WATER RESOURCES RESEARCH, VOL. 23, NO. 1, PAGES 92–104, JANUARY 1987

# The Inverse Problem for Confined Aquifer Flow: Identification and Estimation With Extensions

HUGO A. LOAICIGA

*Department of Geological Sciences, Wright State University, Dayton, Ohio*

MIGUEL A. MARIÑO

*Department of Land, Air, and Water Resources and Department of Civil Engineering, University of California, Davis*

The contributions of this work are twofold. First, a methodology for estimating the elements of parameter matrices in the governing equation of flow in a confined aquifer is developed. The estimation techniques for the distributed-parameter inverse problem pertain to linear least squares and generalized least squares methods. The linear relationship among the known heads and unknown parameters of the flow equation provides the background for developing criteria for determining the identifiability status of unknown parameters. Under conditions of exact or overidentification it is possible to develop statistically consistent parameter estimators and their asymptotic distributions. The estimation techniques, namely, two-stage least squares and three stage least squares, are applied to a specific groundwater inverse problem and compared between themselves and with an ordinary least squares estimator. The three-stage estimator provides the closer approximaton to the actual parameter values, but it also shows relatively large standard errors as compared to the ordinary and two-stage estimators. The estimation techniques provide the parameter matrices required to simulate the unsteady groundwater flow equation. Second, a nonlinear maximum likelihood estimation approach to the inverse problem is presented. The statistical properties of maximum likelihood estimators are derived, and a procedure to construct confidence intervals and do hypothesis testing is given. The relative merits of the linear and maximum likelihood estimators are analyzed. Other topics relevant to the identification and estimation methodologies, i.e., a continuous-time solution to the flow equation, coping with noise-corrupted head measurements, and extension of the developed theory to nonlinear cases are also discussed. A simulation study is used to evaluate the methods developed in this study.

## 1. INTRODUCTION

The estimation of transmissivities, storativities, and other groundwater parameters has received substantial attention in the water resources literature. Several of such approaches to the inverse problem can be found in previous studies by *McLaughlin* [1975], *Cooley* [1977, 1979, 1982], *Neuman and Yakowitz* [1979], *Yakowitz and Duckstein* [1980], *Neuman* [1980], *Yeh and Yoon* [1981], *Yeh et al.* [1983], *Kitanidis and Vomvoris* [1983], *Aboufirassi and Mariño* [1984], *Sadeghipour and Yeh* [1984], *Hoeksema and Kitanidis* [1984], and *Carrera and Neuman* [1986*a*, *b*, *c*], among others. Some problematic aspects of groundwater parameter estimation related to numerical instability, noisy observations, and nonuniqueness are discussed in the works by *Neuman and Yakowitz* [1979] and *Yakowitz and Duckstein* [1980] and have been recognized and reported by several other authors (see, for example, *Yeh et al.* [1983]).

This study has two main objectives: (1) to develop an analytic criterion to establish the identifiability status of the inverse problem for confined groundwater flow and (2) to present linear (i.e., least squares) and maximum likelihood estimation methods for the solution of the inverse problem and to derive the statistical properties of such estimators. A simulation study is used to evaluate the theory developed in this study and to compare the alternative estimation techniques.

The contributions of this study can be summarized as follows.

1. An analytical interpretation of identifiability in the inverse problem based on the relationships between available information and the unknown parameters is developed.
2. Least squares and maximum likelihood methods for groundwater parameter estimation are then presented.
3. Alternative parameter estimators are analyzed on the basis of their statistical properties, i.e., consistency and asymptotic distributions.
4. A unified theory of parameter identifiability, estimation, and statistical properties, is developed.

The remainder of this paper is organized as follows: section 2 contains a description as well as background to the problem; section 3 discusses the identifiability status of the inverse problem for confined aquifer flow; section 4 presents the two-stage least squares estimation technique; section 5 contains the three-stage least squares method for parameter estimation; maximum likelihood estimation is developed in section 6; applications are given in section 7; section 8 contains a summary and discussion of the findings of this study; and section 9 outlines future research needs, e.g., extensions of the theory to unsaturated flow, mass transport, and noisy head observations.

## 2. BACKGROUND AND DESCRIPTION OF THE PROBLEM

### 2.1. *Confined Aquifer Flow: Continuous and Discrete Forms*

The basic equation used in this study is that which describes flow in a heterogeneous and isotropic confined aquifer:

$$\frac{\partial}{\partial x}\left(T\,\frac{\partial \phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(T\,\frac{\partial \phi}{\partial y}\right)+F=S\,\frac{\partial \phi}{\partial t} \tag{1}$$



Paper number 5W4320.
0043-1397/87/005W-4320$05.00

in which $\phi$ denotes the piezometric head (units, $L$); $T = T(x, y)$ denotes transmissivity (units, $L^2/T$); $S = S(x, y)$ represents storativity (dimensionless); and $F$ denotes either a distributed (units, $L/T$) or point (units, $L^3/T$) sink/source, or a combination of both throughout the flow domain. Notice that if $F$ is to represent a point sink/source, it must be multiplied by $\delta(x - x_i, y - y_i)$, the Dirac delta function, in which $(x_i, y_i)$ denotes the location of the sink/source.

Equation (1) is discretized using finite-element methods and expressed as a linear system of differential equations as follows:

$$A^* \boldsymbol{\phi} + B^* \frac{d\boldsymbol{\phi}}{dt} + \mathbf{F}^* + \mathbf{0} \qquad (2)$$

in which the elements of the conductance (or stiffness) matrix $A^*$, the capacity matrix $B^*$, and vector $\mathbf{F}^*$ are obtained by assembling the element equations according to the node numbering selected when the flow domain is divided into a number of subregions or finite elements; the head vector $\boldsymbol{\phi}$ is of size $N \times 1$, where $N$ is the total number of nodal heads in the flow domain; the vector of sinks and sources $\mathbf{F}^*$ is of size $N \times 1$ (although many or even all of its components could be zero, depending upon the sink/source distribution); and the banded matrices $A^*$ and $B^*$ are both of dimension $N \times N$.

The finite-element model has been chosen to form (2) for several reasons: (1) it permits division of the flow domain into subregions over which $T$ and $S$ can be assumed constant, and hence the spatial variability of such parameters is conveniently handled; (2) boundary conditions, in particular flux-type boundary conditions, are easy to manipulate; and (3) the continuous form of (2) can be used, as is shown below, to the derive closed-form solution for $\boldsymbol{\phi}(t)$. Other numerical methods, e.g., finite differences, could also be used as plausible alternatives to the finite-element method.

If boundary conditions are of the flux or mixed type, they would be included automatically into $\mathbf{F}^*$ in (2) by the finite-element method. If the boundary conditions are only prescribed heads, then it is convenient to partition $\boldsymbol{\phi}$ and $\mathbf{F}^*$ into $\boldsymbol{\phi}^T = [\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T]$ and $\mathbf{F}^{*T} = [\mathbf{F}_1^{*T}, \mathbf{F}_2^{*T}]$, in which subscripts 1 and 2 denote the subvectors of prescribed and unknown heads, respectively. Equation (2) can be expressed in partitioned form as

$$\begin{bmatrix} A_{11}^* & A_{12}^* \\ A_{21}^* & A_{22}^* \end{bmatrix}\begin{bmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{bmatrix} + \begin{bmatrix} B_{11}^* & B_{12}^* \\ B_{21}^* & B_{22}^* \end{bmatrix}\begin{bmatrix} \dot{\boldsymbol{\phi}}_1 \\ \dot{\boldsymbol{\phi}}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{F}_1^* \\ \mathbf{F}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{0}_1 \\ \mathbf{0}_2 \end{bmatrix} \qquad (3)$$

Of interest for the developments on identifiability and estimation in this study is the lower subsystem in (3), i.e.,

$$A_{22}^* \boldsymbol{\phi}_2 + B_{22}^* \dot{\boldsymbol{\phi}}_2 + \mathbf{F}_2^* + A_{21}^* \boldsymbol{\phi}_1 + B_{21}^* \dot{\boldsymbol{\phi}}_1 = \mathbf{0}_2 \qquad (4)$$

in which $A_{22}^*$ and $B_{22}^*$ are of size $G \times G$ ($G$ is the dimension of the unknown subvector $\boldsymbol{\phi}_2$); $\boldsymbol{\phi}_2$, $\mathbf{F}_2^*$, and $\boldsymbol{\phi}_1$ are of dimension $G \times 1$, $G \times 1$, and $(N - G) \times 1$, respectively; and $A_{21}^*$ and $B_{21}^*$ are both of size $G \times (N - G)$. For the sake of simplicity in notation, the continuous time linear system in (4) is written as

$$A\boldsymbol{\phi} + B\dot{\boldsymbol{\phi}} + \mathbf{F} = \mathbf{0} \qquad (5)$$

in which $A = A_{22}^*$, $B = B_{22}^*$, $\boldsymbol{\phi} = \boldsymbol{\phi}_2$, $\mathbf{0} = \mathbf{0}_2$, and $\mathbf{F} = \mathbf{F}_2^* + A_{21}^* \boldsymbol{\phi}_1 + B_{21}^* \dot{\boldsymbol{\phi}}_1$; i.e., $\boldsymbol{\phi}$ will denote the $(G \times 1)$ vector of unknown nodal heads, and the matrices $A$ and $B$, as well as the vector $\mathbf{F}$, are appropriately defined after the boundary conditions are properly introduced. Notice that when dealing with flux or mixed type boundary conditions $G$ equals $N$, and the fluxes must be adequately specified in vector $\mathbf{F}^*$ of (2).

It is of relevance for the applications given in section 7 that (5) be explicitly solved for $\boldsymbol{\phi}(t)$, for any time $t$, given an initial condition $\boldsymbol{\phi}(0)$ and known matrices $A$, $B$, and vector $\mathbf{F}$. Such analytic expression is easily derived if (5) is premultiplied by $B^{-1}$ and expressed as

$$\dot{\boldsymbol{\phi}} = C\boldsymbol{\phi}(t) + D\mathbf{F}(t) \qquad \boldsymbol{\phi}(0) = \boldsymbol{\phi}_0 \qquad (6)$$

in which $C = -B^{-1}A$ and $D = -B^{-1}$. In (6), $\mathbf{F}(t)$ takes the role of being a vector of inputs (outputs), where inputs (outputs) can be sources (sinks) and/or the effect of boundary conditions (since boundary conditions, to some extent, govern the groundwater flow). From basic theory of linear systems (see, for example, *Polak and Wong* [1970, pp. 9–17]),

$$\boldsymbol{\phi}(t) = e^{tC}\left[\boldsymbol{\phi}_0 + \int_0^t e^{-\tau C} D\mathbf{F}(\tau)\, d\tau\right] \qquad (7)$$

in which the matrix $e^{tC}$ can be expressed as a function of $t$ by the method of interpolating polynomials (see, for example, *Gantmacher* [1959, pp. 95–129]). Equation (7) will be used in section 7.

The continuous system given in (5) can be discretized as follows:

$$A[\omega\boldsymbol{\phi}_t + (1 - \omega)\boldsymbol{\phi}_{t-1}] + B\,\frac{\boldsymbol{\phi}_t - \boldsymbol{\phi}_{t-1}}{\Delta t} + \mathbf{F}_t^* = \mathbf{u}_t^* \qquad (8)$$

which upon rearrangement becomes

$$\left(\frac{B}{\Delta t} + \omega A\right)\boldsymbol{\phi}_t + \left[A(1 - \omega) - \frac{B}{\Delta t}\right]\boldsymbol{\phi}_{t-1} + \mathbf{F}_t^* = \mathbf{u}_t^* \qquad (9)$$

$$t = 1, 2, \cdots, n$$

in which $0 \le \omega \le 1$ (equation (9) is unconditionally valid, i.e., stable and convergent, for $\frac{1}{2} \le \omega \le 1$) and was set equal to $\frac{1}{2}$ in this study; $\Delta t$ is the selected time step for the simulation of groundwater flow; $n$ is the number of simulation periods; $\mathbf{F}_t^* = \omega\mathbf{F}_t + (1 - \omega)\mathbf{F}_{t-1}$; and $\mathbf{u}_t^*$ is a $G \times 1$ error vector that accounts for modeling errors in approximating the physical process of groundwater flow by (1), measurement errors, etc.

