102323

WATER RESOURCES RESEARCH, VOL. 23, NO. 7, PAGES 1237–1250, JULY 1987

Simultaneous Confidence and Prediction Intervals for Nonlinear Regression Models With Application to a Groundwater Flow Model

ALDO V. VECCHIA

Department of Mathematics, Colorado School of Mines, Golden

RICHARD L. COOLEY

U.S. Geological Survey, Lakewood, Colorado

Methods are presented for computing three types of simultaneous confidence and prediction intervals (exact, likelihood ratio, and linearized) on output from nonlinear regression models with normally distributed residuals. The confidence intervals can be placed on individual regression parameters or on the true regression function at any number of points in the domain of the independent variables, and the prediction intervals can be placed on any number of future observations. The confidence intervals are analogous to simultaneous Scheffè intervals for linear models and the prediction intervals are analogous to the prediction intervals of Hahn (1972). All three types of intervals can be computed efficiently by using the same straightforward Lagrangian optimization scheme. The prediction intervals can be treated in the same computational framework as the confidence intervals by including the random errors as pseudoparameters in the Lagrangian scheme. The methods are applied to a hypothetical groundwater model for flow to a well penetrating a leaky aquifer. Three different data sets are used to demonstrate the effect of sampling strategies on the intervals. For all three data sets, the linearized confidence intervals are inferior to the exact and likelihood ratio intervals, with the latter two being very similar; however, all three types of prediction intervals yielded similar results. The third data set (time drawdown data at only a single observation well) points out many of the problems that can arise from extreme nonlinear behavior of the regression model.

INTRODUCTION

Nonlinear regression modeling has become a valuable tool for investigating complex physical systems in which uncertainties in describing and measuring the system preclude the use of a deterministic modeling scheme. The nonlinear regression model supposes that a set of observations $\{Y_i, 1 \le i \le n\}$ of the physical system are related to a $p \times 1$ vector of unknown parameters $\boldsymbol{\beta}$ through the stochastic model

$$\mathbf{Y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{Y}$ is an $n \times 1$ observation vector, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random variables assumed to satisfy the distributional structure

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}; \sigma^2\boldsymbol{\omega}^{-1}) \tag{2}$$

where $\boldsymbol{\omega}$ is a known $n \times n$, positive definite, reliability matrix. The vector of regression functions $\mathbf{f}(\boldsymbol{\beta})$ will depend on measurements of one or more independent variables, but for notational convenience we do not explicitly denote this dependence in (1). We are concerned mainly with models that are nonlinear in all of the parameters $\boldsymbol{\beta}$. Although the methods presented in this paper remain valid for models that are linear in a subset of the parameters, in such cases more efficient procedures could be developed to take advantage of the partial linearity of the model (see, for instance, *Williams* [1962] and *Halperin* [1963]).

Consider the problem of statistical inference for (1). The investigator is usually interested in estimating a set of scalar functions of the parameters $\{g_1(\boldsymbol{\beta}), g_2(\boldsymbol{\beta}), \cdots\}$, where a typical $g_i(\boldsymbol{\beta})$ might be a single component of $\boldsymbol{\beta}$, the expected value of an observation Y_K at a particular setting of the independent variables (i.e., $f_K(\boldsymbol{\beta})$), or some other function of interest. A point estimate of $g_i(\boldsymbol{\beta})$ is a single random variable $\hat{g}_i$ that is the best estimate of $g_i(\boldsymbol{\beta})$ according to certain criteria. Optimally, one would like to have $\hat{g}_i$ be unbiased ($E\hat{g}_i = g_i(\boldsymbol{\beta})$) and have the smallest variance among all unbiased estimators. If (1) were linear in $\boldsymbol{\beta}$ and $g_i(\boldsymbol{\beta})$ a linear function of $\boldsymbol{\beta}$, then $\hat{g}_i = g_i(\hat{\mathbf{b}})$ is the minimum variance unbiased estimator of $g_i(\boldsymbol{\beta})$, where $\hat{\mathbf{b}}$ is the weighted least squares (WLS) estimator of $\boldsymbol{\beta}$. However, for nonlinear models, the WLS estimator is biased and $g_i(\boldsymbol{\beta})$ may not be a linear function of $\boldsymbol{\beta}$. Hence the statistical properties of $g_i(\hat{\mathbf{b}})$ may be difficult to obtain and, if the model is highly nonlinear, $g_i(\hat{\mathbf{b}})$ may, in fact, be a poor estimator of $g_i(\boldsymbol{\beta})$.

A set of $(1 - \alpha)$ 100% simultaneous confidence intervals for $g_1(\boldsymbol{\beta}), g_2(\boldsymbol{\beta}), \cdots$ is defined as a set of intervals $(\hat{L}_i, \hat{U}_i)$ such that the probability is $1 - \alpha$ that $g_i(\boldsymbol{\beta})$ lies within $(\hat{L}_i, \hat{U}_i)$ for all i simultaneously, no matter how many intervals are computed. Because they include information on model reliability, simultaneous confidence intervals on $g_i(\boldsymbol{\beta})$ are much more informative than point estimates. Hence it may not be crucial to solve the difficult problem of determining statistical properties of point estimators for nonlinear models provided efficient methods for computing simultaneous confidence intervals can be developed. In the framework of a linear regression, simultaneous confidence intervals are known as Scheffè intervals. They can be easily computed (see, for instance, *Graybill* [1976, pp. 198–200]), and the point estimates $g_i(\hat{\mathbf{b}})$ lie in the center of the intervals. For a nonlinear model, $g_i(\hat{\mathbf{b}})$ may not lie in the center of the interval, and formulas for $(\hat{L}_i, \hat{U}_i)$ are not avail-

This paper is not subject to U.S. copyright. Published in 1987 by the American Geophysical Union.

Paper number 6W4647.

31118

able. The usual procedure to obtain approximate intervals is to replace the nonlinear model with a linearized approximation in the vicinity of $\hat{\mathbf{b}}$, then proceed with the linear analysis. In the context of groundwater models, this is the only type of analysis that has been published [*Yeh*, 1986], although *Carrera* [1984, pp. 65–68] has outlined a possible procedure for partially correcting the linear intervals for the effects of nonlinearity.

In this paper, we present new methods of calculating true nonlinear confidence intervals. In addition, we extend the methods to allow computation of simultaneous prediction intervals on any number of future observations from (1). Computation of the confidence and prediction intervals involves finding extreme values of the functions $g_i(\boldsymbol{\beta})$ over a joint confidence region on $\boldsymbol{\beta}$. Three different joint confidence regions are employed, one exact region (the lack of fit region) and two approximate regions (the likelihood ratio and linearized regions). The three regions are reviewed in the next section. Under mild conditions to be explained below, determining extreme values of $g_i(\boldsymbol{\beta})$ over the various joint confidence regions can be reduced to a simple general type of Lagrangian optimization problem with a single constraint. Calculation of simultaneous prediction intervals can also be placed in the same simple framework by including the random errors as pseudoparameters in the Lagrangian formulation. The method that we develop to compute the confidence and prediction intervals can easily be imbedded in a standard nonlinear regression algorithm. In the present study we use the Gauss-Newton optimization method. Because the optimization problem is nonlinear, convergence problems sometimes arise. Thus we present some conditioning and oscillation-dampening procedures to minimize these difficulties in the numerical implementation of the scheme. Finally, we apply the techniques to calculate simultaneous confidence and prediction intervals on the output from a hypothetical groundwater model for flow to a well penetrating a leaky aquifer.

