by the Gauss technique (Gauss, cited in Wilde and Beightler, 1967, p. 299; Hartley, 1961; Marquardt, 1963; Jahns, 1966; and McLaughlin, 1974), which starts at a point close to θ^* and proceeds stepwise toward a minimum by successive improvements.

In the Gauss technique, equation 7 is linearized by expanding $\hat{h}(\theta)$ in a Taylor series about the current trial values of θ and by retaining only the linear terms. The substitution of the Taylor series representation of $\hat{h}(\theta)$ into equation 7 yields

$$(J^{j-1})^T J^{j-1} \Delta\theta = (J^{j-1})^T [h - \hat{h}(\theta^{j-1})] \quad (8)$$

where the superscript $j-1$ designates quantities evaluated at the last trial values and $\Delta\theta \equiv (\Delta\theta_1, \Delta\theta_2, \dots, \Delta\theta_K)^T$ is the vector of parameter corrections. The updated values of the parameters at the j th iteration are given by

$$\theta^j = \theta^{j-1} + \Delta\theta \quad (9)$$

Successive improvements are made until θ converges to θ^* , where convergence is defined as follows: If the value of the objective function has decreased at each of the last N iterations, and if the maximum change in the value of the objective function at successive iterates is less than some specified value, convergence to θ^* is declared. This definition is subjective, and most likely the definition will have to be modified in particular applications.

The objective function is typically polymodal (Jahns, 1966), and in practice, only a local minimum of the objective function can be found. Experience with the Gauss algorithm, however, seems to indicate that, for the steady-state form of the parameter-identification problem, the objective function tends to be unimodal as fewer measured water-level points are used and as fewer parameters are considered.

Levenberg Parameter

If the Taylor series approximation is poor, the Gauss optimization technique will behave erratically because of unwarranted extrapolation. Levenberg (1944) and Marquardt (1963) both proposed a modification of the Gauss technique to prevent parameter corrections that are far out of the region where the Taylor series approximation is still reasonable.

The modification proposed by Levenberg is to constrain the minimization of equation 5. The constraint is

$$|\Delta\theta| = r \quad (10)$$

where r is the radius of a hypersphere small enough that the Taylor series approximation is valid. The matrix form of the normal equations for this problem is (Marquardt, 1963)

$$[(J^{j-1})^T J^{j-1} + \lambda^{j-1} I] \Delta\theta = (J^{j-1})^T [h - \hat{h}(\theta^{j-1})] \quad (11)$$

where λ is the Levenberg parameter and I is the identity matrix. The substitution of equation 10 into equation 11 yields the relation between λ and r (Wilde and Beightler, 1967, p. 302)

$$\lambda^{j-1} = \frac{(J^{j-1})^T [h - \hat{h}(\theta^{j-1})]}{r} \quad (12)$$

Jacoby and others (1972) showed that when λ approaches ∞ or r approaches zero, the direction of the parameter-correction vector that is obtained from equation 11 is identical to the direction that is obtained from the gradient technique for finding the minimum of a function (Wilde and Beightler, 1967, p. 303). When λ approaches zero or r approaches ∞ , the parameter corrections computed from equation 11 are identical to the parameter corrections computed from equation 8. Levenberg (1944) and Marquardt (1963) proposed methods for performing an optimum interpolation at each iteration between the Gauss technique and the gradient technique by choosing an appropriate value of the Levenberg parameter. Satisfactory results can also be obtained, with experience, by making an a priori guess of the proper radius and by computing a new value for the Levenberg parameter at each iteration from equation 12, while holding r constant. The proper radius usually becomes apparent after a few trials.

PARAMETER DISCRETIZATION

In discussing ground-water models the question of the scale of parameter discretization invariably arises. But before the question can be considered, the scale of a model needs to be defined. The definition used here is that the scale of a two-dimensional ground-water model is the size of the area over which the system parameters are considered to be homogeneous.

Two aspects of model scale are to be considered. The first deals with the predictions that are made with a model. Consider the effect on computed hydraulic heads of random perturbations in spatially discretized system parameters. At some very small scale of discretization, these random perturbations will not significantly affect the computed hydraulic heads (Freeze, 1975). However, as the scale of parameter discretization is increased, at some point the random perturbations in the system parameters will significantly affect the computed hydraulic heads. The scale at which this occurs depends on the definition of a significant effect for a particular situation.

The second aspect of the scale of a ground-water model deals with the parameter-identification problem. Consider the effect on the parameter-identification problem of random errors in the measured water levels or in the estimates of the prototype pumpage and recharge. For example, the actual transmissivity of an aquifer might be uniform, but geographic variations of the hydraulic-head gradient occur because of variations in pumpage and recharge. Because of errors in the estimation of pumpage and recharge, however, the variation of the hydraulic-head gradient is interpreted as being caused in part by geographic variations of transmissivity. Similarly, errors in the estimation of hydraulic-head gradient from water-level data might also lead to the interpretation that transmissivity varies geographically. If, when considering the ground-water basin as a whole, errors in measured water levels, pumpage, and recharge are distributed with zero mean, at some very large scale of parameter discretization, these variations have very little effect on the parameter-identification problem. As the scale of parameter discretization is reduced, however, these errors may significantly affect the estimates of the system parameters.

Herein lies the dichotomy. We would like a small scale of parameter discretization for making predictions with the model. We would also like a large scale of parameter discretization for identifying the system parameters.

This disparity can be ameliorated somewhat by aggregating small-scale heterogeneity into surrogate parameters that are heterogeneous on a larger scale. In a practical sense, the small-scale heterogeneity would be at the level of the area of an element used in the finite-element solution of equation 1. The surrogate parameters would represent the aggregation of a group of elements.

The surrogate parameters are introduced into the parameter-identification problem by dividing the domain Ω into K subdomains ω_j ($j = 1, 2, \dots, K$). A subdomain is made up of one or more elements. The geographic variation of transmissivity in ω_j is given by

$$T_i = T_{ij}^{\circ} \theta_j \quad (13)$$

where T_{ij}° is a prescribed base transmissivity for the element i in the subdomain j , and θ_j is a surrogate parameter. The surrogate parameter is a dimensionless factor that multiplies the base transmissivity for each element in a subdomain. The parameter identification problem is then to identify the surrogate parameters that minimize the objective function, and the optimum transmissivity T_i^* is given by the relation

$$T_i^* = T_{ij}^{\circ} \theta_j^* \quad (14)$$

The relative distribution of transmissivity within a subdomain is fixed by the prescription T_{ij}° . The actual distribution of transmissivity within a subdomain is fixed by the identification of θ_j^* . The number and size of subdomains to be used and the distribution of T_{ij}° within each subdomain is prescribed ad hoc.

In the prototype, there is usually little reason to expect that the distribution of transmissivity will not be continuous, except at faults or other geologic discontinuities. The terms "continuous" and "continuity," which have precise meaning in mathematics, are used loosely in the context of the model to mean that the transmissivity of adjacent elements do not differ much. With this definition in mind, the continuity of transmissivity within a subdomain can be assured by assigning proper values to T_{ij}° . But the continuity of transmissivity across the boundaries of adjacent subdomains is not assured.