It is assumed that

$$\mathbf{u}_t^* = \rho\mathbf{u}_{t-1}^* + \mathbf{u}_t \qquad (10)$$

in which

$$E(\mathbf{u}_t) = \mathbf{0} \qquad (11)$$

$$E(\mathbf{u}_t\mathbf{u}_s^T) = [\sigma_{ij}]_{G \times G}\delta_{ts} = \Sigma\, \delta_{ts} \qquad (12)$$

in which $\sigma_{ij}$, $1 \le i, j \le G$ are the elements of $\Sigma$. Equation (10) states that the error term follows an autoregressive time pattern and hence is autocorrelated over time. The unknown scalar coefficient $\rho$ is another parameter in the inverse problem; $\Sigma$ is a $G \times G$ covariance matrix whose elements are denoted by $\sigma_{ij}$, $1 \le i, j < G$; $\delta_{ts}$ is a Kronecker delta; and $\mathbf{u}_t$ is a white noise term. An equivalent error structure to that given in (10) was successfully used by *Carrera and Neuman* [1986*a*, *b*, *c*] in an indirect approach to the inverse problem via maximum likelihood.

### 2.2. *Notational Convention for the Discrete Flow Equation*

The criteria for identifiability set forth in section 3 require the introduction of special notation. Equation (9), combined

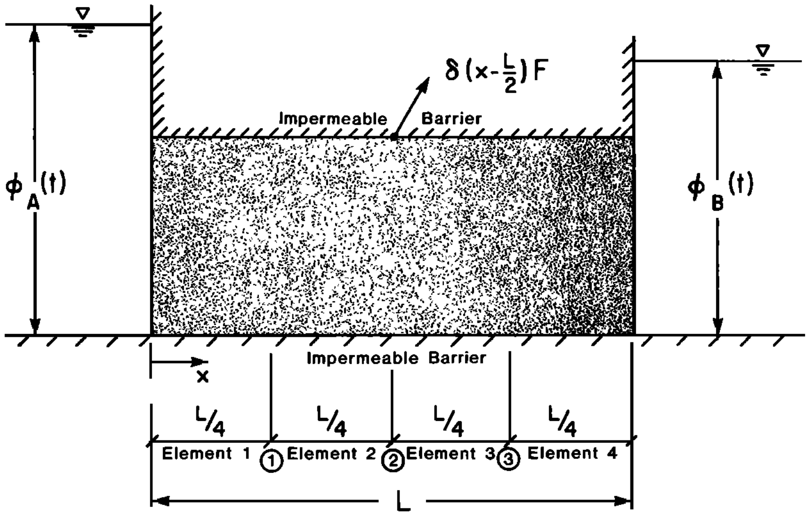


Fig. 1. Confined aquifer subject to time-dependent boundary conditions and a discharge (of units $L^3\ T^{-1}\ L^{-1}$) at $x = L/2$.

with (10), yields

$$\underset{(G \times G)}{\Psi}\ \underset{(G \times 1)}{\boldsymbol{\phi}_t} + \underset{(G \times K)}{\Gamma}\ \underset{(K \times 1)}{\mathbf{x}_t} = \underset{(G \times 1)}{\mathbf{u}_t} \qquad t = 1, 2, \cdots, n \tag{13}$$

in which

$$\Psi = \frac{B}{\Delta t} + \omega A \tag{14}$$

$$\Gamma = \left[ A(1 - \omega) - \frac{B}{\Delta t},\ M,\ -\rho I_{G \times G} \right] \tag{15}$$

$$\mathbf{x}_t^T = [\boldsymbol{\phi}_{t-1}^T, \mathbf{p}_t^T, \mathbf{u}_{t-1}^{*T}] \tag{16}$$

with $\boldsymbol{\phi}_0$ assumed known, in which $M$ and $\mathbf{p}_t$ are a matrix and vector, respectively, determined by the nature of sink/sources and the boundary conditions over the flow domain (see example below). In (16), $\mathbf{u}_0^*$, the error at time zero, is assumed to be equal to zero. Notice that one has to estimate the matrices $\Psi$ and $\Gamma$ (and $\Sigma$) by first assuming that $\rho$ in (10) is zero, and with those estimates generate a sequence of the error term $\mathbf{u}_t^*$, $t = 1, 2, \cdots, n$, via (13). The generated sequence of $\mathbf{u}_t^*$'s is used in (16) to reestimate $\Psi$, $\Gamma$, and $\Sigma$ (with $\rho \neq 0$). It might be necessary to iterate this procedure, although the application in section 7 required only one iteration to satisfy an adequate convergence criterion for the elements of $\Psi$ and $\Gamma$. It is also assumed that the time step $\Delta t$ in (14) and (15) is constant. This assumption is quite appropriate for actual applications in which the data available for the calibration of groundwater parameters are spaced at equal time steps, or can be adapted to a uniform time interval. Since the methods proposed in sections 4–6 do not require the simulation of the groundwater flow equations for the purpose of estimation (i.e., the methods of this paper are of the direct type, as opposed ot the indirect methods), setting $\Delta t$ = const does not lead to any difficulties in their implementation. Clearly, when using indirect methods to solve the inverse problem, it is necessary to solve the entire finite-element equations at each iteration in the search for optimal parameter values, and thus it might be more appropriate to let the time step vary in those methods. The authors are conducting a project to calibrate storativities and transmissivities in the San Joaquin Valley, California, in which the data are available at intervals of 6 months (i.e., $\Delta t = 6$ months). In large-scale applications of this type it is usually convenient and appropriate to set $\Delta t$ = constant.

To illustrate the validity of the terms and the dimensions appearing in (13), the aquifer shown in Figure 1 is used as an example ($\rho$ is set equal to zero for simplicity). The aquifer is confined, of length $L$, subject to time-varying prescribed heads at $x = 0$ and $x = L$; a discharge takes place at a rate $F$ at point $x = L/2$. The flow is one dimensional, in the $x$ direction only. Without loss of generality, a one-dimensional example has been chosen to illustrate the developments of this paper. The properties of the aquifer are its transmissivity $T$ and storativity $S$ (both unknown). The aquifer is divided into four elements each of size $L/4$.

By using linear interpolation functions to form the $(2 \times 2)$ element matrices (their elements can be read directly, e.g., from equations (7) and (8) of *Pinder and Frind* [1972] after dropping the $y$ terms and matching other terms), after assembling the matrices, and condensing out the prescribed heads, one obtains $(B/\Delta t + \omega A)$ = tridiagonal $[\psi_{ij}]_{3 \times 3}$, in which

$$\psi_{ii} = \frac{4\omega}{L}[T^{(i)} + T^{(i+1)}] + \frac{L}{12\Delta t}[S^{(i)} + S^{(i+1)}] \qquad i = 1, 2, 3 \tag{17}$$

$$\psi_{i,i+1} = -\frac{4\omega}{L}[T^{(i+1)}] + \frac{L}{24\Delta t}[S^{(i+1)}] \qquad i = 1, 2 \tag{18}$$

$$\psi_{i,i-1} = \psi_{i-1,i}\ \text{(symmetry)} \qquad i = 2, 3 \tag{19}$$

where $T^{(j)}$ and $S^{(j)}$ denote the values of transmissivity and storativity within the $j$th element ($j = 1, 2, 3, 4$, as seen in Figure 1). The elements of the (symmetric and tridiagonal) matrix $(A(1 - \omega) - B/\Delta t)$ are given by (17)–(19) with $\omega$ replaced by $1 - \omega$ and the plus sign between bracketed terms changed to a minus sign. The $\mathbf{F}_t^*$ vector of (9) is given by

$$\mathbf{F}_t^* = \begin{bmatrix} 0 \\ \bar{F} \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{4}{L} T^{(1)} \bar{\phi}_A \\ 0 \\ -\frac{4}{L} T^{(4)} \bar{\phi}_B \end{bmatrix} + \begin{bmatrix} \frac{L}{24} S^{(1)} \bar{\phi}_A \\ 0 \\ \frac{L}{24} S^{(4)} \bar{\phi}_B \end{bmatrix} \tag{20}$$

in which $\bar{F} = \omega F(t) + (1 - \omega)F(t - 1)$ is the averaged discharge (see Figure 1) at $x = L/2$; $\bar{\phi}_A = \omega\phi_A(t) + (1-\omega)\phi_A(t-1)$; $\bar{\bar{\phi}}_A = [\phi_A(t) - \phi_A(t-1)]/\Delta t$; and similarly for $\bar{\phi}_B$ and $\bar{\bar{\phi}}_B$.

From (17)–(20), $\Psi$, $\Gamma$, $\boldsymbol{\phi}_t$, $\mathbf{x}_t$ and $\mathbf{u}_t$ of (13) for the illustrative example are given by

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} & 0 \\ \psi_{21} & \psi_{22} & \psi_{23} \\ 0 & \psi_{32} & \psi_{33} \end{bmatrix} \tag{21a}$$

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 & \cdot & 0 & \gamma_{15} & 0 & \gamma_{17} & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \cdot & 1 & 0 & 0 & 0 & 0 \\ 0 & \gamma_{32} & \gamma_{33} & \cdot & 0 & 0 & \gamma_{36} & 0 & \gamma_{38} \end{bmatrix}$$

$$\boldsymbol{\phi}_t^T = [\phi_1(t), \phi_2(t), \phi_3(t)]; \qquad \mathbf{u}_t^T = [u_1(t), u_2(t), u_3(t)] \tag{21b}$$

$$\mathbf{x}_t^T = [\phi_1(t-1), \phi_2(t-1), \phi_3(t-1), \bar{F}, \bar{\phi}_A, \bar{\phi}_B, \bar{\bar{\phi}}_A, \bar{\bar{\phi}}_B] \tag{21c}$$

$$\mathbf{p}_t^T = [\bar{F}, \bar{\phi}_A, \bar{\phi}_B, \bar{\bar{\phi}}_A, \bar{\bar{\phi}}_B] \tag{21d}$$

Thus for the example of Figure 1: $G = 3$, $K = 8$, and $K - G = 5$. The elements $\gamma_{15}$, $\gamma_{36}$, $\gamma_{17}$, and $\gamma_{38}$ of the matrix $\Gamma$ are the coefficients multiplying $\bar{\phi}_A$, $\bar{\phi}_B$, $\bar{\bar{\phi}}_A$, and $\bar{\bar{\phi}}_B$ in (20), respectively. After this example, the general notation presented in (13) will describe the time evolution and spatial distribution of confined aquifer flow in either one or two dimensions. Notice that in the notation of (13) the dependent variable $\boldsymbol{\phi}_t$ is a function of the sinks (sources) (i.e., $\bar{F}$), boundary conditions ($\phi_A$, $\phi_B$), and is conditioned on the previous realization $\boldsymbol{\phi}_{t-1}$.

In principle, the coefficient matrices $\Psi$ and $\Gamma$ could be full, although it is known that the groundwater flow equation shows a banded matrix $\Psi$, and the leading submatrix (of size $G \times G$) in $\Gamma$ is also banded. The presence of many zeros in $\Psi$ and $\Gamma$ simplifies the task of estimating their nonzero elements. The unknown covariance matrix $\Sigma$ (see equation (11)) must also be estimated. In summary, the estimation problem consists of obtaining statistically consistent estimators of $\Psi$, $\Gamma$, and $\Sigma$, based on data $\boldsymbol{\phi}_t$ and $\mathbf{x}_t$, $t = 1, 2, \cdots, n$. Recall that an estimator $\hat{\theta}_n$ of an unknown but constant parameter $\tilde{\theta}$ is statistically consistent if

$$\lim_{n\to\infty} P[|\hat{\theta}_n - \tilde{\theta}| > \xi] = 0 \qquad \forall\xi > 0$$

or, equivalently, if $\tilde{\theta}_n$ coincides with the limiting distribution of $\hat{\theta}_n$, i.e., if

$$\operatorname*{plim}_{n\to\infty} \hat{\theta}_n = \tilde{\theta}$$

in which $n$ is the sample size used to compute $\hat{\theta}_n$ (see, for example, *Rao* [1973] or *Bickel and Doksum* [1977, pp. 132–133]). Consistency is a sine qua non property of any estimator: an inconsistent estimator is necessarily based on an incomplete and almost surely faulty theory.