SOME TYPES OF NONLINEAR CONFIDENCE REGIONS

In this section we review several methods for obtaining joint confidence regions on the parameters of (1). Confidence regions on $\boldsymbol{\beta}$ may be classified into exact regions and approximate regions. An exact $(1 - \alpha)$ 100% confidence region for $\boldsymbol{\beta}$ is a region in p-dimensional Euclidean space, say, R_α, that depends on $\mathbf{Y}$ and for which Prob $[\boldsymbol{\beta} \in R_\alpha] = 1 - \alpha$ holds exactly. An approximate $(1 - \alpha)$ 100% confidence region for $\boldsymbol{\beta}$, say, R_α^*, satisfies $\lim_{n\to\infty}$ Prob $[\boldsymbol{\beta} \in R_\alpha^*] = 1 - \alpha$, but for finite n the true confidence level may be different than $1 - \alpha$. We consider first exact regions.

Exact Regions

Let $\mathbf{X}$ be the sensitivity matrix

$$\mathbf{X} = \{\partial f_i(\mathbf{b})/\partial b_j\} \tag{3}$$

which is $n \times p$ and, in general, depends on $\mathbf{b}$. We assume throughout that $\mathbf{X}$ is of full column rank and that $n \geq p + 1$. Define a symmetric idempotent matrix of rank p,

$$\mathbf{P} = \boldsymbol{\omega}^{1/2}\mathbf{X}(\mathbf{X}^T\boldsymbol{\omega}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\omega}^{1/2} \tag{4}$$

where $\boldsymbol{\omega}$ is defined in (2). Note that the marix $\mathbf{I} - \mathbf{P}$ is symmetric idempotent of rank $n - p$, and $\mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{0}$. Hence from *Graybill* [1976, theorems 4.4.1 and 4.5.3],

$$\sigma^{-2}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})]^T\boldsymbol{\omega}^{1/2}\mathbf{P}\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})] \sim \chi^2(p) \tag{5}$$

independently of

$$\sigma^{-2}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})]^T\boldsymbol{\omega}^{1/2}(\mathbf{I} - \mathbf{P})\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})] \sim \chi^2(n - p) \tag{6}$$

where $U \sim \chi^2(k)$ means that the random variable U has a χ^2 distribution with k degrees of freedom. Since the ratio of two independent χ^2 random variables standardized by their degrees of freedom has an F distribution, a $(1 - \alpha)$ 100% exact confidence region for $\boldsymbol{\beta}$ is given by the set of values $\mathbf{b}$ satisfying

$$\frac{[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}\mathbf{P}\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]}{[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}(\mathbf{I} - \mathbf{P})\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]} \leq \frac{p}{n - p} F_\alpha(p, n - p) \tag{7}$$

where $F_\alpha(p, n - p)$ denotes the upper α probability point of an F distribution with p and $n - p$ degrees of freedom.

Basing confidence regions on (7) with $\mathbf{P}$ defined by (3) and (4) is called the lack of fit method. The properties of the lack of fit confidence region and its advantages and disadvantages over other exact methods have been investigated in detail in many previous studies (see, for example, *Donaldson and Schnabel* [1987], *Sundararaj* [1978], *Wallace and Grant* [1977], and *Gallant* [1976]). One advantage is that the boundary of the lack of fit confidence region often coincides closely with a contour of the sum of squares function, $S(\mathbf{b}) = [\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]$, or equivalently, a contour of the likelihood function:

$$\frac{1}{(2\pi\sigma^2)^{n/2}} |\boldsymbol{\omega}|^{1/2} \exp\left\{-\frac{1}{2\sigma^2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]\right\} \tag{8}$$

Thus values of $\mathbf{b}$ that fall in the confidence region tend to result in a smaller sum of squared error and greater likelihood than those values outside the confidence region. Another advantage of the lack of fit method is that the only calculations involving the regression model are the calculations to obtain the sensitivities, which are also needed for some of the commonly used iterative nonlinear regression routines.

Lack of fit confidence regions have two principal drawbacks. First, they automatically include all critical points of $S(\mathbf{b})$, even relative maxima. Hence if $S(\mathbf{b})$ has multiple critical points, the confidence region could consist of several disjoint regions, some of which result in a poor fit to the data. The second drawback is that the sensitivity matrix depends on $\mathbf{b}$, unless the model is linear. This may cause the computational requirements for finding the boundary of the region to become large.

Another method of obtaining an exact or nearly exact confidence region for parameters was suggested by *Hartley* [1964]. Hartley's method is based on obtaining a reparameterization, say, $\boldsymbol{\tau}(\mathbf{b})$, under which the model is nearly linear; that is, for which

$$\mathbf{f}(\mathbf{b}) \simeq \mathbf{X}^*\boldsymbol{\tau}(\mathbf{b}) \tag{9}$$

with $\mathbf{X}^* = \{\partial f_i(\mathbf{b})/\partial \tau_j(\mathbf{b})\}$ independent of $\mathbf{b}$. Hartley points out that the boundary of the confidence region for $\boldsymbol{\tau}(\mathbf{b})$ obtained by this reparameterization method will coincide closely with a contour of $S(\mathbf{b})$, with exact coincidence for models in which the reparameterization is exact. This method has a major computational advantage over the lack of fit method in that $\mathbf{X}^*$ does not depend on $\mathbf{b}$. However, determination of a proper choice for $\boldsymbol{\tau}(\mathbf{b})$ may be quite difficult. If, in a particular case, there is an obvious reparameterization that results in a linear (or nearly linear) model, the investigator might consider this method.

Approximate Regions

An approximate region may be obtained by using the likelihood ratio method. It is based on the fact that for large n, $[(n-p)/p][S(\boldsymbol{\beta}) - S(\hat{\mathbf{b}})]/S(\hat{\mathbf{b}})$ has an approximate F distribution with p and $n-p$ degrees of freedom [see *Rao*, 1973, pp. 417–419]. Thus a $(1-\alpha)$ 100% likelihood ratio confidence region for $\boldsymbol{\beta}$ is given by

$$S(\mathbf{b}) \le S(\hat{\mathbf{b}})\left[\frac{p}{n-p} F_\alpha(p, n-p) + 1\right] \tag{10}$$

The boundary of this region always coincides with a contour of the sum of squares function.