A strategy for obtaining a continuous distribution over the entire model domain is to use objective optimization for the large-scale variations of transmissivity and to use subjective optimization for the small-scale variations. For each subdomain, an initial distribution of T_{ij}° is prescribed that is continuous within and between adjacent subdomains. Optimum values for the surrogate parameters are computed. The new distribution of transmissivity is likely to be discontinuous across the subdomain boundaries. This distribution is then subjectively smoothed to eliminate these discontinuities, and optimum values for the surrogate parameters are again computed. This procedure is repeated until a continuous distribution is obtained.

An important difficulty in the parameter identification problem is that it is virtually impossible to express subjective information in a generalized mathematical form. But the subjective information that is usually available to a hydrologist is important to the parameter-identification problem. An advantage of the above procedure for obtaining a continuous distribution of transmissivity is that it uses the digital computer for those parts of the problem that can be translated into an algorithm, while allowing the introduction of judgments and insights into the problem as they are required.

NUMERICAL EXAMPLE

The Gauss optimization technique was used to identify the distribution of transmissivity for a model of the Whitewater River and Garnet Hill subbasins of the upper Coachella Valley ground-water basin (fig. 1). This study area covers about 400 km². Ground water occurs in unconsolidated and moderately indurated alluvial deposits (Dutcher and Bader, 1963; Farrell, 1964; Proctor, 1968; and Tyley, 1974), which have a saturated thickness of as much as 900 m (Biehler, 1964). A fault in the study area acts as a barrier to ground-water movement; hydraulic-head differentials of as much as 100 m occur across the fault. The effect of the fault on ground-water movement was represented in the model (fig. 2) by low transmissivity in the elements along the fault.

Prior to 1946, the hydraulic heads in the upper Coachella Valley ground-water basin were in quasi-equilibrium with natural recharge. The principal source of natural recharge was the percolation from the channels of ephemeral and intermittent streams that cross the ground-water basin (fig. 3). Optimum transmissivity was computed for this condition. Hydraulic head measurements and estimates of natural recharge were obtained from Tyley (1974).

The model domain was divided into 16 subdomains (fig. 4). Initially $T_{ij}^{\circ} = 4.2 \times 10^{-2}$ m²/s for all elements in subdomains 1, 2, 3, 4, 5, and 6; $T_{ij}^{\circ} = 9.3 \times 10^{-5}$ m²/s for all elements in subdomains 7, 8, 9, 10, and 11; and $T_{ij}^{\circ} = 7.0 \times 10^{-3}$ m²/s for all elements in subdomains 12, 13, 14, 15, and 16. The optimum surrogate parameters θ^* were computed (table 1), and the optimum transmissivities T_i^* were plotted on a map of the study area. This plot revealed discontinuities in the geographic distribution of transmissivity in areas where a continuous distribution reasonably would have been expected. Subjective adjustments were made to the transmissivities to remove these discontinuities. New values of θ^* were computed (table 1). The corresponding values of T_i^* were plotted, and the geographic distribution was examined for physical plausibility. These values (fig. 5) were accepted, and the parameter identification problem was completed.

Figure 6 shows the root-mean-square deviation of the computed hydraulic head from the measured hydraulic head, at each iteration in the initial optimization, for values of r equal to 0.1 (approaching a gradient step), 0.5, 1.0, and ∞ (the Gauss step). For $r = \infty$, the initial rate of convergence is rapid, but at some points the search diverges at successive iterates. For $r = 0.1$, convergence is slow, but stable. For $r = 1.0$, the most rapid rate to final convergence is exhibited; convergence is obtained after about 15 iterations.

An appropriate question to ask at this point is: How good are the estimates of transmissivity for this ground-water basin? If we had a perfect model and perfect knowledge of the prototype recharge and water levels, then the calibration procedure should yield a very close approximation of the true prototype transmissivity. However, our model is only an approximation of the prototype, and we have imperfect knowledge of the prototype recharge and of the corresponding prototype response. As a consequence, the transmissivity estimates probably deviate from the true transmissivity, which leads us back to the original question.

The approach that is often used is to examine the objective function value corresponding to the optimum parameters and also to examine the departure of the optimum parameters from subjective estimates of the optimum parameters. If the objective function and the parameter departures are within acceptable limits, the hydrologist accepts the parameter estimates. The inadequacy of this approach is chiefly in that it does not consider the effect of data errors on the parameter estimates.

A second approach that has been suggested, but seldom used in practice, is split-sample testing. Split-sample testing consists of organizing the available field data into two parts, each representing a different historical period. Optimum parameter estimates are computed using one part of the data. Then, the model is used to simulate conditions during the historical period represented by the second part of the data. The deviations of the computed hydraulic heads from the measured water levels for the second historical period are assumed to be a measure of the predictive reliability of the model.

In many applications split-sample testing is a powerful statistical tool for evaluating models. When used to evaluate ground-water models, however, this technique often fails to provide useful information. This failure occurs because the available field data seldom can be divided into two independent samples.

The point of the above discussion is that a methodology for evaluating the predictive reliability of ground-water models is still in the developing stages. Attention will focus here on what the author considers to be a first step in developing approaches for evaluating ground-water models--conditional evaluations based on an assumed probability model of the occurrence of errors in the independent variables. An unconditional evaluation would probably require the generation of additional information. Just how this information would be generated depends largely on the specifics of a particular situation. It might involve additional primary data collection, the use of independent parameter estimates from secondary data sources, or the application of subjective information through constrained regression, or through Bayesian estimation procedures.

EFFECTS OF DATA ERRORS

Measurements and sampling errors in the independent variables that are used to compute the optimum parameters are reflected in the deviation of the optimum parameters from the true parameters. If the optimum parameters are used in the model to make predictions of the response of the prototype to assumed future stresses, the predicted response will be in error. In the following paragraphs, a technique for analyzing these errors in the predicted response is demonstrated using the model of the upper Coachella Valley ground-water basin.

The upper Coachella Valley ground-water basin is being artificially recharged with imported water. Recharge operations began in 1973 when 9.2 hm^3 of water was recharged from percolation ponds in the northern part of the ground-water basin (fig. 1). Annual recharge will most likely be increased until a maximum annual recharge of 75 hm^3 is reached in about 1990. One effect of the recharge operation will be raised ground-water levels in wells. The maximum rise will occur when the hydraulic head in the ground-water basin reaches the steady-state values. Model predictions indicate that the maximum water-level rise, due only to the artificial recharge, will be about 350 m at the recharge site and about 180 m at Palm Springs, a major population center in the upper Coachella Valley (fig. 1). (Superimposed on the effect of artificial recharge is the effect of pumpage, which was not considered in the example.)