## 3. Conditions for Identification

The identifiability criteria set forth in this section apply to the linear estimation methods of sections 4 and 5. The interest is in the matrices $\Psi$ and $\Gamma$ in (13) that govern the time evolution of piezometric heads and in the covariance matrix $\Sigma$ (see equation (12)). Once $\Psi$ and $\Gamma$ are available, one can simulate, or predict, the response of the aquifer to natural and/or artificial inputs. There are many situations that justify the estimation of the nonzero elements of $\Psi$ and $\Gamma$, as opposed to estimating physical parameters such as transmissivity and storativity. One such situation occurs when there is scarce information about the geological structure and physical properties of an aquifer. It may then be difficult to specify a zonation of the aquifer, by either finite elements or finite differences, that would realistically approximate the actual spatial variability of aquifer properties. In such an instance, the calibration of $\Psi$ and $\Gamma$ with available piezometric head data will provide the necessary information for the simulation of the aquifer response, and the use of such response information in planning studies. For a direct estimate of parameters such as transmissivity and storativity, the reader is referred to section 6, where a nonlinear maximum likelihood method is given.

The problem of identification is that of being able to determine all the (nonzero) elements of the matrices $\Psi$ and $\Gamma$, as well as the covariance matrix $\Sigma$. Notice that we look for estimators of $\psi_{ij}$, $\gamma_{ij}$, and $\sigma_{ij}$ (i.e., the elements of the matrices $\Psi$, $\Gamma$, and $\Sigma$, respectively).

By premultiplying (13) by $\Psi^{-1}$ and solving for $\boldsymbol{\phi}_t$ one obtains

$$\boldsymbol{\phi}_t = (-\Psi^{-1}\Gamma)\mathbf{x}_t + (\Psi^{-1}\mathbf{u}_t) = \Pi\,\mathbf{x}_t + \mathbf{e}_t \qquad t = 1, 2, \cdots, n \tag{22}$$

in which $E(\mathbf{e}_t\mathbf{e}_t^T) = \Psi^{-1}\Sigma(\Psi^{-1})^T$. As is shown in section 4, the $G \times K$ (full and unknown) matrix $\Pi$ can be estimated consistently by ordinary least squares (OLS). The identification problem can then be stated as follows: given a consistent estimator of $\Pi$ $(= -\Psi^{-1}\Gamma)$, is it possible to estimate (consistently) $\Psi$ and $\Gamma$, and if so, are the estimators unique? It is important to point out that if $\Psi$ and $\Gamma$ can be estimated, then immediately one has all the information required to simulate the discretized groundwater flow equations (9) or (13). The estimated covariance $\Sigma$ is useful to determine the properties of the estimators for $\Psi$ and $\Gamma$.

### 3.1. *The Identifiability Problem*

Given that

$$\underset{(G\times K)}{\Pi} = \underset{(G\times G)}{-\Psi^{-1}}\;\underset{(G\times K)}{\Gamma} \tag{23}$$

and that $\Pi$ can be found independently of $\Psi$ and $\Gamma$ (see section 4) it follows that

$$\Psi\,\Pi = -\Gamma \tag{24}$$

The $j$th row ($j = 1, 2, \cdots, G$) of (24) can be written as

$$(\psi_{j1} \quad \psi_{j2} \quad \cdots \quad \psi_{jG})\Pi = -(\gamma_{j1} \quad \gamma_{j2} \quad \cdots \quad \gamma_{jK}) \tag{25}$$

As it was shown in the example of subsection 2.2, some of the elements $\psi_{jk}$ and $\gamma_{jl}$ ($k = 1, 2, \cdots, G$; $l = 1, 2, \cdots, K$) are equal to zero.

The elements of the right- and left-hand side vectors of (25) can be rearranged so that their nonzero elements lead those that are equal to zero. The matrix $\Pi$ can be conformally rearranged so that (25) can be rewritten as

$$(\boldsymbol{\psi}_\Delta^T \quad \mathbf{0}_{\Delta\Delta}^T)\begin{bmatrix} \Pi_{\Delta *} & \Pi_{\Delta **} \\ \Pi_{\Delta\Delta *} & \Pi_{\Delta\Delta **} \end{bmatrix} = -(\boldsymbol{\gamma}_*^T \quad \mathbf{0}_{**}^T) \tag{26}$$

in which $\boldsymbol{\psi}_\Delta$ and $\mathbf{0}_{\Delta\Delta}$ are the $G^\Delta \times 1$ and $(G - G^\Delta) \times 1$ subvectors of nonzero and zero elements in the left-hand side vector of (26); and $\boldsymbol{\gamma}_*$ and $\mathbf{0}_{**}$ are the $K^* \times 1$ and $(K - K^*) \times 1$ subvectors of nonzero and zero elements of the right-hand side vector of (26). The submatrices $\Pi_{\Delta *}$, $\Pi_{\Delta **}$, $\Pi_{\Delta\Delta *}$, and $\Pi_{\Delta\Delta **}$ are of dimensions $G^\Delta \times K^*$, $G^\Delta \times (K - K^*)$, $(G - G^\Delta) \times K^*$, and $(G - G^\Delta) \times (K - K)^*$, respectively, and correspond to a conformally rearranged matrix $\Pi$ as required by the vector partition in (26). $G^\Delta$ and $K^*$ denote the number of nonzero elements in the left- and right-hand side vectors of (25), respectively. Equation (26) leads to the following expressions:

$$\underset{(1 \times G^\Delta)}{\psi_\Delta{}^T} \; \underset{(G^\Delta \times K^*)}{\Pi_{\Delta *}} = \underset{(1 \times K^*)}{-\gamma_*{}^T} \qquad (27)$$

$$\underset{(1 \times G^\Delta)}{\psi_\Delta{}^T} \; \underset{G^\Delta \times (K - K^*)}{\Pi_{\Delta **}} = \underset{1 \times (K - K^*)}{\mathbf{0}_{**}{}^T} \qquad (28)$$

In (27) and (28) it is possible to divide both sides by any of the (nonzero) elements of $\psi_\Delta$ so that one of the elements of $\psi_\Delta$ can be normalized to unity. Then, there are $(G^\Delta - 1) + K^*$ unknown variables in the $j$th equation (i.e., equation (25)). If (28) could be solved for $\psi_\Delta$, then $\gamma^*$ would be immediately determined from (27). From basic matrix theory [*Graybill*, 1983, pp. 149–178] it is known that at least $G^\Delta - 1$ equations are needed to solve for $\psi_\Delta$ in (28). The vector $\psi_\Delta$ has $G^\Delta - 1$ unknown elements, since one of its elements can be normalized to unity, as is stated above. Therefore it is required that

$$K - K^* \geq G^\Delta - 1 \qquad (29)$$

since there are $K - K^*$ columns in $\Pi_{\Delta **}$.

Equation (29) is only a necessary condition for identifiability, because even if it is satisfied, the columns of $\Pi_{\Delta **}$ may not be linearly independent. A necessary and sufficient condition for the identification of $\psi_\Delta$ and $\gamma_*$ in the $j$th equation is that the number of linearly independent columns of $\Pi_{\Delta **}$ be equal to $G^\Delta - 1$, i.e.,

$$\text{rank}\,(\Pi_{\Delta **}) = G^\Delta - 1 \qquad (30)$$

To summarize, the identification status of the $j$th equation (see equation (25)), $j = 1, 2, \cdots, G$ must belong to one of the possible cases below.

1. If $K - K^* > G^\Delta - 1$ and $\text{rank}(\Pi_{\Delta **}) = G^\Delta - 1$, the $j$th equation is overidentified, meaning that one can solve for the unknown elements $\psi_{jk}$ and $\gamma_{jl}$ by different, consistent methods. The overidentification condition implies that there are more independent equations than there are unknown parameters. The two-stage and three-stage least squares methods of sections 4 and 5 are applicable to overidentified equations.

2. If $K - K^* = G^\Delta - 1$ and rank $(\Pi_{\Delta **}) = G^\Delta - 1$, there exists exact identification. It is possible to solve uniquely for $\psi_\Delta$ and $\gamma_*$ from (27) and (28). Under exact identification, the two-stage and three-stage least square methods, as well as the maximum likelihood estimators yield identical parameter estimators.

3. If $K - K^* \geq G^\Delta - 1$ and rank $(\Pi_{\Delta **}) < G^\Delta - 1$, or if $K - K^* < G^\Delta - 1$, the equation is underidentified. In this case, it is not possible to estimate consistently the parameters in $\psi_\Delta$ and $\gamma_*$. This is equivalent to saying there are more unknowns than there are (independent) equations to estimate them. It is shown in subsection 3.2 that the problem of estimating $\Psi$ and $\Gamma$ in confined aquifer problems is most likely to be overidentified, so that there are alternative methods to obtain consistent estimators.

### 3.2. *An Example of Determining Identifiability*

*Judge et al.* [1982, p. 358] show that the rank condition expressed in (30) can be checked without having to compute $\Pi_{\Delta **}$ if (30) is written in its equivalent form

$$\begin{aligned}\text{rank}(\Pi_{\Delta **}) &= \text{rank}(\Psi_{\Delta\Delta} : \Gamma_{**}) - (G - G^\Delta)\\ &= G^\Delta - 1 \end{aligned} \qquad (31)$$

in which $\Psi_{\Delta\Delta}$ and $\Gamma_{**}$ are submatrices of the matrices $\Psi$ and $\Gamma$, respectively, corresponding to the variables omitted from the $j$th equation but included in the other rows in (24). $\Psi_{\Delta\Delta}$ and $\Gamma_{**}$ are of dimensions $(G - 1) \times (G - G^\Delta)$ and $(G - 1) \times (K - K^*)$, respectively (see example below).

The application of (29) and (31) is illustrated with the system introduced in subsection 2.2 (see equation (21)). By arbitrarily choosing the first equation, one obtains for the left- and right-hand side vectors of (26),

$$(\psi_\Delta{}^T \quad \mathbf{0}_{\Delta\Delta}{}^T) = (\psi_{11} \quad \psi_{12} \quad 0)$$

$$-(\gamma_*{}^T \quad \mathbf{0}_{**}{}^T) = -(\gamma_{11} \quad \gamma_{12} \quad \gamma_{15} \quad \gamma_{17} \quad 0\;0\;0\;0)$$

respectively, so that $K - K^* = 8 - 4 = 4$, and $G^\Delta - 1 = 1$; therefore $K - K^* > G^\Delta - 1$, and condition (29) is satisfied. The rank condition, as specified in (31), can be tested as follows:

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} & : & 0 \\ \psi_{21} & \psi_{22} & : & \psi_{23} \\ 0 & \psi_{32} & : & \psi_{33} \end{bmatrix}$$

and the reordered $\Gamma$ matrix is

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{15} & \gamma_{17} & : & 0 & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 & : & \gamma_{23} & 1 & 0 & 0 \\ 0 & \gamma_{32} & 0 & 0 & : & \gamma_{33} & 0 & \gamma_{36} & \gamma_{38} \end{bmatrix}$$

Thus

$$(\Psi_{\Delta\Delta} : \Gamma_{**}) = \begin{bmatrix} \psi_{23} & : & \gamma_{23} & 1 & 0 & 0 \\ \psi_{33} & : & \gamma_{33} & 0 & \gamma_{36} & \gamma_{38} \end{bmatrix}$$

and $\text{rank}(\Psi_{\Delta\Delta} : \Gamma_{**}) = 2$. Since $G - G^\Delta = 3 - 2 = 1$, then $\text{rank}(\Pi_{\Delta **}) = 2 - 1 = 1$; also, $G^\Delta - 1 = 2 - 1 = 1$; therefore $\text{rank}(\Pi_{\Delta **}) = \text{rank}(\Psi_{\Delta\Delta} : \Gamma_{**}) - (G - G^\Delta) = 1$, which equals $G^\Delta - 1 = 1$. The first equation of the system represented in (21) is therefore overidentified.