The simplest method for obtaining an approximate confidence region is the linearization method. It is based on approximating the nonlinear function $\mathbf{f}(\mathbf{b})$ contained in $S(\mathbf{b}) = [\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]$ by its first-order Taylor series expansion in the neighborhood of $\hat{\mathbf{b}}$, the weighted least squares estimator of $\boldsymbol{\beta}$. That is,

$$\mathbf{f}(\mathbf{b}) \simeq \mathbf{f}(\hat{\mathbf{b}}) + \hat{\mathbf{X}}(\mathbf{b} - \hat{\mathbf{b}}) \tag{11}$$

where $\hat{\mathbf{X}}$ is the $n \times p$ sensitivity matrix evaluated at $\mathbf{b} = \hat{\mathbf{b}}$. The linearized confidence region is derived by using (11) in (10) and is given by all $\mathbf{b}$ for which

$$(\mathbf{b} - \hat{\mathbf{b}})^T\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}(\mathbf{b} - \hat{\mathbf{b}}) \le ps^2F_\alpha(p, n-p) \tag{12}$$

where $s^2 = S(\hat{\mathbf{b}})/(n-p)$. The boundary of (12) forms an ellipsoid centered about $\hat{\mathbf{b}}$.

Based on Monte Carlo studies (see, for instance, *Donaldson and Schnabel* [1987]) the actual probability that $\boldsymbol{\beta}$ lies within the likelihood ratio region (10) generally coincides closely to the specified probability level of $(1-\alpha)$. In contrast, the actual probability that $\boldsymbol{\beta}$ lies within the linearized region (12) may differ considerably from $(1-\alpha)$. The key to this difference between the two regions is that the boundary of (10) is a true probability contour (that is, a contour of equation (8)), while the boundary of (12) is not a true probability contour because it involves assuming that the contours of (8) are elliptical in the vicinity of $\hat{\mathbf{b}}$ [*Draper and Smith*, 1981, pp. 472–473].

NONLINEAR CONFIDENCE AND PREDICTION INTERVALS

In a nonlinear regression analysis, the investigator is interested not only in obtaining confidence regions on model parameters, but also in obtaining simultaneous confidence intervals on various functions of the model parameters, such as the expected value of the dependent variable (i.e., $\mathbf{f}(\boldsymbol{\beta})$) at several points in the domain of the independent variables. In addition, simultaneous confidence intervals on several future observations of the dependent variable Y or of some other quantity such as a flow rate may be desired. The latter intervals are called simultaneous prediction intervals, since they place probability limits on the values of a set of random variables, rather than fixed quantities. In this section, we show how the exact and approximate joint confidence regions on parameters developed in the previous section may be used to derive simultaneous confidence intervals and prediction intervals. In the following sections, we give operational methods for actually computing the nonlinear intervals.

Exact Intervals

Let $f_K(\boldsymbol{\beta})$ be the true value of the regression function at any fixed point in the domain of the independent variables. An exact simultaneous $(1-\alpha)$ 100% confidence interval on $f_K(\boldsymbol{\beta})$ is given by

$$\left(\min_{\mathbf{b}} f_K(\mathbf{b}), \max_{\mathbf{b}} f_K(\mathbf{b})\right) \qquad \mathbf{b} \in R_\alpha \tag{13}$$

where $\mathbf{b} \in R_\alpha$ signifies that $\mathbf{b}$ may range over the entire exact confidence region defined by (3), (4), and (7). The resulting interval is a simultaneous, rather than individual, interval because (13) can be computed at any number of points to yield simultaneous $(1-\alpha)$ 100% confidence intervals on the regression function. Hence the probability that an individual value $f_K(\boldsymbol{\beta})$ falls within its interval may exceed $(1-\alpha)$.

To derive simultaneous prediction intervals, consider first the case of placing a prediction interval on a single future value of the dependent variable at point K:

$$Y_K = f_K(\boldsymbol{\beta}) + \varepsilon_K \tag{14}$$

To simplify the derivations that follow, we assume that the random error ε_K is independent of $\boldsymbol{\varepsilon}$ in (1); that is,

$$\operatorname{Var}\begin{bmatrix}\boldsymbol{\varepsilon}\\ \varepsilon_K\end{bmatrix} = \sigma^2\begin{bmatrix}\boldsymbol{\omega} & 0\\ 0 & \omega_K\end{bmatrix}^{-1} \tag{15}$$

In this case, $\omega_K\sigma^{-2}\varepsilon_K{}^2$ has a χ^2 distribution with one degree of freedom and is clearly independent of (5) and (6). Therefore the quantity

$$\{(p+1)^{-1}\{[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})]^T\boldsymbol{\omega}^{1/2}\mathbf{P}\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})] + \omega_K\varepsilon_K{}^2\}\} \cdot \{(n-p)^{-1}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})]^T\boldsymbol{\omega}^{1/2}(\mathbf{I} - \mathbf{P})\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})]\}^{-1} \tag{16}$$

has an F distribution with $p+1$ and $n-p$ degrees of freedom and a $(1-\alpha)$ 100% prediction region for the vector $(\boldsymbol{\beta}^T, \varepsilon_K)^T$ is given by

$$\frac{[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}\mathbf{P}\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})] + \omega_K e_K{}^2}{[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}(\mathbf{I} - \mathbf{P})\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]} \le \frac{p+1}{n-p} F_\alpha(p+1, n-p) \tag{17}$$

where e_K is the dummy value for ε_K. A $(1-\alpha)$ 100% prediction interval for Y_K is then obtained by finding extreme values of $f_K(\mathbf{b}) + e_K$ over the region (17). Note that the resulting prediction interval is conservative. That is, the probability that an individual observation Y_K falls within its prediction interval may exceed $(1-\alpha)$ because the underlying interval on $f_K(\boldsymbol{\beta})$ is simultaneous. Methods for calculating the interval are explained in the following section.

To obtain simultaneous prediction intervals on m future observations, the right-hand side of (17) needs to be replaced with a value such that the probability is $(1-\alpha)$ that the statement holds simultaneously for all m points. The correct value may be obtained quite simply via Monte Carlo simulations as follows.

1. Generate a set of mutually independent random variables $\{U, V, W_1, \cdots, W_m\}$, where $U \sim \chi^2(p)$, $V \sim \chi^2(n-p)$, and $W_i \sim \chi^2(1)$ for $1 \le i \le m$.
2. Compute $M = (U + \max_{1 \le i \le m} W_i)/V$.
3. Repeat (1) and (2) a large number of times and find the value $M_\alpha(p, n, m)$ such that a proportion $(1-\alpha)$ of the computed values of M are less than $M_\alpha(p, n, m)$. The resulting value of $M_\alpha(p, n, m)$ may be substituted for the right-hand side of (17) for each of the m points, and (17) may then be used to determine extreme values for each of the observations. Once again, the details are explained in the next section.