These water-level rises were computed from the model using the transmissivity estimates shown in figure 5. If different estimates of natural recharge had been used in the parameter-identification problem, different transmissivity estimates would have been obtained, and different predictions of the water-level response to artificial recharge would have been made. The independent variables of the parameter-identification problem are random variables, and transmissivity estimates and water-level predictions, which are functions of the independent variables, are also random variables. Conclusions about the effects on the modeling process of errors in the independent variables should be based on the statistical nature of the transmissivity estimates and of water-level predictions. A Monte Carlo simulation (Benjamin and Cornell, 1970, p. 124) of the modeling process provides a method for examining the statistical nature of these quantities, in spite of data limitations in the real world, which preclude a direct examination. The method is described stepwise as follows:

Step 1: Random estimates of the local recharge rate were generated. These estimates were considered to come from log-normal distributions with means equal to the estimates used in the original numerical example (fig. 3) and coefficient of variation equal to 25 percent. The assumed distribution of the local recharge estimates was derived heuristically and is undoubtedly situationally specific.

Step 2: Transmissivity estimates were computed using the new recharge estimates obtained in step 1. Except for the new recharge estimates, the configuration of the parameter identification problem was as in the original numerical example. Identical subdomains and measured water levels were used. The relative geographic distribution of the transmissivity within subdomains was the same as shown in figure 5.

Step 3: The model was used to compute steady-state hydraulic head for artificial recharge. The transmissivity estimates from step 2 were used in the model for this step.

Steps 1 through 3 were repeated 20 times, and the results are given in tables 2, 3, and 4. The range, mean, and coefficient of variation of the generated natural-recharge estimates are given in table 2 for each recharge site. With reference to the parameter-identification problem, the surrogate parameter estimates and the root-mean-square deviation of computed hydraulic heads from measured water levels are given in table 3 for each execution of step 2. The mean coefficient of variation of the surrogate parameter estimates are also given in table 3 for each surrogate parameter. With reference to the prediction problem, the water levels at the recharge site and at Palm Springs are given in table 4, for each execution of step 3. Additionally, the means and coefficients of variation of these water levels are given in table 4.

Initial values of surrogate parameters for the parameter-identification problem were, for each recharge data set, the values given in table 1, which gives the optimum surrogate parameters for the original numerical example. Although individual optimum-parameter values obtained from the Monte Carlo simulation deviate as much as 60 percent from initial values, the means of the various estimates of individual parameters are nearly identical to the values given in table 1. The coefficients of variation of these estimates range from 6 to 17 percent. The upper limit of this range is less than the coefficient of variation used in the generation of recharge estimates.

The magnitude of the deviation of the surrogate parameters tends to be related to the magnitude of the deviation of the discharge through the ground-water basin. The local discharge through the ground-water basin increases in the downstream direction, owing to the local contributions of geographically distributed recharge. By introducing random deviations in the local recharge estimates, random deviations are produced in the local discharge through the ground-water basin. Because of the effects of adding a series of independent random variables, however, the relative deviation of the discharge through the ground-water basin tends to decrease where the discharge results from about equal contributions from many recharge sites. Concomitantly, the surrogate parameters with the largest coefficient of variation tend to represent or be near areas with high local recharge. In these areas the local discharge through the ground-water basin includes mostly the discharge contribution from the local recharge, and the deviation of the local surrogate parameters results mostly from the deviation of a single recharge estimate.

On the basis of data in table 4, the standard deviation of computed hydraulic head resulting from variations in artificial recharge is about 27 m near the recharge site and is about 14 m near Palm Springs. The coefficient of variation of the computed hydraulic head is about 8 percent at each location. Furthermore, the coefficient of variation of computed hydraulic head is about the same as the average coefficient of variation of the surrogate parameters. But, as in the case of the surrogate parameters, the coefficient of variation of computed hydraulic head is smaller than the coefficient of variation of the recharge estimates. A point to be made here is that the root-mean-square deviations of measured water level from computed hydraulic head obtained from the parameter-identification step of the Monte Carlo simulation are all nearly zero, relative to the range of water levels that occur in the upper Coachella Valley ground-water basin (table 3). But the predicted hydraulic heads are distributed with a coefficient of variation of 8 percent. Therefore, the fit of the model to the water levels used in the parameter-identification problem is not in this case an indicator of the predictive reliability of the model.

A conditional answer can now be given to the earlier question: How good are the estimates of transmissivity for the upper Coachella Valley ground-water basin? The answer is given by the description of the distribution of surrogate parameter estimates. Given that the recharge estimates are independently distributed log-normal random variables with means equal to the recharge estimates used in the original numerical example and coefficients of variation equal to 25 percent, the transmissivity estimates are distributed with means and coefficients of variation as given in table 3. A conditional answer in terms of probability distributions can be given also to the more important question: How good are the predictions of hydraulic-head changes due to artificial recharge? Assuming that error in estimated ground-water recharge is the principal source of error in predicted hydraulic head, the estimated hydraulic-head changes near the recharge site and near Palm Springs are distributed with coefficients of variation of 8 percent. The 90-percent confidence interval for the hydraulic-head estimates is from 316 to 384 m near the recharge site and from 171 to 207 m near Palm Springs.

SUMMARY AND CONCLUSIONS

The Gauss optimization technique can be used to identify the parameters of a ground-water model. The success in identifying parameters, however, as measured by the fit of the model to the data used to identify parameters, is not a measure of the predictive reliability of the model. For 20 different random sets of recharge estimates, the Gauss optimization algorithm reduced the root-mean-square water-level residual to virtually zero. A zero root-mean-square water-level residual might incorrectly be interpreted as indicating nearly perfect parameter estimates. However, corresponding to each random set of recharge estimates was a different set of parameter estimates. Clearly all the sets of parameter estimates cannot be correct, and in this case, the root-mean-square water-level residual gives little insight into possible errors in predictions that might be made with the model.

A method for evaluating errors in the predictions of future water levels due to errors in recharge estimates was demonstrated. The method involves a Monte Carlo simulation of the parameter-identification problem and of the prediction problem. The steps in the method are: (1) To prescribe the distribution of the recharge estimates, (2) to use this distribution to generate random sets of recharge estimates, (3) to use the Gauss optimization technique to identify the corresponding set of parameter estimates for each set of recharge estimates, (4) to make the corresponding set of hydraulic-head predictions for each set of parameter estimates, and (5) to examine the distribution of hydraulic-head predictions and to draw appropriate conclusions. Similarly, the method can be used independently or simultaneously to estimate the effect on hydraulic-head predictions of errors in the measured water levels that are used in the parameter-identification problem.

Errors in hydraulic-head predictions result not only from errors in the recharge estimates but also from errors in the conceptual model of the ground-water basin. Errors resulting from the conceptual model are not treated by the Monte Carlo simulation. An assumption of the method is that errors in the conceptual model do not significantly affect the parameter-identification problem or the prediction problem.