The reader can verify that for the second equation in the equation system (21), overidentification holds with $K - K^* = 8 - 4 > G^\Delta - 1 = 3 - 1$ and rank $(\Pi_{\Delta **}) = 2 - (3 - 3) = G^\Delta - 1 = 3 - 1$. The third equation in the equation system (21) is overidentified with $K - K^* = 8 - 4 > G^\Delta - 1 = 2 - 1$, and $\text{rank}(\Pi_{\Delta **}) = 2 - (3 - 2) = 2 - 1 = G^\Delta - 1 = 2 - 1$.

It is apparent that for groundwater problems of larger size either in one or two dimensions, the sparsities of the matrices $\Psi$ and $\Gamma$ will be larger than those in (21). The densities of those matrices (i.e., the proportion of nonzero elements) will become smaller as the dimension $G$ increases. The effect of this is an overabundance of information (taking the form of equations) about relatively few parameters on each of the $G$ equations of the system given by (24). Consequently, the parameters (i.e., the nonzero elements) of the matrices $\Psi$ and $\Gamma$ cannot be uniquely estimated, as is shown above. There is a condition of overidentification. The estimation methods of sections 4 and 5 yield consistent estimators for the parameters $\psi_{jk}$ and $\gamma_{jl}$.

## 4. ESTIMATION BY ONE SINGLE-EQUATION METHOD

### 4.1. *The Two-Stage Least Squares Method (2SLS)*

The 2SLS method was originally developed by *Basmann* [1957] in the context of econometric models First, notice that

(13) can be rewritten for all time periods, at once, as follows:

$$\Psi[\boldsymbol{\phi}_1 \cdots \boldsymbol{\phi}_n] + \Gamma[\mathbf{x}_1 \cdots \mathbf{x}_n] = [\mathbf{u}_1 \cdots \mathbf{u}_n] \tag{32}$$

The system of (32) contains $G$ equations, each equation corresponding to one of the rows of the matrix $\Psi$ (say, the $j$th) times the matrix of the $\boldsymbol{\phi}$'s, plus the $j$th row of $\Gamma$ times the matrix of the $\mathbf{x}$'s being equal to the $j$th row of the right-hand side of (32). The next step is to choose the $j$th row equation and normalize the $\psi_{jj}$ parameter to unity by dividing the entire $j$th equation by $\psi_{jj}$, an arbitrary choice (in the sequel, the normalized coefficients, $\psi_{ij}^* = \psi_{ij}/\psi_{jj}$ and $\gamma_{jl}^* = \gamma_{jl}/\psi_{jj}$ are represented by $\psi_{ij}$ and $\gamma_{jl}$, respectively, to simplify the notation, and from the context it should be obvious whether the raw or normalized coefficients are being used in the equations). By taking the transpose of the (normalized) $j$th equation (so that the parameters are ordered columnwise) one obtains

$$\boldsymbol{\phi}_j = \Phi_j \boldsymbol{\psi}_j + X_j \boldsymbol{\gamma}_j + \mathbf{u}_j \tag{33}$$

in which

$$\boldsymbol{\phi}_j^T = [\phi_j(1), \cdots, \phi_j(n)]_{1 \times n}$$

$$\boldsymbol{\psi}_j^T = [-\psi_{j1}, \cdots, -\psi_{j,j-1}, -\psi_{j,j+1}, \cdots, -\psi_{jG^\Delta}]_{1 \times (G^\Delta - 1)}$$

$$\boldsymbol{\gamma}_j^T = [-\gamma_{j1}, -\gamma_{j2}, \cdots, -\gamma_{jK*}]_{1 \times K*}$$

$$\mathbf{u}_j^T = [u_j(1), \cdots, u_j(n)]_{1 \times n}$$

$$\Phi_j = \begin{bmatrix} \phi_1(1) & \cdots & \phi_{j-1}(1) & \phi_{j+1}(1) & \cdots & \phi_{G^\Delta}(1) \\ \phi_1(1) & \cdots & \phi_{j-1}(2) & \phi_{j+1}(2) & \cdots & \phi_{G^\Delta}(2) \\ \vdots & & \vdots & \vdots & & \vdots \\ \phi_1(n) & \cdots & \phi_{j-1}(n) & \phi_{j+1}(n) & \cdots & \phi_{G^\Delta}(n) \end{bmatrix}_{n \times (G^\Delta - 1)}$$

$$X_j = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_{K*}(1) \\ x_1(2) & x_2(2) & \cdots & x_{K*}(2) \\ \vdots & \vdots & & \vdots \\ x_1(n) & x_2(n) & \cdots & x_{K*}(n) \end{bmatrix}_{n \times K*}$$

In (33), $E(\mathbf{u}_j\mathbf{u}_j^T) = \sigma_{jj}I_{nn}$, according to the assumptions given in (11) and (12). Also, notice that matrices $\phi_j$ and $X_j$ contain the variables associated with nonzero coefficients; thus their respective column dimensions are $G^\Delta - 1$ and $K^*$ (see discussion following equation (26)).

Equation (33) can be written in the usual linear model form,

$$\boldsymbol{\phi}_j = Z_j \boldsymbol{\beta}_j + \mathbf{u}_j \tag{34}$$

where $Z_j = [\Phi_j X_j]$ and $\boldsymbol{\beta}_j^T = [\boldsymbol{\psi}_j^T \boldsymbol{\gamma}_j^T]$.

It is tempting to apply the ordinary least squares (OLS) method to (34) to obtain

$$\hat{\boldsymbol{\beta}}_j = (Z_j^T Z_j)^{-1} Z_j^T \boldsymbol{\phi}_j \tag{35}$$

in which the design matrix $Z_j$ as well as the heads $\boldsymbol{\phi}_j$ are observable data. Furthermore, one could estimate $\boldsymbol{\beta}_j$ similarly for $j = 1, 2, \cdots, G$, i.e., for each equation, one at a time (hence the name single-equation method), to estimate the nonzero elements of $\Psi$ and $\Gamma$. One inconvenience though is that $Z_j$ is a stochastic matrix (it contains the matrix $\phi_j$) whose first $G^\Delta - 1$ columns are correlated with the error vector $\mathbf{u}_j$. This is easily shown by taking one of the columns of $\Phi_j$, say, $\boldsymbol{\phi}_k$, and taking expectations,

$$E(\boldsymbol{\phi}_k \mathbf{u}_j^T) = E[(Z_k \boldsymbol{\beta}_k + \mathbf{u}_k)(\mathbf{u}_j^T)] = E(Z_k \boldsymbol{\beta}_k \mathbf{u}_j^T) + \sigma_{kj} I_{nn}$$

which is clearly nonzero even if one assumes that $\sigma_{kj} = 0$ due to the complex nonzero term $E(Z_k \boldsymbol{\beta}_k \mathbf{u}_j^T)$. It is well known that a correlated design matrix and error term lead to inconsistent OLS estimators (see, for example, *Kmenta* [1971, pp. 298–304]).

The two-stage least squares method transforms (33) so that the resulting design matrix becomes asymptotically uncorrelated with the (transformed) error term. The details are given in Appendix A, where it is shown that (33) can be expressed as

$$\boldsymbol{\phi}_j = \hat{Z}_j \boldsymbol{\beta}_j + \mathbf{w}_j \tag{36}$$

in which the transformed design matrix $\hat{Z}_j$ and error term $\mathbf{w}_j$ (defined in Appendix A) are asymptotically uncorrelated. Therefore the OLS method applied to (36) has all the asymptotic properties of the standard linear model, in particular, it is consistent. The 2SLS estimator for $\boldsymbol{\beta}_j$ is thus

$$\tilde{\boldsymbol{\beta}}_j = (\hat{Z}_j^T \hat{Z}_j)^{-1} \hat{Z}_j^T \boldsymbol{\phi}_j \tag{37}$$

The first stage of the 2SLS method consists of finding the matrix $\hat{Z}_j$ (see Appendix A) and the second stage in computing $\tilde{\boldsymbol{\beta}}_j$ by (37).

### 4.2. *Properties of the 2SLS Method*

The asymptotic covariance of $\tilde{\boldsymbol{\beta}}_j$ is estimated by

$$\begin{aligned} \operatorname{cov}(\tilde{\boldsymbol{\beta}}_j) &= (\hat{Z}_j^T \hat{Z}_j)^{-1} \hat{Z}_j^T (\hat{\sigma}_{jj} I_{nn}) \hat{Z}_j (\hat{Z}_j^T \hat{Z}_j)^{-1} \\ &= \hat{\sigma}_{jj} (\hat{Z}_j^T \hat{Z}_j)^{-1} \end{aligned} \tag{38}$$

The unknown $\sigma_{jj}$ is estimated by the well-known consistent estimator

$$\hat{\sigma}_{ij} = \frac{(\boldsymbol{\phi}_i - \hat{Z}_i \tilde{\boldsymbol{\beta}}_i)^T (\boldsymbol{\phi}_j - \hat{Z}_j \tilde{\boldsymbol{\beta}}_j)}{n - (G^\Delta - 1 + K^*)} \tag{39}$$

with $i = j$, which is the usual error sum of squares, divided by the corresponding degrees of freedom, in classical linear model theory.

From large-sample theory of least squares (see, for example, *Rao* [1973]) it follows that the asymptotic distribution of $(n)^{1/2}(\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)$ is normal, zero mean, with covariance equal to $\sigma_{jj}\, \operatorname{plim}_{n \to \infty} [n^{-1}(\hat{Z}_j^T \hat{Z}_j)]^{-1}$, so that $\tilde{\boldsymbol{\beta}}_j$ is consistent.

The 2SLS method can be applied to each equation ($j = 1, 2, \cdots, G$) to yield consistent estimators for all the nonzero elements of $\Psi$ and $\Gamma$ and the error covariance matrix $[\sigma_{ij}] = \Sigma$. An application of the 2SLS method is provided in section 7.

## 5. Estimation by One System Equation Method

### 5.1. *The Three-Stage Least Squares Method (3SLS)*

The 2SLS method is applied to each equation, one by one. It is possible to estimate all the $\boldsymbol{\beta}_j$'s ($j = 1, 2, \cdots, G$) simultaneously (hence the name system equation), with a gain in asymptotic efficiency (see, for example, *Bickel and Doksum*, [1977, pp. 137–141]). Such efficiency gain is due to the fact that in the 2SLS method the cross correlation among equation disturbances is not considered, since it deals with one equation at a time.

Adopting the same notation of subsection 4.1 (see equation (36)), the system of $G$ equations can be expressed in augmented form by stacking them as follows:

$$\begin{bmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \vdots \\ \boldsymbol{\phi}_G \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 & 0 & \cdots & 0 \\ 0 & \hat{Z}_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \hat{Z}_G \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_G \end{bmatrix} + \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_G \end{bmatrix} \tag{40}$$

or in compact notation,

$$\boldsymbol{\phi} = \hat{Z}\boldsymbol{\beta} + \mathbf{w} \tag{41}$$

in which $\boldsymbol{\phi}$ and $\mathbf{w}$ are both of dimension $nG \times 1$, and $\boldsymbol{\beta}$ is of dimension $[\sum_{j=1}^{G} (G_j^{\Delta} - 1 + K_j^*)] \times 1$, in which $G_j^{\Delta}$ and $K_j^*$ are the number of nonzero parameters in the $j$th row of the $\Psi$ and $\Gamma$ matrices, respectively (see equations (25) and (26)). The dimension of the matrix $\hat{Z}$ in (41) is $nG \times [\sum_{j=1}^{G} (G_j^{\Delta} - 1 + K_j^*)]$.