Approximate Intervals

A $(1-\alpha)$ 100% approximate simultaneous confidence interval for $f_K(\boldsymbol{\beta})$ can be obtained by finding extreme values of $f_K(\mathbf{b})$ over the likelihood ratio region (10) or the linearized region (12). The resulting intervals hold simultaneously for any number of points. To obtain approximate prediction intervals based on the likelihood ratio method, note that (10) is based on the fact that $\sigma^{-2}(S(\boldsymbol{\beta}) - S(\hat{\mathbf{b}}))$ is approximately $\chi^2(p)$ distributed (for large n) independently of $\sigma^{-2}S(\hat{\mathbf{b}})$ which is approximately $\chi^2(n-p)$ distributed. Hence by similar arguments to those used in obtaining (16),

$$\frac{(p+1)^{-1}[S(\boldsymbol{\beta}) - S(\hat{\mathbf{b}}) + \omega_K \varepsilon_K{}^2]}{(n-p)^{-1}S(\hat{\mathbf{b}})} \tag{18}$$

has an approximate F distribution with $p+1$ and $n-p$ degrees of freedom. Hence a $(1-\alpha)$ 100% approximate prediction region for $(\boldsymbol{\beta}^T, \varepsilon_K)^T$ then becomes

$$S(\mathbf{b}) + \omega_K e_K{}^2 \le S(\hat{\mathbf{b}})\left[\frac{p+1}{n-p} F_\alpha(p+1, n-p) + 1\right] \tag{19}$$

To obtain approximate simultaneous prediction intervals for m future values, (19) should be replaced by

$$S(\mathbf{b}) + \omega_K e_K{}^2 \le S(\hat{\mathbf{b}})[M_\alpha(p, n, m) + 1] \tag{20}$$

where $M_\alpha(p, n, m)$ is obtained from steps (1)–(3) of the above algorithm.

Confidence regions based on linear approximations involve replacing the nonlinear model (1) with a linear approximation in the neighborhood of $\hat{\mathbf{b}}$. Simultaneous prediction intervals for the resultant linearized models maybe obtained using results of previous studies, such as *Lieberman* [1961] or *Hahn* [1972]. For an explanation of how these methods may be applied in a groundwater modeling framework, see *Cooley and Naff* [1985, chapter 5].

General Method for Computing Confidence Intervals

In this section we describe briefly a general method for using a joint confidence region on the parameters to compute nonlinear confidence intervals on various functions of the parameters. In the following section, the general method is specialized for application to the nonlinear confidence intervals developed in the previous section. It is also straightforward to apply the method to nonlinear prediction intervals. The procedures used here are similar to those used in a previous paper [*Cooley and Vecchia*, 1987] to obtain intervals on output from hand-calibrated groundwater models.

Let a scalar function of parameters for which a confidence interval is desired be given by $g(\boldsymbol{\beta})$, and let a problem-dependent continuous vector function of the parameters and data be given by $\mathbf{q}(\mathbf{Y}; \boldsymbol{\beta})$. (The term "problem-dependent" is used to describe a quantity whose precise definition depends on the particular situation to which the general method is to be applied.) Finally, define a $(1-\alpha)$ 100% confidence region for $\boldsymbol{\beta}$ by

$$\text{Prob}\,[\mathbf{q}^T(\mathbf{Y}; \boldsymbol{\beta})\mathbf{A}\mathbf{q}(\mathbf{Y}; \boldsymbol{\beta}) \le d_{1-\alpha}{}^2] = 1-\alpha \tag{21}$$

where $\mathbf{A}$ is a problem-dependent, symmetric matrix that is not a function of $\boldsymbol{\beta}$, and $1-\alpha$ is the probability level. The problem-dependent quantity $d_{1-\alpha}{}^2$ is defined relative to the other terms in (21) to make the probability statement true. Then, a $(1-\alpha)$ 100% simultaneous confidence interval on $g(\boldsymbol{\beta})$ may be computed as

$$\left[\min_{\mathbf{b}} g(\mathbf{b}), \max_{\mathbf{b}} g(\mathbf{b})\right] \tag{22}$$

subject to

$$\mathbf{q}^T(\mathbf{Y}; \mathbf{b})\mathbf{A}\mathbf{q}(\mathbf{Y}; \mathbf{b}) \le d_{1-\alpha}{}^2 \tag{23}$$

where $\mathbf{b}$ is a dummy set of parameters. For notational convenience, from here on we will drop the explicit functional dependence on $\mathbf{Y}$ and write $\mathbf{q}(\mathbf{b})$ instead of $\mathbf{q}(\mathbf{Y}; \mathbf{b})$. Under the mild condition that the gradient of $g(\mathbf{b})$ with respect to $\mathbf{b}$ be nonzero within the closed region (23) [see *Cooley and Vecchia*, 1987] the complicated nonlinear optimization problem posed by (22) and (23) may be replaced by the Lagrangian problem of finding the extreme values of

$$L(\mathbf{b}, \lambda') = g(\mathbf{b}) + \lambda'\mathbf{q}^T(\mathbf{b})\mathbf{A}\mathbf{q}(\mathbf{b}) \tag{24}$$

subject to

$$\mathbf{q}^T(\mathbf{b})\mathbf{A}\mathbf{q}(\mathbf{b}) = d_{1-\alpha}{}^2 \tag{25}$$

where λ' is the Lagrange multiplier.

Differentiating (24) with respect to $\mathbf{b}$ and setting the result to zero yields

$$\mathbf{Z}^T\mathbf{A}\mathbf{q}(\mathbf{b}) = -\lambda\mathbf{a} \tag{26}$$

where

$$\mathbf{Z} = \{Z_{ij}\} = \left\{\frac{\partial q_i(\mathbf{b})}{\partial b_j}\right\} \tag{27}$$

$$\mathbf{a} = \{a_i\} = \left\{\frac{\partial g(\mathbf{b})}{\partial b_i}\right\} \tag{28}$$

$$\lambda = \frac{1}{2\lambda'} \tag{29}$$

Simultaneous solution of (25) and (26) yields two sets of extreme values $\mathbf{b} = \mathbf{b}_e$ from which the desired confidence interval may be calculated.