REFERENCES CITED

- Biehler, S., 1964, Geophysical study of the Salton trough of southern California: Pasadena, California Institute of Technology, Ph. D. thesis, 139 p.
- Benjamin, J. R., and Cornell, C. A., 1970, Probability, statistics, and decision for civil engineers: New York, McGraw-Hill, 684 p.
- Bruch, J. C., Jr., Lam, C. H., and Simundich, T. M., 1974, Parameter identification in field problems: Water Resources Research, v. 10, no. 1, p. 73-79.
- Coats, K. H., Dempsey, J. R., and Henderson, H. H., 1970, A new technique for determining reservoir description from field performance data: Society of Petroleum Engineers Journal, v. 10, no. 1, p. 66-74.
- Dutcher, L. C., and Bader, J. S., 1963, Geology and hydrology of Agua Caliente Spring, Palm Springs, California: U.S. Geological Survey Water-Supply Paper 1605, 43 p.
- Emsellem, Y., and de Marsily, G., 1971, An automatic solution for the inverse problem: Water Resources Research, v. 7., no. 5, p. 1264-1283.
- Farrell, C. R., 1964, Coachella Valley investigation: California Department of Water Resources Bulletin 108, 145 p.
- Freeze, R. A., 1975, A stochastic-conceptual analysis of one-dimensional ground-water flow in nonuniform homogeneous media: Water Resources Research, v. 11, no. 5, p. 725-741.
- Frind, E. O., and Pinder, G. F., 1973, Galerkin solution of the inverse problem for aquifer transmissivity: Water Resources Research, v. 7, no. 5, p. 1397-1410.
- Hartley, H. O., 1961, The modified Gauss-Newton method for fitting of non-linear regression functions by least-squares: Technometrics, v. 3, no. 2, p. 269-280.
- Jacoby, S. L. S., Kowalils, J. S., and Piyyo, J. T., 1972, Iterative methods for nonlinear optimization problems: Englewood Cliffs, New Jersey, Prentice-Hall, 276 p.
- Jacquard, P., and Jain, C., 1965, Permeability distribution from field pressure data: Society of Petroleum Engineers Journal, v. 5, no. 4, p. 281-294.
- Jahns, H. O., 1966, A rapid method for obtaining a two-dimensional reservoir description from well response data: Society of Petroleum Engineers Journal, v. 6, no. 4, p. 315-327,
- Kleinecke, D., 1971, Use of linear programing for estimating geohydrologic parameters of ground-water basins: Water Resources Research, v. 7, no. 2, p. 367-374.
- Kruger, W. D., 1961, Determining areal permeability distributions by calculations: Journal of Petroleum Technology, v. 13, no. 6, p. 691-696.
- Levenberg, K., 1944, A method for the solution of certain non-linear problems in least squares: Quarterly Applied Mathematics, v. 2, p. 164-168.
- Lovell, R. E., 1971, Collective adjustment of the parameters of the mathematical model of a large aquifer: Tucson, Arizona University, Report 4, Department of Hydrology and Department of Systems Engineering, 87 p.
- Marquardt, D. W., 1963, An algorithm for least-squares estimation of nonlinear parameters: Journal of the Society for Industrial Applied Mathematics, v. 11, no. 2, p. 431-441.

- McLaughlin, D., 1974, Investigation of alternative procedures for estimating ground-water basin parameters: Walnut Creek, California, Water Resources Engineers, 185 p.
- Nelson, R. W., 1968, In place determination of permeability distributions for heterogeneous porous media through analysis of energy dissipation: Society of Petroleum Engineers Journal, v. 8, no. 1, p. 33-42.
- Nelson, R. W., and McCollum, W. L., 1969, Transient energy dissipation methods of measuring permeability distributions in heterogenous porous materials: U.S. Geological Survey Open-File Report, 77 p.
- Neuman, S. P., 1973, Calibration of distributed parameter ground-water flow models viewed as a multiple-objective decision process under uncertainty: Water Resources Research, v. 9, no. 4, p. 1006-1021.
- Pinder, G. H., and Frind, E. O., 1972, Application of Galerkin procedure to aquifer analysis: Water Resources Research, v. 8, no. 1, p. 108-120.
- Proctor, R. J., 1968, Geology of the Desert Hot Springs-Upper Coachella Valley area, California: California Division of Mines and Geology, Special Report 94, 50 p.
- Sager, B., 1973, Calibration and validation of aquifer models: Tucson, Arizona University Report 17, Department of Hydrology and Department of Systems Engineering, 175 p.
- Slater, G. E., and Durrer, E. J., 1971, Adjustment of reservoir simulation models to match field performance: Society of Petroleum Engineers Journal, v. 11, no. 3, p. 295-305.
- Stallman, R. W., 1956, Numerical analyses of regional water levels to define aquifer hydrology: American Geophysical Union Transcripts, v. 37, no. 4, p. 451-460.
- Tyley, S. J., 1974, Analog model study of the ground-water basin of the upper Coachella Valley, California: U.S. Geological Survey Water-Supply Paper 2027, 77 p.
- Vemuri, V., and Karplus, W. J., 1969, Identification of nonlinear parameters of ground-water basins by hybrid computation: Water Resources Research, v. 5, no. 1, p. 172-185.
- Wilde, D. J., and Beightler, C. S., 1967, Foundations of optimization: Englewood Cliffs, New Jersey, Prentice-Hall, 480 p.
- Yeh, W. W-G., and Tauxe, G. W., 1971, A proposed technique for identification of unconfined aquifer parameters: Journal of Hydrology, no. 12, p. 117-127.