*Zellner and Theil* [1962] proposed to apply the generalized least squares technique to (41) to yield the 3SLS estimator,

$$\tilde{\boldsymbol{\beta}} = [\hat{Z}^T(\hat{\Sigma}\Theta I_{nn})^{-1}\hat{Z}]^{-1}[\hat{Z}^T(\hat{\Sigma}\Theta I_{nn})^{-1}]\boldsymbol{\phi} \tag{42}$$

in which $\hat{\Sigma}\Theta I_{nn}(=[\hat{\sigma}_{ij}I_{nn}])$ is the Kronecker product of $\hat{\Sigma}$ and the identity matrix $I_{nn}$, and $\hat{\Sigma}$ denotes the matrix $[\hat{\sigma}_{ij}]$ (of dimension $G \times G$), where the covariance estimators $\hat{\sigma}_{ij}$'s are obtained from the error sum of squares computed from using the 2SLS estimators, i.e.,

$$\hat{\sigma}_{ij} = \frac{(\boldsymbol{\phi}_i - \hat{Z}_i\tilde{\boldsymbol{\beta}}_i)^T(\boldsymbol{\phi}_j - \hat{Z}_j\tilde{\boldsymbol{\beta}}_j)}{n - \max\,[G_i^{\Delta} - 1 + K_i^*,\, G_j^{\Delta} - 1 + K_j^*]} \tag{43}$$

The covariance $\hat{\Sigma}\Theta I_{nn}$ (of dimension $nG \times nG$) is an estimator of the asymptotic covariance of $\mathbf{w}$ in (41), and its structure (i.e., the Kronecker product) is a consequence of the covariance assumptions made in (11) and (12). In the implementation of the 3SLS estimator it is assumed that the number of observations $n$ exceeds the number of equations $G$ to avoid the singularity of $\hat{\Sigma}\Theta I_{nn}$.

The calculation of $\tilde{\boldsymbol{\beta}}$ in (42) requires first the computation of the 2SLS estimators $\tilde{\boldsymbol{\beta}}_j$ $(j = 1, 2, \cdots, G)$ to estimate $\hat{\Sigma}\Theta I_{nn}$, and subsequently the generalized least squares (42); hence its name, 3SLS estimator.

### 5.2. *Properties of the 3SLS Method*

The linearity of the estimator $\tilde{\boldsymbol{\beta}}$ in (42) leads to the following estimator of its asymptotic covariance:

$$\text{Cov}\,(\tilde{\boldsymbol{\beta}}) = [\hat{Z}^T(\hat{\Sigma}^{-1}\Theta I_{nn})\hat{Z}]^{-1} \tag{44}$$

(notice that $(\hat{\Sigma}\Theta I_{nn})^{-1} = \hat{\Sigma}^{-1}\Theta I_{nn}$). The asymptotic distribution of $(n)^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is normal, zero mean, with covariance equal to

$$\operatorname*{plim}_{n\to\infty}\, [n^{-1}\hat{Z}^T(\hat{\Sigma}^{-1}\Theta I_{nn})\hat{Z}]^{-1}$$

implying that $\tilde{\boldsymbol{\beta}}$ is a consistent estimator. The 3SLS method is illustrated in section 7.

## 6. MAXIMUM LIKELIHOOD ESTIMATION

### 6.1. *The Negative Log-Likelihood Function*

The estimation methods presented in subsections 5.1 and 5.2 are linear. They yield the matrices that govern the discretized equation of groundwater flow (see equations (9) and (13)). It was shown that due to overidentification (the number of linearly independent equations exceeds the number of unknown parameters) the parameters are not unique (i.e., there are alternative consistent estimators). The criterion for determining the identification status was based on analyzing the groundwater flow equation as a linear simultaneous equation system. It is possible to determine directly transmissivities, storativities, and the autoregressive parameter $\rho$ (see equation (10)) if one introduces a nonlinear criterion to be maximized with respect to the set of unknown parameters. One such criterion is the maximum likelihood (ML) method.

Assuming that $\mathbf{u}_t$ in (13) has a multivariate normal distribution, the likelihood function corresponding to the process of (13) is

$$L = \frac{|\Psi|^n}{(2\pi)^{nG/2}}\,|\Sigma|^{-n/2} \cdot \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)^T\Sigma^{-1}(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)\right\} \tag{45}$$

For estimation purposes it is convenient to use the negative of the logarithm of (45) to obtain the following negative log-likelihood function:

$$f = \frac{nG}{2}\ln(2\pi) + \frac{n}{2}\ln|\Sigma| - n\ln|\Psi| + \frac{1}{2}\sum_{t=1}^{n}(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)^T\Sigma^{-1}(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t) \tag{46}$$

The unknown matrices $\Psi$ and $\Gamma$ have elements that are functions of transmissivities and storativities (see equations (17) and (20)). Therefore one must minimize (46) with respect to a parameter vector $\boldsymbol{\theta}$ $(q \times 1)$ whose elements are the unknown transmissivities and storativities, which vary within the subdomains of the finite-element spatial discretization, as well as the autoregressive parameter $\rho$.

It is convenient to simplify (46) by taking its derivative with respect to $\Sigma$, equating to zero and solving for $\Sigma$, to obtain the estimator $\hat{\Sigma}$:

$$\hat{\Sigma} = \frac{1}{n}\left[\sum_{t=1}^{n}(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)^T\right] \tag{47}$$

By substituting (47) into (46) one obtains the following expression for the negative log-likelihood function $f$:

$$f = \frac{nG}{2}\ln(2\pi) + \frac{n}{2}\ln\left|\frac{1}{n}\sum_{t=1}^{n}(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)(\Psi\boldsymbol{\phi}_t + \Gamma\mathbf{x}_t)^T\right| - n\ln|\Psi| + \frac{n^2}{2} = C + \frac{n}{2}\ln|\hat{\Sigma}| - n\ln|\Psi| \tag{48}$$

in which

$$c = \frac{nG}{2}\ln(2\pi) + \frac{n^2}{2} = \text{const} \tag{49}$$

Equation (48) is minimized with respect to the parameter vector $\boldsymbol{\theta}$ by the Newton-Raphson method [see *Gill et al.*, 1981]. One advantage of using the Newton-Raphson method to minimize (48) is that the gradient and Hessian matrix of $f$ are computable in closed form (i.e., there is no need for numerical differentiation in the search algorithm) as shown in Appendix B. In addition, the rate of convergence of the Newton-Raphson is quadratic. An application of the ML approach is given in section 7.

### 6.2. *Properties of ML Estimators*

Suppose $\tilde{\boldsymbol{\theta}}$ denotes the true but unknown vector of parameters. Let

$$I(\boldsymbol{\theta}) = E(\partial^2 f/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T) \tag{50}$$

$$i(\boldsymbol{\theta}) = \partial^2 f/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T \tag{51}$$

in which the expectation in (50) is with respect to $\boldsymbol{\phi}_t$. The matrices given in (50) and (51) are the Fisher information and the sample information matrices, respectively. For a sample size $n$ sufficiently large, the distribution of the ML estimator

$\boldsymbol{\theta}^*$ is approximately [see, for example, *Efron and Hinkley*, 1978]

$$\boldsymbol{\theta}^* \sim N(\tilde{\boldsymbol{\theta}}, I^{-1}(\tilde{\boldsymbol{\theta}})) \tag{52}$$

in which $N$ denotes the multivariate normal distribution. From (50) follows that the expression

$$(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T I^{-1}(\boldsymbol{\theta}^*)(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \le \chi_\alpha^2(q) \tag{53}$$

in which $\chi_\alpha^2(q)$ is the $(1 - \alpha)$th percentile of a chi-squared variable with $q$ degrees of freedom, represents an ellipsoid in the $q$-dimensional $\boldsymbol{\theta}$-space centered at $\boldsymbol{\theta}^*$; the probability that this random ellipsoid covers the true parameter $\tilde{\boldsymbol{\theta}}$ is $1 - \alpha$. Equation (52) allows the construction of a hypothesis test. Let

$$\begin{aligned} H_0&: \boldsymbol{\theta} = \boldsymbol{\theta}^0 \\ H_1&: \boldsymbol{\theta} \neq \boldsymbol{\theta}^0 \end{aligned} \tag{54}$$

in which $\boldsymbol{\theta}^0$ is the specified value in the null hypothesis $H_0$. The null hypothesis $H_0$ is rejected at a significance level $\alpha$ if

$$(\boldsymbol{\theta}^0 - \boldsymbol{\theta}^*)^T I^{-1}(\boldsymbol{\theta}^0)(\boldsymbol{\theta}^0 - \boldsymbol{\theta}^*) > \chi_\alpha^2(q) \tag{55}$$

In practice, $I(\ )$ (see equation (50)) may be difficult to obtain, so the sample information matrix $i(\ )$ (see equation (51)) replaces $I(\ )$ in (52), (53), and (55).

### 6.3. *Caveats on Least Squares and ML Estimation*

Under overidentification conditions (see section 3.1), the 2SLS and 3SLS methods provide consistent estimators for the parameter matrices $\Psi$ and $\Gamma$ (and $\Sigma$). The implementation of these methods requires only the use of ordinary least squares and generalized least squares subroutines that are extremely efficient (and stable) in solving the system of equations on the unknown parameters. For identified (see section 3.1) inverse problems it can be shown that the nonzero elements of the matrices $\Psi$ and $\Gamma$ are identically estimated (i.e., obtain the same numerical values) by either the 2SLS, 3SLS, or the ML methods.

The ML approach given in this paper deals directly with the unknown parameters $T$, $S$, and $\rho$ by minimizing a nonlinear function (i.e., the negative log-likelihood function, equation (48)) with respect to $T$, $S$, and $\rho$. Our computational experience in this study (see section 7) shows that the ML method results in a highly stable estimation process. The search algorithm shows quadratic convergence, and this is partly due to the fact that the log-likelihood function (see equation 48) is convex on the parameter space $\boldsymbol{\theta}$. Such convexity follows from the properties of the family of exponential distributions, of which the normal is a member.

## 7. APPLICATION OF METHODS

The implementation of the estimation methods discussed in sections 4–6 is illustrated with the confined aquifer of Figure 1. This example, although simple in nature, is intended to illustrate the theoretical developments presented in this paper. A large-scale application to the San Joaquin groundwater basin is in preparation. Table 1 contains the data used to generate the piezometric heads used in the estimation.

### 7.1. *Generation of Head Values*

Based on the data in Table 1, the head values at nodes 1, 2, and 3 (Figure 1) were generated by the exact analytical solution given in (7). Piezometric heads were computed for $t = 1, 2, \cdots, 20$, where the time index is in days. The solution of (7) led to the following closed-form expression:

$$\phi_i(t) = A_i + B_i t + C_i e^{\lambda_1 t} + D_i e^{\lambda_2 t} + E_i e^{\lambda_3 t} \tag{56}$$

$i = 1, 2, 3$; $t > 0$. The constant coefficients $A_i$, $B_i$, $C_i$, $D_i$, and $E_i$, as well as the eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$ of the matrix $C$ (see equations (6) and (7)) are given in Table 2. The negative values of the $\lambda_i$'s indicate that if the boundary heads are held constant, even if pumping continues indefinitely at a constant rate, the heads at the interior nodes will reach a steady state after relatively short times (i.e., the system is stable). This follows from the fact that $e^{\lambda_i t}$ approaches zero exponentially with $t$, and that $B_i$ becomes zero when the boundary conditions, as well as the sinks/sources, are kept constant.

### 7.2. *Ordinary Least Squares Estimators*

For the purpose of comparison, the OLS estimators (see equation (35)) were computed. It was shown in subsection 4.1 that the OLS estimators are inconsistent estimators. However, they are readily computable, and even though their asymptotic properties are unappealing, they can provide a quick approximation to the parameters $\boldsymbol{\beta}_j$ ($j = 1, 2, 3$) in (33). Table 3 displays the values of the estimators, as well as their corresponding standard errors (within parentheses). From the values of the parameter vectors $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ the matrices $\Psi$ and $\Gamma$ (see equation (21)) can be constructed and used to simulate the piezometric heads (the same approach is used to simulate the heads after computing the 2SLS or the 3SLS estimators).