As in the previous paper [*Cooley and Vecchia*, 1987], we adopt an iterative solution method for solving (25) and (26) by replacing the nonlinear quantity $\mathbf{q}(\mathbf{b})$ with its linear approximation about $\mathbf{b}_0$ [see *Cooley and Vecchia*, 1987, equations (11) and (16)] to yield

$$\mathbf{Z}_0{}^T\mathbf{A}\mathbf{Z}_0(\mathbf{b}_e - \mathbf{b}_0) = -\mathbf{Z}_0{}^T\mathbf{A}\mathbf{q}(\mathbf{b}_0) - \lambda\mathbf{a}_0 \tag{30}$$

$$\lambda = \pm\{[d_{1-\alpha}{}^2 - \mathbf{q}^T(\mathbf{b}_0)\mathbf{A}\mathbf{q}(\mathbf{b}_0) + \mathbf{q}^T(\mathbf{b}_0)\mathbf{A}^T\mathbf{Z}_0(\mathbf{Z}_0{}^T\mathbf{A}\mathbf{Z}_0)^{-1} \cdot \mathbf{Z}_0{}^T\mathbf{A}\mathbf{q}(\mathbf{b}_0)]/\mathbf{a}_0{}^T(\mathbf{Z}_0{}^T\mathbf{A}\mathbf{Z}_0)^{-1}\mathbf{a}_0\}^{1/2} \tag{31}$$

where $\mathbf{b}_0$ is a fixed point and $\mathbf{Z}_0$ and $\mathbf{a}_0$ are evaluated from (27) and (28) with $\mathbf{b} = \mathbf{b}_0$. Recursive solution of (30) and (31) is obtained by setting $\mathbf{b}_0 = \mathbf{b}_e$ at the end of each iteration and resolving for $\mathbf{b}_e$ again on the next iteration, until convergence is achieved. Two separate problems must be solved to obtain the two sets of values $\mathbf{b}_e$, one using the plus sign in (31), the other using the minus sign. Because the optimization problem is nonlinear, it is possible that solutions to (25) and (26) corresponding to local extremes in (24) may exist. To guard against obtaining such a local extreme, the optimization problem should be solved with several different starting values $\mathbf{b}_0$ to check that the same extreme sets $\mathbf{b}_e$ are obtained from each run.

Application of the General Method

The general method of the previous section is now applied to finding a simultaneous (in contrast to individual) confidence

TABLE 1. Definitions for Variables in (30) and (31)

	Exact Region, Equation (36)	Likelihood Region, Equation (37)	Linearized Region, Equation (38)
$\mathbf{q}(\mathbf{b})$	$\mathbf{Y} - \mathbf{f}(\mathbf{b})$	$\mathbf{Y} - \mathbf{f}(\mathbf{b})$	$\mathbf{b} - \hat{\mathbf{b}}$
$\mathbf{A}$	$\boldsymbol{\omega}^{1/2}[\mathbf{P}_0 - D^2(\mathbf{I} - \mathbf{P}_0)]\boldsymbol{\omega}^{1/2}$	$\boldsymbol{\omega}$	$\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}$
$d_{1-\alpha}{}^2$	0	$S(\hat{\mathbf{b}})(D^2 + 1)$	$S(\hat{\mathbf{b}})D^2$
$g(\mathbf{b})$	$b_i, f_K(\mathbf{b})$*	$b_i, f_K(\mathbf{b})$*	$b_i, f_K(\mathbf{b})$*
$\mathbf{Z}_0$	$-\mathbf{X}_0$	$-\mathbf{X}_0$	$\mathbf{I}_{(p \times p)}$
$\mathbf{a}_0$	$\mathbf{i}, \mathbf{X}_{0K}{}^{T*}$	$\mathbf{i}, \mathbf{X}_{0K}{}^{T*}$	$\mathbf{i}, \hat{\mathbf{X}}_K{}^{T*}$

*The first entry is for an interval on β_i, and the second entry is for an interval on $f_K(\boldsymbol{\beta})$.

interval on any parameter β_i or on $f_K(\boldsymbol{\beta})$, the mean of the dependent variable at a fixed point K in the domain of the independent variables. In particular, for a groundwater flow model, $f_K(\boldsymbol{\beta})$ is the hydraulic head computed at point K. We can also find confidence intervals on other output quantities from models, but only intervals on β_i and $f_K(\boldsymbol{\beta})$ are considered in the present paper.

Equations used to compute the confidence intervals are based on equations for the boundaries of the confidence regions on parameters that have been put in the form of (25). By using these equations and the quantities on which the confidence intervals are desired (β_i and $f_K(\boldsymbol{\beta})$), definitions of $\mathbf{q}(\mathbf{b})$, $\mathbf{A}$, $d_{1-\alpha}{}^2$, and $g(\mathbf{b})$ can readily be obtained by inspection. Based on these definitions, specialized forms of (30) and (31) are then computed using definitions (27) and (28).

Boundaries of Confidence Regions

The boundary of exact region (7), with $\mathbf{X}$ being the $n \times p$ sensitivity matrix, satisfies the equation

$$\frac{[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}\mathbf{P}\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]}{[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}(\mathbf{I} - \mathbf{P})\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})]} = \frac{p}{n-p} F_\alpha(p, n-p) \tag{32}$$

or

$$[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}[\mathbf{P} - D^2(\mathbf{I} - \mathbf{P})]\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})] = 0 \tag{33}$$

where

$$D^2 = \frac{p}{n-p} F_\alpha(p, n-p) \tag{34}$$

Equations (25) and (33) have the same general form. However, if (33) were used to calculate confidence intervals, then, contrary to assumption, $\mathbf{A}$, which would be given by $\boldsymbol{\omega}^{1/2}[\mathbf{P} - D^2(\mathbf{I} - \mathbf{P})]\boldsymbol{\omega}^{1/2}$, would be a function of $\mathbf{b}$, because the sensitivities contained in $\mathbf{P}$ are functions of $\mathbf{b}$. This problem may be avoided by fixing the point at which the sensitivities are computed in each iteration. By choosing this point as $\mathbf{b}_0$, it will be seen in the development to follow that an interval based on the lack of fit region results. The matrix $\mathbf{P}$ computed at point $\mathbf{b}_0$ will be referred to as $\mathbf{P}_0$ and is defined as

$$\mathbf{P}_0 = \boldsymbol{\omega}^{1/2}\mathbf{X}_0(\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0)^{-1}\mathbf{X}_0{}^T\boldsymbol{\omega}^{1/2} \tag{35}$$

where $\mathbf{X}_0$ is the $n \times p$ sensitivity matrix evaluated at $\mathbf{b} = \mathbf{b}_0$.

Boundaries for the exact, likelihood, and linearized regions may be stated as follows. From (33) (with $\mathbf{P} = \mathbf{P}_0$) the boundary of the exact region is

$$[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}[\mathbf{P}_0 - D^2(\mathbf{I} - \mathbf{P}_0)]\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})] = 0 \tag{36}$$

The boundary of likelihood region (10) is

$$[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b})] = S(\hat{\mathbf{b}})\left[\frac{p}{n-p} F_\alpha(p, n-p) + 1\right] = S(\hat{\mathbf{b}})(D^2 + 1) \tag{37}$$

Finally, the boundary of linearized region (12) is

$$(\mathbf{b} - \hat{\mathbf{b}})^T\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}(\mathbf{b} - \hat{\mathbf{b}}) = ps^2F_\alpha(p, n-p) = S(\hat{\mathbf{b}})D^2 \tag{38}$$