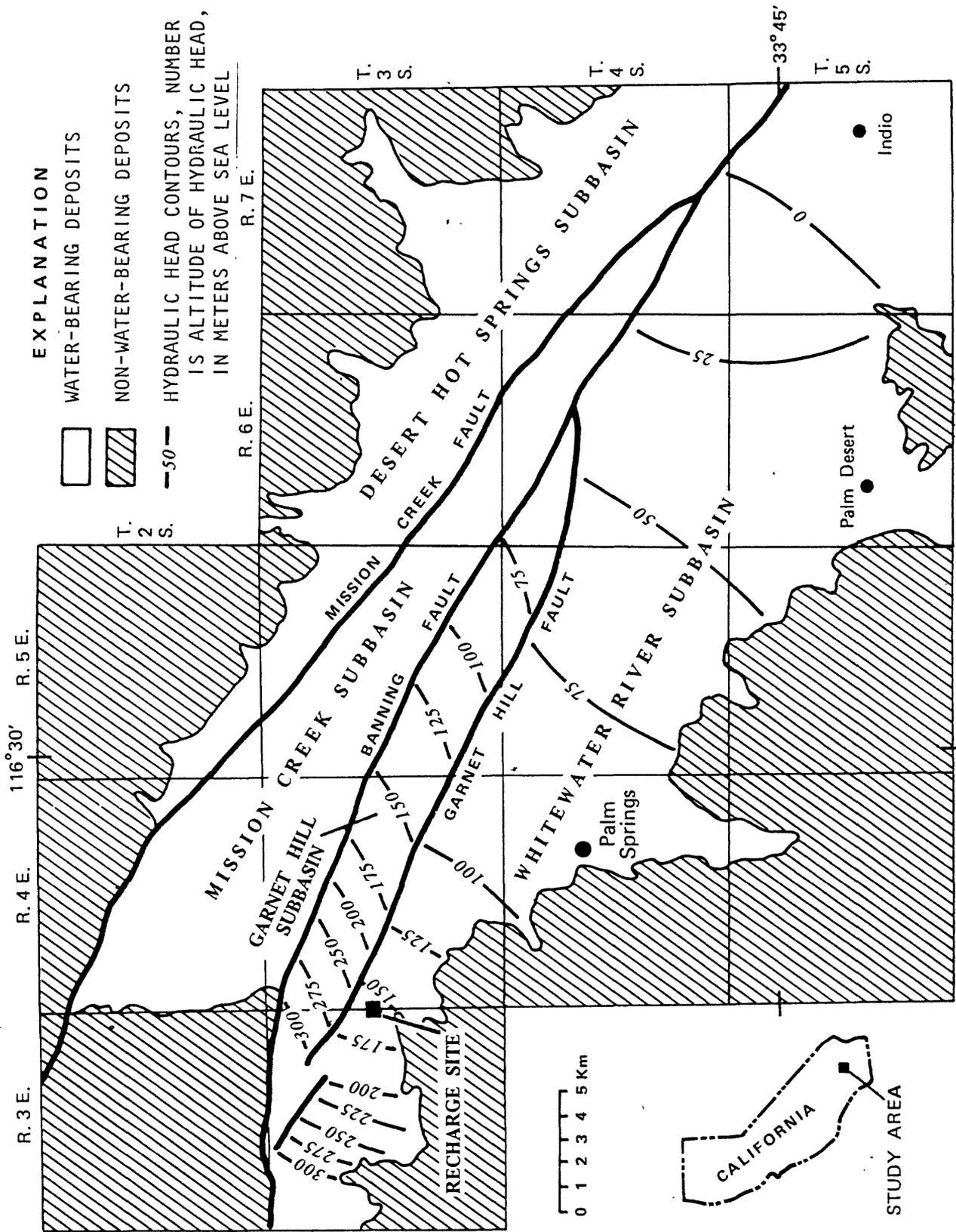


FIGURE 1. - Ground-water subbasins, generalized geology, and steady-state water-level contours for the upper Coachella Valley ground-water basin, California.

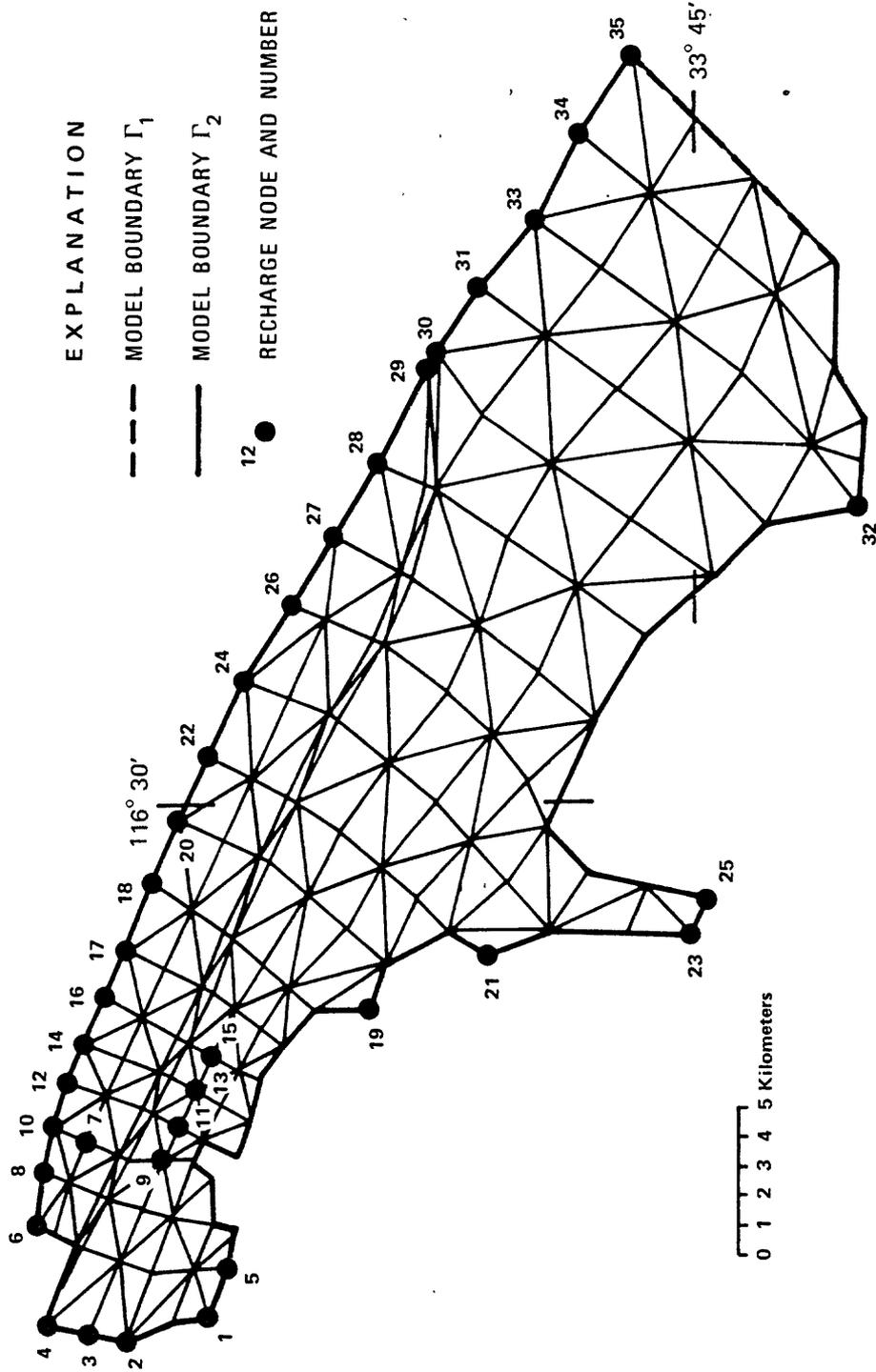


FIGURE 2. - Element layout for model of the Whitewater River and Garnet Hill subbasins.