### 7.3. *Two-Stage and Three-Stage Least Squares Estimators (2SLS, 3SLS)*

The 2SLS and 3SLS estimators (see equations (37) and (42), respectively) and their standard errors are shown in Table 3, together with the OLS estimators and the true parameter values. It can be appreciated from Table 3 that the 3SLS estimators provide a very good approximation to the parameters except for the first and second components of $\boldsymbol{\beta}_3$. Overall, the OLS and 2SLS estimators provide a good estimate of the true parameters, although they show larger deviations from the actual values than those shown by 3SLS, except for two of the components of $\boldsymbol{\beta}_3$ (i.e., the first and second components). In particular, the OLS and 2SLS estimators are quite similar, although the latter appear to give an overall better resemblance of the true parameter values than the former.

The standard errors show that in general, the 2SLS exhibit a smaller standard error than the OLS and 3SLS estimators, the former showing the larger standard errors of all three alternative estimators. The presumed (asymptotic) efficiency gained by the 3SLS due to the joint estimation of all equations is not reflected in this limited-size sample experiment. Gain in efficiency is likely to be effective in reducing small-sample standard errors when the covariances $\sigma_{ij}$ are known, which is not the case in this study. Head values generated by means of the OLS, 2SLS, and 3SLS (i.e., $Z_j\hat{\boldsymbol{\beta}}_j$, $\hat{Z}_j\hat{\boldsymbol{\beta}}$, and $\hat{Z}\hat{\boldsymbol{\beta}}$, respectively) reproduced very closely the actual values obtained from the analytical continuous time solution of (56). Table 4 shows the actual (i.e., continuous-time) and simulated (using equation (36)) piezometric heads for periods $t = 1, 2, \cdots, 20$.

The computation of the OLS and 2SLS requires the inversion of matrices $Z_j^T Z_j$ and $\hat{Z}_j^T \hat{Z}_j$, respectively (see equations (35) and (37)). Often, the columns of the matrices $Z_j$ and $\hat{Z}_j$

TABLE 1. Data for the Example Aquifer

| Element $i$ | Transmissivity $T$, m$^2$/day | Storativity $S$ | Length $l$, m |
|---|---|---|---|
| 1 | 500 | $12 \times 10^{-3}$ | 500 |
| 2 | 500 | $12 \times 10^{-3}$ | 500 |
| 3 | 500 | $12 \times 10^{-3}$ | 500 |
| 4 | 500 | $12 \times 10^{-3}$ | 500 |

| Matrix $A$ (Equation (5)) | Vector F (Equation (5)) | Matrix $B$ (Equation (5)) |
|---|---|---|
| $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ m/day | $\begin{bmatrix} 0 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} F \\ \phi_A \\ \phi_B \\ \dot{\phi}_A \\ \dot{\phi}_B \end{bmatrix}$ | $B = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$ m |

Here, $\phi_A(t) = 80 + t$, m; $\dot{\phi}_A = 1$, m/day; $\phi_B(t) = 100 - t$, m; $\dot{\phi}_B = -1$, m/day; $F = 10$, m$^3$ m$^{-1}$ day$^{-1}$; $\Delta t = 1$ day; and $\omega = 0.5$.

show multicollinearity; i.e., some of their columns are almost linearly dependent, so that $Z_j$ or $\hat{Z}_j$ do not have full column rank and the inverses of $Z_j^TZ_j$ or $\hat{Z}_j^T\hat{Z}_j$ are not computable numerically. This was the case in our application. The problem can be readily solved by perturbing, say, $Z_j^TZ_j$, by adding to it a diagonal matrix $kI$ ($I$ is the identity matrix). Thus one utilizes $(Z_j^TZ_j + kI)^{-1}$. From the theory of ridge regression [see, for example, *Hoerl and Kennard* [1970] or *Marquardt and Snee* [1975]], a small $k$ number (in our case $k$ was set at the value $k = 1$) stabilizes the inversion while leaving the estimators basically unaffected by the perturbation. Multicolinearity is to be expected in groundwater problems in which the heads at adjacent nodes are columns of the regression matrix $Z_j$ (or $\hat{Z}_j$), for such columnar values are likely to be quite similar.

### 7.4. *Maximum Likelihood (ML) Estimators*

The ML approach presented in section 6 was used to estimate $T$, $S$, and $\rho$ for the aquifer of Figure 1. Several initial estimators $T^{(0)}$ and $S^{(0)}$ were tried to start the search for optimal estimators. In all cases, a quadratic rate of convergence was observed to the same minima. Table 5 shows one of the convergence paths to the optimal estimators $T^* = 462$ m$^2$/day, $S^* = 0.0110$, and $\rho^* = 0.08$. The actual parameter values (see Table 1) are $T = 500$ m$^2$/day, $S = 0.012$, and $\rho = 0.10$. The standard errors of estimators (within parentheses in Table 5) were estimated by the square root of the diagonal terms of the sample information matrix (see equation (51)). The maximum likelihood estimators $T^*$, $S^*$, and $\rho^*$ were used to estimate $\Psi$ and $\Gamma$ through the relationships given in (17)–(20). Having the ML estimates $\Psi^*$ and $\Gamma^*$ for $\Psi$ and $\Gamma$, respectively, piezometric heads were simulated by

$$\hat{\boldsymbol{\phi}}_t = \Pi^* \mathbf{x}_t \tag{57}$$

in which

$$\Pi^* = -(\Psi^*)^{-1}\Gamma^* \tag{58}$$

(see equation (22). The simulated head values were essentially the same as those obtained from the 2SLS and 3SLS estimates, already shown in Table 4.

## 8. SUMMARY AND CONCLUSIONS

The theory and methods presented in this paper apply to linear, distributed-parameter, and unsteady systems, of which the flow in a confined aquifer is a typical example. It has been shown that it is possible to derive a continuous-time, closed-form solution for the linear system of differential equations corresponding to the finite-element (spatial) discretization of the confined-aquifer flow equation. The computation of the transition matrix $e^{tC}$ requires moderate numerical and analytical effort and is conveniently obtained by the method of interpolating polynomials. On the other hand, once the closed-form solution of the continuous groundwater flow equation is obtained, one can easily establish the convergence rate to a steady state, whether such steady state exists (stability), and the numerical value of the steady state head distribution. Piezometric heads can be generated for any time and for any node straightforwardly.

Upon the time discretization and weighting of consecutive (i.e., at times $t$ and $t - 1$) of the continuous system equation, one obtains a discrete linear system with unknown stationary parameters. The first important question is whether the matrices governing the discretized flow equations can be estimated consistently (in a statistical sense). The answer depends on their identifiability status, i.e., on whether each of the rows of the system equation is itself under-, exactly, or overidentified. Underidentification leads to inconsistent estimation, exact identification implies unique and consistent estimation, and overidentification is synonymous to consistent but nonunique estimation.

By means of an example it was argued that the inverse problem in groundwater is almost surely overidentified. As a consequence, three linear estimation techniques (OLS, 2SLS, and 3SLS) were utilized to obtain the elements of the matrices $\Psi$ and $\Gamma$ governing the discretized flow equation. As a subproduct of the linear estimation techniques, their variance-covariance matrices and asymptotic distributions are also available.

The simplest linear method, i.e., OLS, is computed with

TABLE 2. Parameters of the Analytic Solution

| Node $i$ | $A_i$ | $B_i$ | $C_i$ | $D_i$ | $E_i$ | $\lambda_i$ |
|---|---|---|---|---|---|---|
| 1 | 78.5036 | 0.5000 | −1.0356 | 1.5000 | 6.0319 | −1.3204 |
| 2 | 80.0056 | 0.0000 | 1.4639 | 0.0000 | 8.5306 | −0.5000 |
| 3 | 91.5037 | −0.5000 | −1.0356 | −1.5000 | 6 0319 | −0 1082 |

TABLE 3. Values of Parameters and Standard Errors

| Method | $\beta_1$ 1 | 2 | 3 | 4 | 5 | $\beta_2$ 1 | 2 | 3 | 4 | 5 | 6 | $\beta_3$ 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.10880 | −0.38920 | −0.25932 | −0.27312 | 0.08673 | −0.05797 | −0.03821 | −0.17117 | −0.55131 | −0.18538 | 0.10239 | 0.15485 | −018028 | −0.51807 | −0.28130 | 0.16312 |
| | (0.02532) | (0.03138) | (0.02981) | (0.009783) | (0.04931) | (0.03420) | (0.029100) | (0.032510) | (0.01831) | (0.01516) | (0.01712) | (0.09312) | (0.07311) | (0.11769) | (0.053211) | (0.08416) |
| 2SLS | 0.09853 | −0.42612 | −0.24951 | −0.29003 | 0.144812 | −0.06931 | 0.15077 | 0.15216 | −0.51644 | −0.18361 | 0.10101 | −0.05192 | −015131 | −0.46154 | −0.29832 | 0.06397 |
| | (0.033320) | (0.03810) | (0.016513) | (0.011341) | (0.033617) | (0.33178) | (0.03647) | (0.02341) | (0.01706) | (0.03164) | (0.014507) | (0.15632) | (0.16313) | (0.12137) | (0.06749) | (0.22998) |
| 3SLS | 0.06347 | −0.61542 | −0.24514 | −0.19167 | 0.06413 | 0.11368 | 0.09134 | −0.28736 | −0.62139 | −0.27672 | 0.10932 | 0.03412 | 038935 | −0.59633 | −0.23131 | 0.10976 |
| | (0.05631) | (0.07632) | (0.06632) | (0.03126) | (0.019323) | (0.10431) | (0.03657) | (0.10827) | (0.02132) | (0.02132) | (0.00156) | (0.21438) | (0.27131) | (0.18360) | (0.09130) | (0.10382) |
| True value | 0.1 | −0.6 | −0.3 | −0.2 | 0.2 | 0.1 | 0.1 | −0.3 | −0.6 | −0.3 | 0.2 | 0.1 | −0.3 | −0.6 | −0.2 | 0.2 |

Standard errors are given in parentheses.

TABLE 4. Actual and Forecasted Piezometric Heads

| Time, $t$ | Actual Node 1 | Actual Node 2 | Actual Node 3 | Simulated Node 1 | Simulated Node 2 | Simulated Node 3 |
|---|---|---|---|---|---|---|
| 0 | 85.00 | 90.00 | 95.00 | | | |
| 1 | 85.05 | 88.05 | 95.23 | 85.05 | 87.99 | 95.29 |
| 2 | 84.04 | 86.98 | 94.74 | 84.67 | 86.95 | 94.78 |
| 3 | 84.68 | 86.20 | 94.01 | 84.62 | 86.22 | 94.07 |
| 4 | 84.61 | 85.55 | 93.21 | 84.61 | 85.52 | 93.29 |
| 5 | 84.64 | 84.97 | 93.39 | 84.64 | 84.95 | 92.47 |
| 6 | 84.73 | 84.46 | 91.58 | 84.75 | 84.46 | 91.66 |
| 7 | 84.88 | 84.01 | 90.79 | 84.90 | 84.02 | 90.82 |
| 8 | 85.07 | 83.06 | 90.01 | 85.11 | 83.13 | 90.10 |
| 9 | 85.29 | 83.23 | 89.27 | 85.32 | 83.26 | 89.32 |
| 10 | 85.56 | 82.90 | 88.54 | 85.58 | 82.92 | 88.59 |
| 11 | 85.84 | 82.60 | 87.83 | 85.88 | 82.63 | 87.88 |
| 12 | 86.15 | 82.33 | 87.15 | 86.16 | 82.35 | 87.20 |
| 13 | 86.48 | 82.10 | 86.48 | 86.50 | 82.11 | 86.52 |
| 14 | 86.83 | 81.88 | 85.83 | 86.83 | 81.89 | 85.89 |
| 15 | 87.19 | 81.69 | 85.19 | 87.19 | 81.70 | 85.21 |
| 16 | 87.57 | 81.52 | 84.57 | 87.55 | 81.52 | 84.57 |
| 17 | 87.69 | 81.36 | 83.96 | 87.57 | 81.36 | 83.94 |
| 18 | 88.36 | 81.22 | 83.36 | 88.35 | 81.19 | 83.35 |
| 19 | 88.78 | 81.10 | 82.78 | 88.74 | 81.06 | 82.74 |
| 20 | 89.20 | 80.99 | 82.20 | 89.18 | 80.95 | 82.16 |

Piezometric heads are in meters.

minimal effort, but due to the correlation of its regression matrix with the error term, its large sample properties are not desirable; in particular, it is not consistent. The 2SLS and 3SLS are consistent estimators that can be computed with standard least squares and generalized least squares regressions, respectively. The application of this study showed that the 3SLS offered a better overall approximation to the true parameters, although despite its unique joint estimation feature, it showed larger standard error of estimators than the OLS and 2SLS methods. It turns out that the supposed gain in asymptotic efficiency of the 3SLS method (due to joint estimation) materializes when the unknown covariances $\sigma_{ij}$ are known. Since in actuality the $\sigma_{ij}$'s are unknown, the gain in asymptotic efficiency may not materialize in limited-size sample estimates. Clearly, our limited-size experiment points to the fact that the standard errors of 3SLS estimators are not necessarily smaller than those of OLS and 2SLS estimators.