Equations for Confidence Intervals

Table 1 gives definitions for quantities in (30) and (31) based on comparison of (25) and boundaries (36)–(38). All quantities in Table 1 except $\mathbf{i}$ have been defined previously. Vector $\mathbf{i}$ is defined as

$$\mathbf{i} = [0, 0, \cdots, 1, 0, \cdots]^T \tag{39}$$

where the one is in row i. For all regions, $\mathbf{Z}_0$ was derived by straightforward differentiation of $\mathbf{q}(\mathbf{b})$ using (27) evaluated at $\mathbf{b} = \mathbf{b}_0$. For both the exact and likelihood regions $\mathbf{a}_0$ was similarly obtained by differentiating b_i or $f_K(\mathbf{b})$ directly using (28). For the linearized region, b_i was also differentiated directly for the interval on β_i, but, because the linearized region is based on linearized model (11), $\mathbf{a}_0$ for the interval on $f_K(\boldsymbol{\beta})$ had to be obtained as follows:

$$\mathbf{a}_0 = \left\{\frac{\partial f_K(\mathbf{b})}{\partial b_i}\bigg|_{\mathbf{b}=\mathbf{b}_0}\right\} \simeq \left\{\frac{\partial}{\partial b_i}[f_K(\hat{\mathbf{b}}) + \hat{\mathbf{X}}_K(\mathbf{b} - \hat{\mathbf{b}})]\right\} = \hat{\mathbf{X}}_K{}^T \tag{40}$$

By substituting the appropriate quantities from Table 1 into (30) and (31), the equations for both the exact and likelihood regions are derived as

$$\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0(\mathbf{b}_e - \mathbf{b}_0) = \mathbf{X}_0{}^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)] - \lambda\mathbf{a}_0 \tag{41}$$

$$\lambda = \pm\{[d_{1-\alpha}{}^2 - (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\mathbf{A}(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)) + (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\boldsymbol{\omega}^{1/2} \cdot \mathbf{P}_0\boldsymbol{\omega}^{1/2}(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))]/[\mathbf{a}_0{}^T(\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0)^{-1}\mathbf{a}_0]\}^{1/2} \tag{42}$$

Appropriate definitions for $\mathbf{A}$ and $\mathbf{a}_0$ are obtained from Table 1 depending on the type of interval desired.

The confidence interval on β_i is obtained as the two values of $b_i{}^e$ from the two solutions of (41) and (42) (one solution using the plus sign in (42) and one using the minus sign), and, similarly, the confidence interval on $f_K(\boldsymbol{\beta})$ is obtained as the two values of $f_K(\mathbf{b}_e)$. The values of $f_K(\mathbf{b}_e)$ are directly calculated following solution of (41) and (42). As a final note, when the iterative process based on the exact region converges, $\mathbf{b}_e \simeq \mathbf{b}_0$ so that $\mathbf{P}_0$ is computed nearly at $\mathbf{b}_e$. This agreement of $\mathbf{b}$ used in $\mathbf{f}(\mathbf{b})$ and $\mathbf{b}$ used in $\mathbf{P}$ means that the resulting interval corresponds with the standard lack of fit region defined by (7).

The linearized interval on β_i is

$$\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}(\mathbf{b}_e - \mathbf{b}_0) = -\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}(\mathbf{b}_0 - \hat{\mathbf{b}}) - \lambda\mathbf{i}$$

or

$$\mathbf{b}_e = \hat{\mathbf{b}} - \lambda(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\mathbf{i} \tag{43}$$

$$\lambda = \pm\{d_{1-\alpha}{}^2/[\mathbf{i}^T(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\mathbf{i}]\}^{1/2} \tag{44}$$

The ith equation in the combination of (43) and (44) gives the desired confidence interval on β_i, or

$$b_i{}^e = \hat{b}_i \pm d_{1-\alpha}[\mathbf{i}^T(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\mathbf{i}]^{1/2} \tag{45}$$

Equations analogous to (43) and (44) are obtained for a confidence interval on $f_K(\boldsymbol{\beta})$:

$$\mathbf{b}_e = \hat{\mathbf{b}} - \lambda(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}_K{}^T \tag{46}$$

$$\lambda = \pm\{d_{1-\alpha}{}^2/[\hat{\mathbf{X}}_K(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}_K{}^T]\}^{1/2} \tag{47}$$

By using the linearized model

$$f_K(\mathbf{b}_e) \simeq f_K(\hat{\mathbf{b}}) + \hat{\mathbf{X}}_K(\mathbf{b}_e - \hat{\mathbf{b}}) \tag{48}$$

and premultiplying (46) by $\hat{\mathbf{X}}_K$, (46) and (47) can be combined to give

$$f_K(\mathbf{b}_e) \simeq f_K(\hat{\mathbf{b}}) \pm d_{1-\alpha}[\hat{\mathbf{X}}_K(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}_K{}^T]^{1/2} \tag{49}$$

Note that (49) is the standard Scheffè interval for a linear model (see, for example, *Graybill* [1976, p. 200]).

Numerical Implementation

The linearized intervals may be directly computed so no further discussion concerning numerical implementation of them is needed. The exact and likelihood intervals are nonlinear, and convergence of the iterative procedure is not always obtained with (41) and (42). In particular, the matrix $\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0$ sometimes became ill-conditioned and even, on occasion, singular for the test problem employed. Thus conditioning and oscillation-dampening procedures had to be used to aid convergence of the ill-conditioned cases. The cases that were singular had to be addressed differently, which is explained further on.

Matrix $\mathbf{X}^T\boldsymbol{\omega}\mathbf{X}$ often has entries that are of greatly differing magnitudes, which can lead to significant accumulation of round-off error during solution of (41) and (42) [*Draper and Smith*, 1981, p. 258]. To reduce this source of round-off error, (41) and (42) are scaled. The scaling is accomplished by defining scaling matrix $\mathbf{C}$ as the diagonal matrix whose diagonal elements are composed of the square roots of the inverses of the diagonal elements of $\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0$. Then, by letting

$$\boldsymbol{\delta} = \mathbf{C}^{-1}(\mathbf{b}_e - \mathbf{b}_0) \tag{50}$$

$$\mathbf{S}_0 = \mathbf{X}_0\mathbf{C} \tag{51}$$

$$\mathbf{p}_0 = \mathbf{C}\mathbf{a}_0 \tag{52}$$

(41) and (42) are modified to become

$$\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0\boldsymbol{\delta} = \mathbf{S}_0{}^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)] - \lambda\mathbf{p}_0 \tag{53}$$

$$\lambda = \pm\{[d_{1-\alpha}{}^2 - (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\mathbf{A}(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)) + (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\boldsymbol{\omega}^{1/2}\mathbf{P}_0\boldsymbol{\omega}^{1/2}(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))]/[\mathbf{p}_0{}^T(\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0)^{-1}\mathbf{p}_0]\}^{1/2} \tag{54}$$

The result of the scaling is that matrix $\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0$ has retained the symmetry of $\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0$ but has diagonal elements that are all unity.