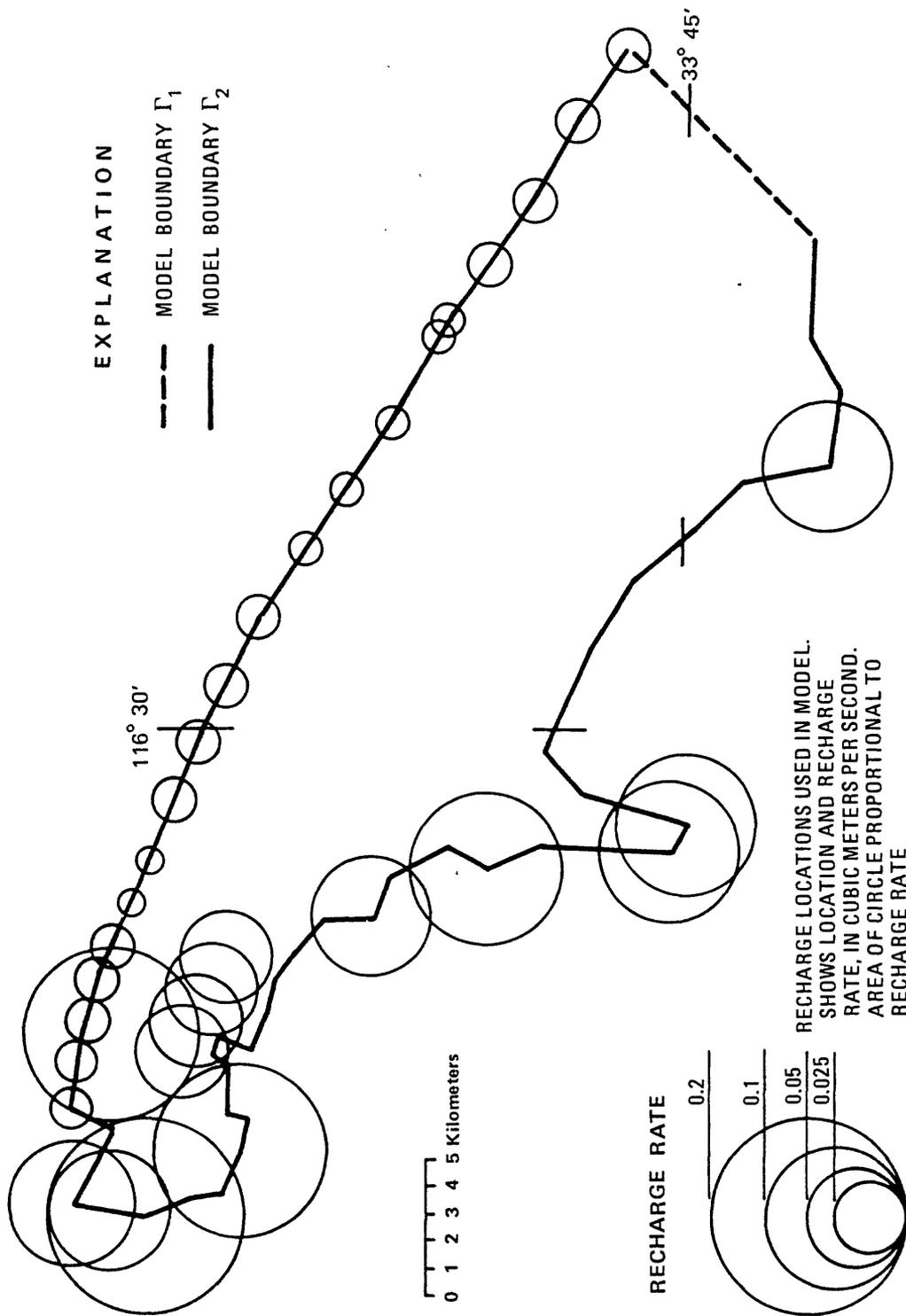


FIGURE 3.- Geographic distribution of natural recharge to Whitewater River and Garnet Hill subbasins.

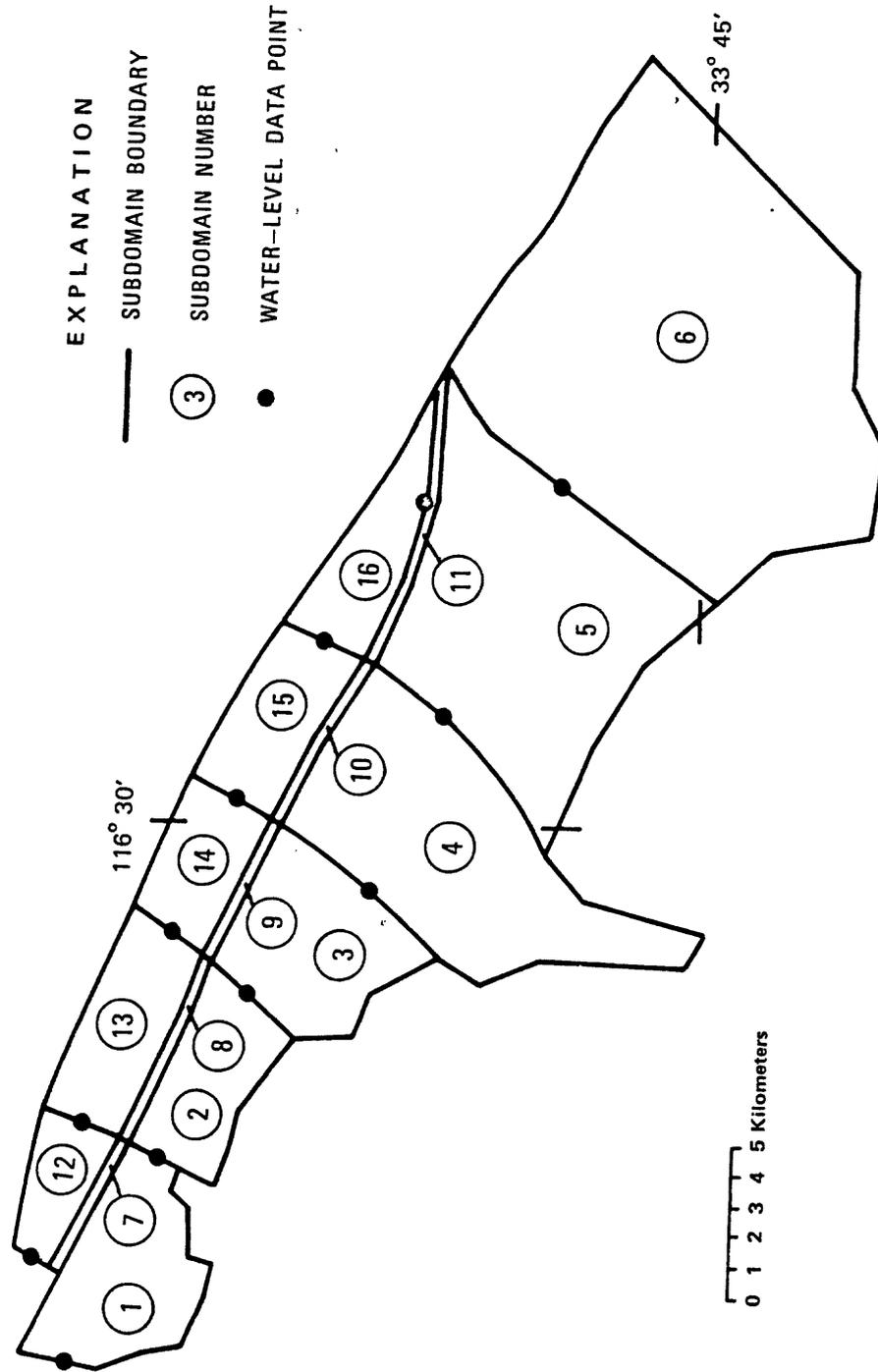


FIGURE 4. - Subdomains and water-level points used in computing optimum transmissivity from steady-state conditions.