A maximum likelihood approach has been developed in this paper. The ML method yields directly estimates of transmissivities, storativities, and other statistical parameters. The asymptotic distribution, confidence ellipsoids, and test of hypothesis for ML estimators have also been derived. Due to the (global) convexity of the negative log-likelihood function, the

TABLE 5. Results of Maximum Likelihood Estimation

| Iteration | Transmissivity, $m^2$/day | Storativity | Autoregressive Parameter $\rho$ | Negative Log-Likelihood Function |
|---|---|---|---|---|
| 0 | 350 | 0.0060 | 0.10 | 11.98 |
| 1 | 455 | 0.0081 | 0.03 | 7.96 |
| 2 | 446 | 0.0096 | 0.06 | 7.10 |
| 3 | 453 | 0.0010 | 0.07 | 6.45 |
| 4 | 459 | 0.0015 | 0.08 | 6.30 |
| 5 | 462 | 0.0110 | 0.08 | 5.83 |
| | (47.8) | (0.00390) | (0.001) | |

Standard errors of ML estimators are given in parentheses. Actual values are $T = 500$ $m^2$/day, $S = 0012$, and $\rho = 0.10$.

ML estimators for $T$, $S$, and $\rho$ are global minima. A quadratic rate of convergence was observed in the search algorithm (i.e., Newton-Raphson method). A sensitivity analysis on the initial estimators showed convergence in all cases to the same minimizing point. Simulated heads were for all practical purposes equal when using the 2SLS, 3SLS, and ML estimates for $\Psi$ and $\Gamma$.

The theory and the application of this work leads to the following conclusions.

1. It is feasible and useful to obtain a closed-form (analytical) solution to the continuous-time flow equation for confined aquifers. Stability and convergence properties of the piezometric heads are readily established from the continuous-time solution.

2. The structure of the discretized flow equation lends itself for a general analysis of the identification status of the inverse problem. Nonuniqueness and consistent estimation are features of the inverse problem in groundwater when approached with linear estimation techniques.

3. Of the competing methods, OLS is inconsistent but yet may yield useful and easy to compute estimates capable of producing accurate simulated heads. 2SLS and 3SLS are appealing from the standpoint of their asymptotic properties, specifically, being consistent and asymptotically normal. For a limited sample test case, 3SLS showed the smaller bias but also the larger standard errors, indicating that for small samples, the (asymptotic) efficiency gain of joint estimation is of little relevance.

4. Since the 2SLS estimator requires an ordinary least squares stage for its implementation (and, in turn, 3SLS requires the 2SLS estimates), it is straightforward to design software to compute OLS, 2SLS, and 3SLS at once, using standard least squares and generalized least squares subroutines.

5. It has been shown that the ML technique is applicable to the discrete confined-aquifer flow equation, leading to direct estimates of transmissivities and storativities. Asymptotic properties and standard errors of estimates are also available in the ML method. Quadratic convergence, consistent estimators, and global optimality are attractive features of the ML approach developed in this work.

## 9. Future Research Needs and Possible Extensions

The methods presented in this paper are not limited to confined aquifer flow. It is possible to linearize the unconfined flow equation (see, for example, *Mariño and Luthin* [1982]) and apply the identifiability and estimation techniques developed above.

Nonlinear processes, such as unsaturated flow and solute transport, can also be dealt with conveniently by dividing the field into subregions or elements so that the spatial variability of parameters (such as diffussivity) can be adequately represented. By treating the parameters of diffusion (or diffusion-convection) processes as the unknowns and letting the field variable (e.g., tension heads or solute concentrations) be the known data, it is possible to write a discrete time system equation relating parameters and the field variable as in (13). Application of the methods of this paper is then feasible (see, for example, *Loaiciga and Mariño* [1986]).

A more complicated situation arises when the field variable is observed with errors. It is apparent that (9) can be expressed as

$$\boldsymbol{\phi}_t = P\boldsymbol{\phi}_{t-1} + \mathbf{q}_t + \boldsymbol{\xi}_t \tag{59}$$

in which the transition matrix $P$, the inputs $\mathbf{q}_t$, and the error term $\boldsymbol{\xi}_t$ follow immediately from the terms in (9). To the first-order, discrete, "state" equation (59) an observation process is added; i.e.,

$$\mathbf{z}_t = M\boldsymbol{\phi}_t + \mathbf{v}_t \tag{60}$$

which denotes that a linear combination (i.e., $M\boldsymbol{\phi}_t$) of heads (or any other field variable) is observed with error (distributed according to the white noise sequence $\mathbf{v}_t$). Given the measurements $\mathbf{z}_t$ and $\forall t$ it is possible to obtain the smoothed estimators of $\boldsymbol{\phi}_t$, $\forall t$, and the parameters of the state space model (59) and (60), namely, $P$, the unknown covariances $\Sigma_1$ and $\Sigma_2$ of the error terms $\boldsymbol{\xi}_t$, and $\mathbf{v}_t$, respectively, as well as the initial conditions, i.e., the expected value and covariance of $\boldsymbol{\phi}_0$. The estimation is possible by means of the so-called expectation-maximization algorithm (see *Shumway and Stoffer* [1982] and *Loaiciga and Mariño* [1985]).

## Appendix A: Derivation of the 2SLS Equation

Equation (36) is obtained by first transposing (32) to yield

$$\begin{bmatrix} \boldsymbol{\phi}_1^T \\ \vdots \\ \boldsymbol{\phi}_n^T \end{bmatrix} \Psi^T = - \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} \Gamma^T + \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \tag{A1}$$

or in compact form, after postmultiplying by $(\Psi^T)^{-1}$,

$$\underset{(n \times G)}{\Phi} = X[-\Gamma^T(\Psi^T)^{-1}] + V = \underset{(n \times K)}{X} \; \underset{(K \times G)}{R} + \underset{(n \times G)}{V} \tag{A2}$$

in which $V$ equals the product of the right-hand side matrix in (A1) times $(\Psi^T)^{-1}$. Incidentally, $R$ equals $\Pi^T$, $\Pi$ being defined in (23). The multivariate regression model of (A2) can be solved for $R$ as follows (see, for example, *Anderson* [1984]),

$$\hat{R} = (X^TX)^{-1}X^T\Phi \tag{A3}$$

in which $\hat{R}$ denotes an estimator for $R$. Let $\hat{R}_j$ be the following submatrix of $\hat{R}$;

$$\hat{R}_j = [\mathbf{r}_1, \cdots, \mathbf{r}_{j-1}, \mathbf{r}_{j+1}, \cdots, \mathbf{r}_G\Delta] \tag{A4}$$

$\hat{R}_j$ equals the matrix $\hat{R}$ with its $j$th column suppressed. It follows from (A2) that

$$\Phi - V = XR \tag{A5}$$

Since the matrix of disturbances $V$ is unobservable, one can approximate the left-hand side of (A5) by

$$\Phi - \hat{V} = X\hat{R} \tag{A6}$$

in which $\hat{V}$ is the matrix of residuals obtained from the multivariate regression in (A3), i.e.,

$$\hat{V} = \Phi - X\hat{R} \tag{A7}$$

From (A6) it is clear that by deleting the $j$th column,

$$\Phi_j - \hat{V}_j = X\hat{R}_j \tag{A8}$$

Finally, (33) can be transformed to

$$\begin{aligned} \boldsymbol{\phi}_j &= [\Phi_j - \hat{V}_j]\boldsymbol{\psi}_j + X_j\boldsymbol{\gamma}_j + (\mathbf{u}_j + \hat{V}_j\boldsymbol{\psi}_j) \\ &= [X\hat{R}_j \vdots X_j]\begin{bmatrix} \boldsymbol{\psi}_j \\ \boldsymbol{\gamma}_j \end{bmatrix} + \mathbf{w}_j \\ &= \hat{Z}_j\boldsymbol{\beta}_j + \mathbf{w}_j \end{aligned} \tag{A9}$$

Since $\Phi_j - \hat{V}_j$ and $\mathbf{u}_j + \hat{V}_j\psi_j$ $(=\mathbf{w}_j)$ are asymptotically uncorrelated (the probability limit of $\Phi_j - \hat{V}_j$ converges to $XR_j$, which is uncorrelated with $\mathbf{w}_j$), the OLS method applied to (A9) yields consistent estimators of $\boldsymbol{\beta}_j$, as is shown in section 4. Equation (A9) is identical to the expression in (36), which is used to derive the 2SLS estimators.

## Appendix B: Gradient and Hessian of the Negative Log-Likelihood Function

In the implementation of the Newton-Raphson method, the gradient of $f$ (see equation (48)) $\nabla f$ and the matrix of second derivatives of $f$ with respect to the parameters $\theta_i$, $G$, are needed. In order to compute $\nabla f$ and $G$ at $\boldsymbol{\theta}_k$, the following matrix derivative results are useful [*Graybill*, 1983]:

$$\frac{\partial \ln |A|}{\partial \theta_i} = \operatorname{Tr}\left[A^{-1}\frac{\partial A}{\partial \theta_i}\right] \qquad |A| > 0 \tag{B1}$$

$$\frac{\partial A^{-1}}{\partial \theta_i} = -A^{-1}\frac{\partial A}{\partial \theta_i}A^{-1} \tag{B2}$$

$$\frac{\partial^2 \ln |A|}{\partial \theta_i^2} = \operatorname{Tr}\left[-A^{-1}\frac{\partial A}{\partial \theta_i}A^{-1}\frac{\partial A}{\partial \theta_i} + A^{-1}\frac{\partial^2 A}{\partial \theta_i^2}\right] \tag{B3}$$

in which $A$ should be replaced by $\hat{\Sigma}$ or $\Psi$ (see equation (48)) in actual computations.

By means of (B1)–(B3), the first and second derivatives of the negative log-likelihood function $f$ are

$$\frac{\partial f}{\partial \theta_i} = \frac{n}{2}\operatorname{Tr}\left(\hat{\Sigma}^{-1}\frac{\partial \hat{\Sigma}}{\partial \theta_i}\right) - n\operatorname{Tr}\left(\Psi^{-1}\frac{\partial \Psi}{\partial \theta_i}\right) \tag{B4}$$

$$\frac{\partial^2 f}{\partial \theta_i^2} = \frac{n}{2}\operatorname{Tr}\left[-\hat{\Sigma}^{-1}\frac{\partial \hat{\Sigma}}{\partial \theta_i}\hat{\Sigma}^{-1}\frac{\partial \hat{\Sigma}}{\partial \theta_i} + \hat{\Sigma}^{-1}\frac{\partial^2 \hat{\Sigma}}{\partial \theta_i^2}\right] - n\operatorname{Tr}\left[-\Psi^{-1}\frac{\partial \Psi}{\partial \theta_i}\Psi^{-1}\frac{\partial \Psi}{\partial \theta_i}\right] \tag{B5}$$

$$\frac{\partial^2 f}{\partial \theta_j \partial \theta_i} = \frac{n}{2}\operatorname{Tr}\left[-\hat{\Sigma}^{-1}\frac{\partial \hat{\Sigma}}{\partial \theta_j}\hat{\Sigma}^{-1}\frac{\partial \hat{\Sigma}}{\partial \theta_i} + \hat{\Sigma}^{-1}\frac{\partial^2 \hat{\Sigma}}{\partial \theta_j \partial \theta_i}\right] - n\operatorname{Tr}\left[-\Psi^{-1}\frac{\partial \Psi}{\partial \theta_j}\Psi^{-1}\frac{\partial \Psi}{\partial \theta_i}\right] \qquad j \neq i \tag{B6}$$

The gradient $\nabla f$ is the $(q \times 1)$ vector whose elements are $\partial f/\partial \theta_i$, and the matrix of second derivatives $(G)$ is the $(q \times q)$ matrix whose elements are $\partial^2 f/\partial \theta_i \partial \theta_j$.