The method employed for the ill-conditioned cases was derived by adding a term to $\mathbf{q}^T(\mathbf{b})\mathbf{A}\mathbf{q}(\mathbf{b})$ that drops out when the process converges. The effect of this term is to transform matrix $\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0$ to become $\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I}$, where k is a parameter to be determined. Thus if we define

$$\mathbf{q}(\mathbf{b}) = \begin{bmatrix} \mathbf{Y} - \mathbf{f}(\mathbf{b}) \\ \mathbf{b}_0 - \mathbf{b} \end{bmatrix} \tag{55}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^* & \mathbf{0}_{(p\times p)} \\ \mathbf{0}_{(p\times p)} & k\mathbf{C}^{-2} \end{bmatrix} \tag{56}$$

where $\mathbf{A}^*$ is either $\boldsymbol{\omega}^{1/2}[\mathbf{P}_0 - D^2(\mathbf{I} - \mathbf{P}_0)]\boldsymbol{\omega}^{1/2}$ or $\boldsymbol{\omega}$, as appropriate, then

$$\mathbf{Z}_0 = \begin{bmatrix} -\mathbf{X}_0 \\ -\mathbf{I} \end{bmatrix} \tag{57}$$

By using (55) through (57) in (30) and (31) and scaling the results, the following equations are derived to replace (53) and (54):

$$(\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I})\boldsymbol{\delta} = \mathbf{S}_0{}^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)] - \lambda\mathbf{p}_0 \tag{58}$$

$$\lambda = \pm\{[d_{1-\alpha}{}^2 - (\mathbf{Y}-\mathbf{f}(\mathbf{b}_0))^T\mathbf{A}^*(\mathbf{Y}-\mathbf{f}(\mathbf{b}_0)) + (\mathbf{Y}-\mathbf{f}(\mathbf{b}_0))^T\boldsymbol{\omega}\mathbf{S}_0(\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I})^{-1}\mathbf{S}_0{}^T\boldsymbol{\omega}(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))] \div [\mathbf{p}_0{}^T(\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I})^{-1}\mathbf{p}_0]\}^{1/2} \tag{59}$$

Depending on the values in $\mathbf{b}_0$, the numerator in (59) may become negative so that λ is imaginary. This means that

$$d_{1-\alpha}{}^2 - [\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)]^T\mathbf{A}^*[\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)] < 0 \tag{60}$$

Equation (60) implies that $S(\mathbf{b}_0)$ is much too large and should be reduced. This reduction can be accomplished by setting $\lambda = 0$ for the affected iteration.

Parameter k is similar to the Marquardt parameter [*Marquardt*, 1963]. It is computed using the philosophy and method given in the work by *Cooley and Naff* [1985, pp. 167–170]. That is, if the angle between the negative of the gradient (the right-hand side of equation (58)) and the search vector $\boldsymbol{\delta}$ is greater than a specified angle (usually around 80°–85°), then $k > 0$ is computed to make the angle smaller. Dampening of oscillations was accomplished by using the underrelaxation method given by *Cooley and Vecchia* [1987].

The final method used to control the effects of ill-conditioning was to employ the recommendations of *Stewart* [1973, pp. 225–228] concerning precision. All manipulations involving matrices $\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0$ and $\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I}$, including assembly, were performed using double precision arithmetic, whereas the remaining arithmetic operations including computing $\mathbf{X}_0$ were performed using single precision.

Note that the final algorithm is just a modified (scaled, conditioned, and dampened) Gauss-Newton algorithm in which the gradient of the sum of squares surface has been augmented by the term $\pm\lambda\mathbf{p}_0$. Thus instead of converging to a point where the gradient is zero, the procedure converges to a point where the gradient has the value $\pm\lambda\mathbf{p}_0$. For this reason, the same computer program can be used to compute both $\hat{\mathbf{b}}$ and $\mathbf{b}_e$.

Computation of Prediction Intervals

Exact and likelihood confidence regions on $(\boldsymbol{\beta}^T, \varepsilon_K)^T$ are given by (17) and (19), respectively. Boundaries of these regions can be written in the following forms. For the exact region

$$[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}^{1/2}[\mathbf{P}_0 - D^2(\mathbf{I} - \mathbf{P}_0)]\boldsymbol{\omega}^{1/2}[\mathbf{Y} - \mathbf{f}(\mathbf{b})] + \omega_K e_K{}^2 = 0 \tag{61}$$

and for the likelihood region

$$[\mathbf{Y} - \mathbf{f}(\mathbf{b})]^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b})] + \omega_K e_K{}^2 = S(\hat{\mathbf{b}})(D^2 + 1) \tag{62}$$

where

$$D^2 = \frac{p+1}{n-p}F_\alpha(p+1, n-p) \qquad m = 1$$
$$D^2 = M_\alpha(p, n, m) \qquad m > 1 \tag{63}$$

TABLE 2. Equations for Confidence and Prediction Intervals

	Exact Region	Likelihood Region	Linearized Region
Interval on β_i	(58), (59)	(58), (59)	(45)
Interval on $f_K(\boldsymbol{\beta})$	(58), (59)	(58), (59)	(49)
Interval on Y_K	(71), (72)	(71), (72)	(75)

Note that ε_K at some probability level α can be regarded as a fixed value, just like the values in $\boldsymbol{\beta}$. Hence we may consider ε_K to be a psuedoparameter as was done by *Cooley and Vecchia* [1987]. Based on this idea we augment parameter set $\mathbf{b}$ to include e_K. We must also augment the equation set (1) to include (14), which yields the following definitions:

$$\mathbf{q}(\mathbf{b}, e_K) = \begin{bmatrix} \mathbf{Y} - \mathbf{f}(\mathbf{b}) \\ e_K \end{bmatrix} \tag{64}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^* & \mathbf{0}_{(n \times 1)} \\ \mathbf{0}_{(1 \times n)} & \omega_K \end{bmatrix} \tag{65}$$

Because the prediction interval is on Y_K,

$$g(\mathbf{b}, e_K) = f_K(\mathbf{b}) + e_K = Y_K \tag{66}$$

Differentiation of $\mathbf{q}(\mathbf{b}, e_K)$ and $g(\mathbf{b}, e_K)$ to obtain $\mathbf{Z}_0$ and $\mathbf{a}_0$ must be with respect to both $\mathbf{b}$ and e_K. Thus from (64) and (66),

$$\mathbf{Z}_0 = \begin{bmatrix} -\mathbf{X}_0 & \mathbf{0}_{(n \times 1)} \\ \mathbf{0}_{(1 \times n)} & 1 \end{bmatrix} \tag{67}$$

$$\mathbf{a}_0 = \begin{bmatrix} \mathbf{X}_{0K}{}^T \\ 1 \end{bmatrix} \tag{68}$$