EXPLANATION

- 50— LINE OF EQUAL TRANSMISSIVITY. NUMBER SHOWN IS 1000 TIMES THE TRANSMISSIVITY, IN METERS SQUARED PER SECOND
- ↔0.1↔ TRANSMISSIVITY OF A SEGMENT OF THE FAULT. NUMBER IS TRANSMISSIVITY, IN METERS SQUARED PER SECOND. ARROWS INDICATE THE EXTENT OF THE SEGMENT

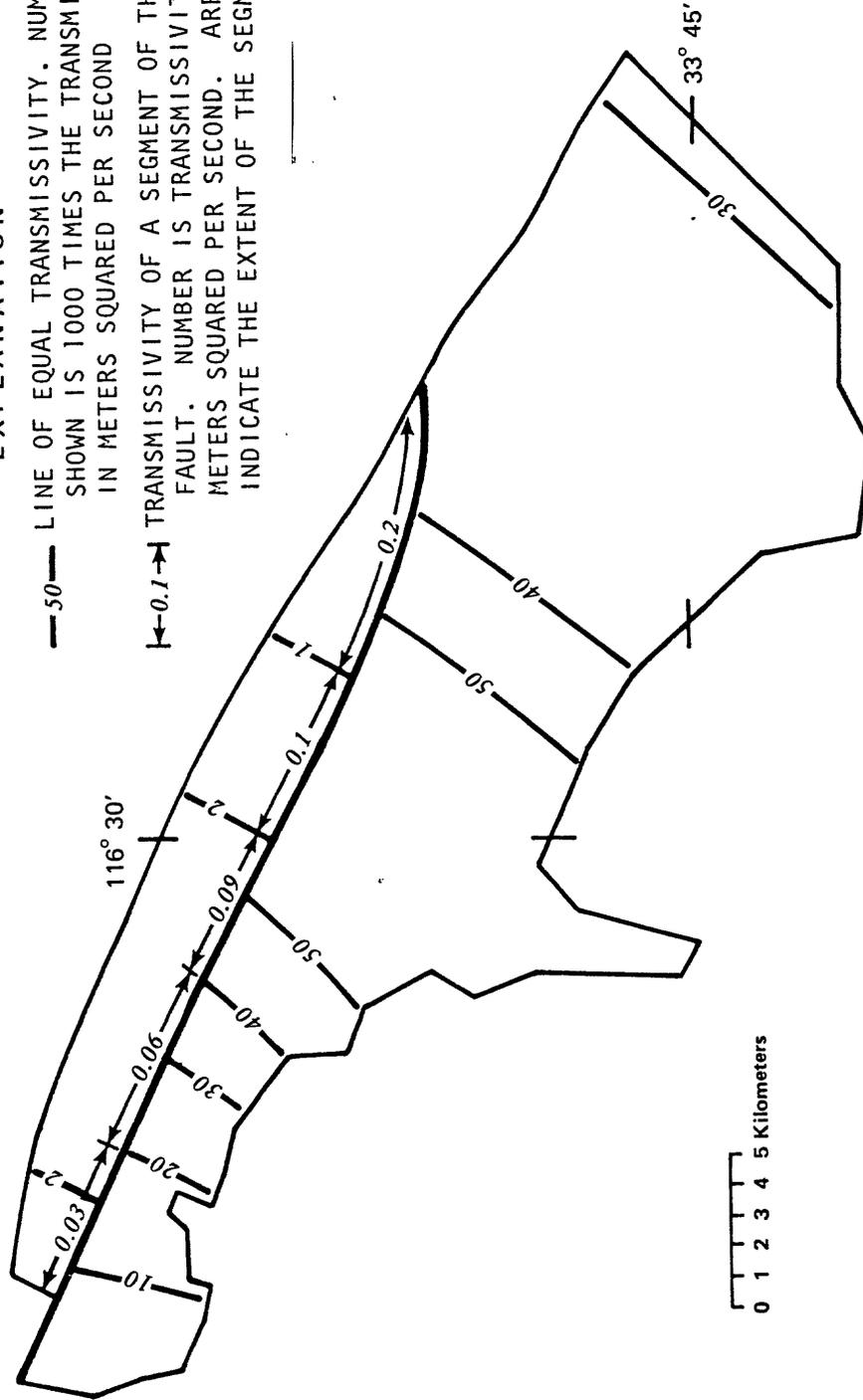


FIGURE 5. - Optimum transmissivity obtained by objectively adjusting large-scale variations of transmissivity and by subjectively adjusting small-scale variations of transmissivity.

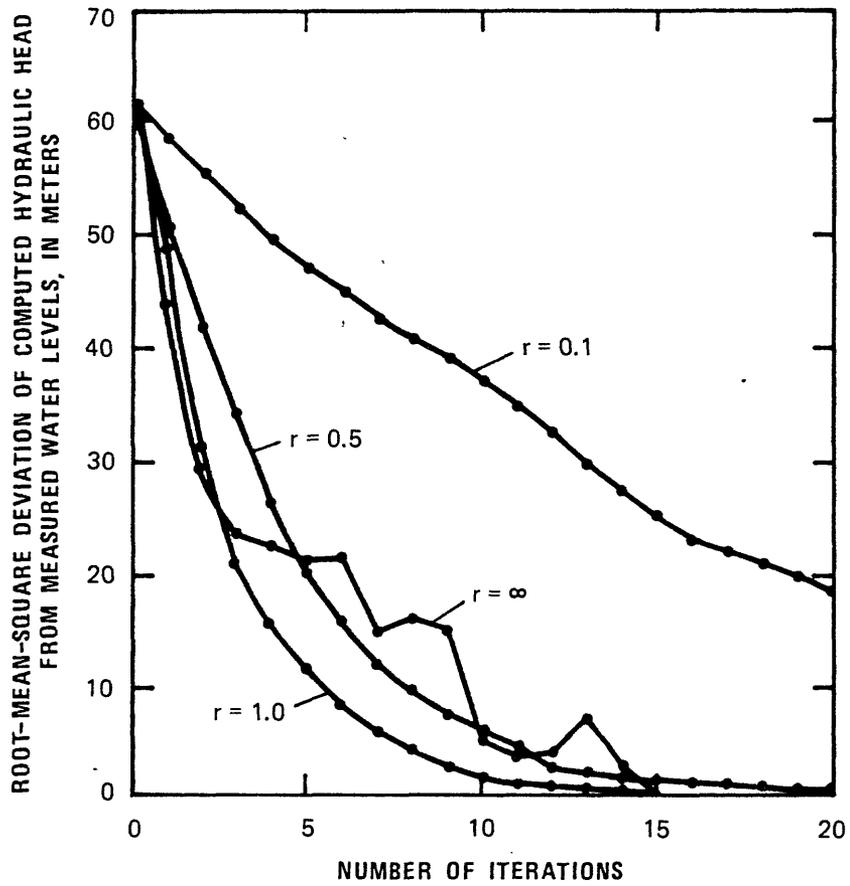


FIGURE 6. - Convergence of the Gauss optimization technique for different search radii, r.