## Notation

- $A$ $G \times G$ matrix in the continuous-time groundwater flow equation.
- $B$ $G \times G$ matrix in the continuous-time groundwater flow equation.
- $C$ $G \times G$ matrix in the continuous-time groundwater equation; the value of its characteristic roots determine the stability of the continuous-time flow process.
- $\mathbf{e}_t$ $G \times 1$ error term vector in the discretized groundwater equation.
- $F$ pumping rate.
- $\mathbf{F}_t$ $G \times 1$ vector of inputs in the discretized groundwater flow equation at time $t$.
- $f$ negative log-likelihood function.
- $G$ number of structural equations.
- $G^{\Delta}$ number of nonzero structural parameters in matrix $\Psi$ of any structural equation.
- $I(\ )$ Fisher information matrix.
- $i(\ )$ sample information matrix.
- $j$ index to denote anyone of the structural equations $(j = 1, 2, \cdots, G)$.
- $K$ column dimension of the structural matrix $\Gamma$.
- $K^*$ number of nonzero parameters in matrix $\Gamma$ of any structural equation.
- $L$ log-likelihood function.
- $n$ number of time periods.
- $R$ $K \times G$ matrix of parameters in the multivariate regression of heads $(R = \Pi^T)$.
- $\hat{R}$ $K \times G$ estimator matrix of $R$.
- $\hat{R}_j$ $K \times (G^{\Delta} - 1)$ submatrix of $\hat{R}$ in the $j$th structural equation.
- $S$ storativity.
- $S^{(i)}$ storativity within the $i$th element.
- $T$ transmissivity.
- $T^{(i)}$ transmissivity within the $i$th element.
- $t$ time index $(t = 1, 2, \cdots, n)$.
- $\mathbf{u}_t$ $G \times 1$ error vector in the discretized flow equation at time $t$.
- $\mathbf{u}_j$ $n \times 1$ error term in the $j$th structural equation.
- $V$ $n \times G$ error matrix in the multivariate regression model for heads.
- $\hat{V}$ $n \times G$ estimator for $V$.
- $\mathbf{v}_t$ $G \times 1$ error term in the observation process of the state-space model at time $t$.
- $\mathbf{w}_j$ $n \times 1$ error term in the transformed $j$th structural equation of the 2SLS method.
- $\mathbf{w}$ $nG \times 1$ error term in the 3SLS method.
- $X$ $n \times K$ regression matrix in the multivariate regression model for heads.
- $X_j$ $n \times K^*$ regression submatrix in the $j$th structural equation.
- $\mathbf{x}_t$ $K \times 1$ vector of predetermined variables in the discretized flow equation at time $t$.
- $Z_j$ $n \times (G^{\Delta} - 1 + K^*)$ regression matrix in the $j$th structural equation.
- $\hat{Z}$ $nG \times (\sum_{j=1}^{G} G_j^{\Delta} - 1 + K_j^*)$ regression matrix in the 3SLS method.
- $\hat{Z}_j$ $n \times (G^{\Delta} - 1 + K^*)$ transformed regression matrix in the $j$th structural equation of the 2SLS method.
- $\boldsymbol{\beta}$ $(\sum_{j=1}^{G} (G_j^{\Delta} - 1 + K_j^*)) \times 1$ parameter vector in the 3SLS method.
- $\boldsymbol{\beta}_j$ $(G^{\Delta} - 1 + K^*) \times 1$ vector of parameters in the $j$th structural equation.
- $\hat{\boldsymbol{\beta}}_j$ $(G^{\Delta} - 1 + K^*) \times 1$ estimator of $\boldsymbol{\beta}_j$ in the OLS method.
- $\tilde{\boldsymbol{\beta}}_j$ $(G^{\Delta} - 1 + K^*) \times 1$ estimator of $\boldsymbol{\beta}_j$ in the 2SLS method.
- $\bar{\boldsymbol{\beta}}$ $(\sum_{j=1}^{G} (G_j^{\Delta} - 1 + K_j^*)) \times 1$ estimator of $\beta$ in the 3SLS method.
- $\Gamma$ $G \times K$ matrix of structural parameters.
- $\boldsymbol{\gamma}_j$ $K^* \times 1$ vector of parameters (the $j$th row of the matrix $\Gamma$).
- $\gamma_{ij}$ $ij$th element of $\Gamma$.
- $\boldsymbol{\theta}$ parameter vector $(q \times 1)$.
- $\boldsymbol{\theta}^*$ ML parameter vector.
- $\tilde{\boldsymbol{\theta}}$ true and unknown parameter vector.
- $\lambda_i$ $i$th eigenvalue of the $G \times G$ matrix $C$ of the continuous-time flow equation.
- $\boldsymbol{\xi}_t$ $G \times 1$ error term in the state equation.
- $\Pi$ $G \times K$ matrix of parameters in the discrete time flow equation.
- $\rho$ parameter in the autoregressive error term.
- $\Sigma$ $G \times G$ covariance matrix of $\mathbf{u}_t$.

$\sigma_{ij}$ $ij$th element of the covariance matrix $\Sigma$.
$\hat{\sigma}_{ij}$ estimator of $\sigma_{ij}$.
$\Phi$ $n \times G$ matrix of piezometric heads in the multivariate regression model.
$\Phi_j$ $n \times (G^\Delta - 1)$ matrix of piezometric heads in the $j$th structural equation.
$\phi_j$ $n \times 1$ vector of piezometric heads in the $j$th structural equation.
$\phi_t$ $G \times 1$ vector of piezometric heads at time $t$.
$\dot{\phi}_t$ $G \times 1$ time derivative of vector $\phi_t$.
$\chi$ chi-squared random variable.
$\Psi$ $G \times G$ matrix of structural parameters.
$\psi_{ij}$ $ij$th element of $\Psi$.
$\psi_j$ $(G_j^\Delta - 1) \times 1$ vector of parameters (the $j$th row of the matrix $\Psi$).
$\omega$ weighting factor in the discretized equation.

*Acknowledgment.* The research leading to this report was supported by the University of California, Water Resources Center, as part of Water Resources Center Project UCAL-WRC-W-634.

## REFERENCES

Aboufirassi, M., and M. A. Mariño, A geostatistically based approach to the identification of aquifer transmissivities in Yolo Basin, California, *Math. Geol.*, *16*(2), 125–137, 1984.
Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, 2nd. ed., John Wiley, New York, 1984.
Basmann, R. L., A generalized classical method of linear estimation of coefficients in a structural equation, *Econometrica*, *25*, 77–83, 1957.
Bickel, P. J., and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Topics*, Holden-Day, San Francisco, Calif., 1977.
Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum likelihood method incorporating prior information, *Water Resour. Res.*, *22*(2), 199–210, 1986*a*.
Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 2, Uniqueness, stability, and solution algorithms, *Water Resour. Res.*, *22*(2), 211–227, 1986*b*.
Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 3, Application to synthetic and field data, *Water Resour. Res.*, *22*(2), 228–242, 1986*c*.
Cooley, R. L., A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 1, Theory and numerical properties, *Water Resour. Res.*, *13*(2), 318–324, 1977.
Cooley, R. L., A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 2, Applications of statistical analysis, *Water Resour. Res.*, *15*(3), 603–617, 1979.
Cooley, R. L., Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 1, Theory, *Water Resour. Res.*, *18*(4), 965–976, 1982.
Efron, B., and D. V. Hinkley, Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika*, *65*, 457–487, 1978.
Gantmacher, F. R., *The Theory of Matrices*, vol. 1, Chelsea, New York, 1959.
Gill, P. E., W. Murray, and M. H. Wright, *Practical Optimization*, Academic, Orlando, Fla., 1981.
Graybill, F. A., *Matrices with Applications in Statistics*, 2nd ed., Wadsworth International Group, Belmont, Calif., 1983.
Hoeksema, R. J., and P. K. Kitanidis, An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resour. Res.*, *20*(7), 1003–1020, 1984.
Hoerl, A. E., and R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, *12*(1), 69–82, 1970.
Judge, G. G., R. C. Hill, W. E. Giffiths, H. Lütkepohl, and T.-C. Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley, New York, 1982.
Kitanidis, P. K., and E. G. Vomvoris, A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations, *Water Resour. Res.*, *19*(3), 677–690, 1983.
Kmenta, J., *Elements of Econometrics*, MacMillan, New York, 1971.
Loaiciga, H. A., and M. A. Mariño, An approach to parameter estimation and stochastic control in water resources with an application to reservoir operation, *Water Resour. Res.*, *21*(11), 1575–1584, 1985.
Loaiciga, H. A., and M. A. Mariño, Estimation and inference in the inverse problem, in *Proceedings of Water Forum '86: World Water Issues in Evolution*, pp. 973–980, American Society of Civil Engineers, New York, 1986.
Mariño, M. A., and J. N. Luthin, *Seepage and Groundwater*, *Developments in Water Science*, vol. 13, Elsevier Science, New York, 1982.
Marquardt, D. W., and R. D. Snee, Ridge regression in practice, *Am. Stat.*, *29*(1), 3–20, 1975.
McLaughlin, D., Investigation of alternative procedures for estimating groundwater basin parameters, report prepared for the Office of Water Resources and Technology, Water Resour. Eng., Walnut Creek, Calif., 1975.
Neuman, S. P., and S. Yakowitz, A statistical approach to the inverse problem of aquifer hydrology, 1, Theory, *Water Resour. Res.*, *15*(4), 845–860, 1979.
Numan, S. P., A statistical approach to the inverse problem of aquifer hydrology, 3, Improved solution method and added perspective, *Water Resour. Res.*, *16*(2), 331–346, 1980.
Pinder, G. F., and E. O. Frind, Application of Galerkin's procedure to aquifer analysis, *Water Resour. Res.*, *8*(1), 108–120, 1972.
Polak, E., and E. Wong, *A First Course on Linear Systems*, Van Nostrand-Reinhold, New York, 1970.
Rao, C. R., *Linear Statistical Inference and Its Applications*, 2nd ed., John Wiley, New York, 1973.
Sadeghipour, J., and W. W.-G. Yeh, Parameter identification of groundwater aquifer models: A generalized least squares approach, *Water Resour. Res.*, *20*(7), 971–979, 1984.
Shumway, R. H., and D. S. Stoffer, An approach to time series smoothing and forecasting using the EM algorithm, *J. Time Ser.*, *3*(4), 253–264, 1982.
Yakowitz, S., and L. Duckstein, Instability in aquifer identification: Theory and case studies, *Water Resour. Res.*, *16*(6), 1045–1064, 1980.
Yeh, W. W.-G., and Y. S. Yoon, Aquifer parameter identification with optimum dimension in parametrization, *Water Resour. Res.*, *17*(3), 664–672, 1981.
Yeh, W. W.-G., Y. S. Yoon, and K. S. Lee, Aquifer parameter identification with kriging and optimum parametrization, *Water Resour. Res.*, *19*(1), 225–233, 1983.
Zellner, A., and H. Theil, Three stage least squares: Simultaneous estimation of simultaneous equations, *Econometrica*, *30*, 54–78, 1962.

H. A. Loaiciga, Department of Geological Sciences, Wright State University, Dayton, OH 45435.

M. A. Mariño, Department of Land, Air, and Water Resources and Department of Civil Engineering, University of California, Davis, CA 95616.

(Received October 4, 1985;
revised July 17, 1986;
accepted August 26, 1986.)