Table 1. Optimum surrogate parameters computed from field data

Subdomain No. ¹	Optimum parameters (dimensionless)	
	First optimization ²	Second optimization ³
1	0.288	0.288
2	.679	.678
3	.953	1.16
4	1.76	1.34
5	1.01	1.01
6	.755	.755
7	.344	.347
8	.639	.648
9	.961	.992
10	1.20	1.25
11	1.86	1.85
12	.227	.276
13	.380	.377
14	.381	.371
15	.221	.222
16	.0553	.0560

¹Locations shown in figure 4.

²Initial value equaled 1.0.

³Initial value equaled optimum value from first optimization.

Table 2. Summary statistics on random recharge estimates

Natural recharge site No. ¹	Natural recharge (cubic meters per second)			
	Maximum	Minimum	Mean	Coefficient of variation
1	0.16	0.059	0.086	0.27
2	.29	.11	.20	.24
3	.10	.046	.077	.19
4	.12	.035	.074	.27
5	.10	.050	.078	.19
6	.012	.0047	.0090	.22
7	.26	.12	.17	.26
8	.016	.0051	.0096	.27
9	.078	.020	.043	.30
10	.015	.0064	.010	.26
11	.058	.026	.041	.23
12	.019	.0054	.011	.30
13	.086	.026	.046	.29
14	.018	.0072	.011	.30
15	.058	.028	.040	.20
16	.0061	.0024	.0034	.26
17	.0048	.0018	.0033	.21
18	.015	.0069	.010	.23
19	.11	.049	.075	.23
20	.010	.0051	.0086	.18
21	.18	.058	.11	.30
22	.015	.0048	.010	.26
23	.15	.068	.099	.24
24	.012	.0057	.0091	.18
25	.14	.047	.093	.24
26	.0092	.0037	.0058	.27
27	.0071	.0031	.0049	.24
29	.0055	.0035	.0047	.11
30	.0078	.0039	.0052	.20
31	.015	.0055	.010	.23
32	.13	.056	.088	.20
33	.016	.0051	.010	.28
34	.014	.0063	.010	.20
35	.012	.0066	.0095	.18

¹Locations shown in figure 2.

Table 3. Optimum surrogate parameter values (dimensionless) computed using random recharge estimates (table 2)

Recharge data set No.	Parameter number															Model fit ¹ (meters)	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		16
1	0.33	0.76	1.3	1.4	1.1	0.79	0.34	0.63	0.96	1.2	1.8	0.27	0.39	0.38	0.23	0.55	0.02
2	.32	.70	1.2	1.4	1.0	.75	.31	.59	.98	1.4	1.8	.25	.34	.34	.20	.060	.00
3	.30	.67	1.1	1.3	.94	.70	.27	.47	.83	1.1	2.0	.21	.32	.34	.22	.053	.07
4	.24	.59	1.0	1.2	.91	.69	.31	.58	.95	1.3	1.7	.24	.33	.34	.21	.060	.01
5	.30	.68	1.2	1.3	.97	.72	.32	.61	1.0	1.3	2.1	.25	.34	.34	.22	.055	.02
6	.33	.73	1.2	1.4	1.0	.76	.27	.48	.85	1.2	1.9	.20	.31	.34	.21	.055	.02
7	.27	.69	1.2	1.4	1.1	.79	.51	.97	1.3	1.4	1.8	.40	.51	.43	.25	.059	.27
8	.32	.71	1.3	1.4	1.1	.79	.27	.47	.83	1.2	2.1	.22	.32	.34	.21	.052	.03
9	.32	.76	1.2	1.4	1.0	.77	.30	.55	.94	1.3	1.8	.23	.32	.35	.21	.059	.01
10	.30	.68	1.3	1.4	1.0	.76	.30	.54	.92	1.2	1.8	.24	.34	.33	.22	.057	.00
11	.34	.75	1.2	1.5	1.2	.88	.41	.80	1.15	1.4	1.6	.32	.42	.39	.24	.060	.04
12	.29	.69	1.1	1.3	1.0	.74	.28	.51	.86	1.1	2.1	.24	.32	.35	.22	.051	.02
13	.22	.58	1.1	1.3	.97	.72	.54	.99	1.3	1.4	1.6	.41	.51	.43	.26	.060	.12
14	.24	.62	1.2	1.3	.95	.71	.42	.75	1.0	1.2	1.6	.34	.44	.42	.23	.058	.00
15	.29	.69	1.1	1.3	1.0	.76	.33	.61	.95	1.2	1.6	.27	.37	.37	.22	.059	.00
16	.27	.62	1.2	1.3	.98	.73	.35	.65	.97	1.2	1.8	.28	.38	.38	.22	.056	.00
17	.29	.68	1.1	1.4	1.0	.77	.31	.60	.99	1.3	1.9	.22	.34	.34	.23	.056	.00
18	.25	.60	1.1	1.3	.96	.72	.37	.68	.98	1.2	1.8	.30	.41	.40	.23	.054	.02
19	.26	.66	1.2	1.3	.98	.75	.51	.95	1.2	1.3	1.7	.39	.50	.45	.25	.057	.35
20	.31	.70	1.2	1.4	1.0	.78	.28	.51	.89	1.3	1.7	.23	.32	.33	.20	.061	.00
Mean	.29	.68	1.2	1.3	1.0	.76	.35	.65	.99	1.2	1.8	.28	.38	.37	.22	.057	
Coefficient of variation	.033	.053	.067	.070	.055	.041	.082	.16	.13	.077	.17	.064	.065	.038	.016	.0029	.070

¹Root-mean-square deviation of computed hydraulic head from measured water levels after six iterations.

Table 4. Hydraulic-head responses predicted with model using optimum surrogate parameters computed using random recharge estimates (table 3)

Recharge data set no.	Water level (meters)	
	Artificial recharge site	Palm Springs
1	317	176
2	344	189
3	374	208
4	406	217
5	363	200
6	338	188
7	331	177
8	328	176
9	320	180
10	342	183
11	291	152
12	350	192
13	385	201
14	379	204
15	344	187
16	372	197
17	343	183
18	383	202
19	357	193
20	332	180
Mean	350	189
Coefficient of variation	0.077	0.075