By using (64) through (68) in (30) and (31), we obtain

$$\begin{aligned} \mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0(\mathbf{b}_e - \mathbf{b}_0) &= \mathbf{X}_0{}^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)] - \lambda\mathbf{X}_{0K}{}^T \\ \omega_K e_K{}^e &= -\lambda \end{aligned} \tag{69}$$

$$\begin{aligned} \lambda = \pm\{&[d_{1-\alpha}{}^2 - (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\mathbf{A}^*(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)) \\ &+ (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\boldsymbol{\omega}^{1/2}\mathbf{P}_0\boldsymbol{\omega}^{1/2} \\ &\cdot (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))]/[\mathbf{X}_{0K}(\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0)^{-1}\mathbf{X}_{0K}{}^T + \omega_K{}^{-1}]\}^{1/2} \end{aligned} \tag{70}$$

where $e_K{}^e$ is an extreme value of e_K corresponding to $\mathbf{b}_e$. Scaling and conditioning may be applied in the present case in an analogous fashion to that given previously. These procedures modify (69) and (70) to become

$$\begin{aligned} (\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I})\boldsymbol{\delta} &= \mathbf{S}_0{}^T\boldsymbol{\omega}[\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)] - \lambda\mathbf{S}_{0K}{}^T \\ \omega_K e_K{}^e &= -\lambda \end{aligned} \tag{71}$$

$$\begin{aligned} \lambda = \pm\{&[d_{1-\alpha}{}^2 - (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\mathbf{A}^*(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0)) \\ &+ (\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))^T\boldsymbol{\omega}\mathbf{S}_0\,(\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I})^{-1}\mathbf{S}_0{}^T\boldsymbol{\omega}(\mathbf{Y} - \mathbf{f}(\mathbf{b}_0))] \\ &\div [\mathbf{p}_0{}^T(\mathbf{S}_0{}^T\boldsymbol{\omega}\mathbf{S}_0 + k\mathbf{I})^{-1}\mathbf{p}_0 + \omega_K{}^{-1}]\}^{1/2} \end{aligned} \tag{72}$$

Note that the only difference in form between (59) and (72) is the addition of $\omega_K{}^{-1}$ to the denominator of (72). Also, note that the second of (71) has to be solved for $e_K{}^e$ only after convergence of the iterative procedure. The extreme value of Y_K, and $Y_K{}^e$, is then obtained as

$$Y_K{}^e = f_K(\mathbf{b}_e) + e_K{}^e \tag{73}$$

The boundary of the linearized region can be written using (62) as

$$(\mathbf{b} - \hat{\mathbf{b}})^T\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}(\mathbf{b} - \hat{\mathbf{b}}) + \omega_K e_K{}^2 = S(\hat{\mathbf{b}})D^2 \tag{74}$$

because $S(\mathbf{b}) - S(\hat{\mathbf{b}}) = (\mathbf{b} - \hat{\mathbf{b}})^T\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}}(\mathbf{b} - \hat{\mathbf{b}})$. By following the same procedure as used to obtain (49), we obtain

$$Y_K{}^e \simeq f_K(\hat{\mathbf{b}}) \pm d_{1-\alpha}[\hat{\mathbf{X}}_K(\hat{\mathbf{X}}^T\boldsymbol{\omega}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}_K{}^T + \omega_K{}^{-1}]^{1/2} \tag{75}$$

Note that (75) has the form of the intervals given by *Lieberman* [1961].

Summary

Table 2 gives the equations to solve to obtain the various confidence and prediction intervals. In the equations cited in this table, $\mathbf{p}_0$ is given by (52), and the appropriate values for $\mathbf{a}_0$ in (52) are given in Table 1. Also, $\mathbf{A}^*$ is given by $\boldsymbol{\omega}^{1/2}[\mathbf{P}_0 - D^2(\mathbf{I} - \mathbf{P}_0)]\boldsymbol{\omega}^{1/2}$ for the exact intervals and by $\boldsymbol{\omega}$ for the likelihood intervals. Finally, $d_{1-\alpha}{}^2$ is given in Table 1, and D^2 is defined by (34) for the confidence intervals and by (63) for the prediction intervals.

SINGULAR CASES

On several occasions involving computation of exact and likelihood intervals for the test problem used here, solutions for the extreme sets of parameters $\mathbf{b}_e$ became singular. That is, although $\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0$ was full rank, p, in the vicinity of $\hat{\mathbf{b}}$, as the iterations progressed, the matrix became progressively more poorly conditioned until at some point it became for all practical purposes, singular. A special technique had to be derived to address these cases.

If the rank of $\mathbf{X}_0{}^T\boldsymbol{\omega}\mathbf{X}_0$ is $p - l$, where $l > 0$, then there is no unique solution for $\mathbf{b}_e$. In this case, the solution is unique for only $p - l$ linearly independent combinations of the p original parameters. A particular solution may be obtained by holding l of the p parameters fixed, then solving for the remaining $p - l$ parameters. In this way a particular exact or likelihood interval can be obtained. By changing the values of the parameters held fixed during solution, a new particular exact interval can be found. This idea may be used as a basis for a method of determining whether or not a problem is truly

TABLE 3. Numerical Information for Example Problem

Quantity	Value
Aquifer transmissivity, T	864 ft^2/day
Aquifer storage coefficient, S	0.0001.
Aquifer leakance, $R = K'/b'$	0.000002 day^{-1}
Discharge for pumping period 1, Q_1, for $0 < t \leq 90$ days	19,008 ft^3/day
Discharge for pumping period 2, Q_2, for $90 < t \leq 150$ days	0
Distances from pumping well, r	100 and 600 ft
Observation times, t	0.5, 5, 10, 30, 60, 90, 90.5, 95, 100, 120, and 150 days
Observed drawdowns, Y, at $r = 100$ ft	11.32, 16.85, 17.60 16.68, 20.06, 18.10, 6.03, 2.16, 1.21, 0.04, and −0.25 ft
Observed drawdowns, Y, at $r = 600$ ft	4.88, 8.14, 11.67, 12.54 12.69, 11.53, 7.47, 2.44, 2.28, 0.87, and 0.38 ft

One foot = 30.48 cm; 1 ft^2 = 0.093 m^2; 1 ft^3 = 0.028 m^3.

TABLE 4. Regression Estimates and Confidence Intervals on Parameters

Regression Estimate	Linear Interval	Likelihood Interval	Exact Interval
		Full Data Set	
T, 799.50 ft^2/d	(639.89, 959.11)	(664.64, 986.91)	(667.73, 982.38)
S, 0.15184×10^{-3}	$(-0.17122 \times 10^{-4}, 0.32080 \times 10^{-3})$	$(0.39939 \times 10^{-4}, 0.39012 \times 10^{-3})$	$(0.40942 \times 10^{-4}, 0.38349 \times 10^{-3})$
R, 0.47535×10^{-5} d^{-1}	$(-0.27668 \times 10^{-5}, 0.12274 \times 10^{-4})$	$(0.71146 \times 10^{-6}, 0.17736 \times 10^{-4})$	$(0.75205 \times 10^{-6}, 0.17136 \times 10^{-4})$
		Reduced Data Set	
T, 776.38 ft^2/d	(580.60, 972.16)	(616.37, 1007.9)	(625.98, 995.69)
S, 0.14975×10^{-3}	$(-0.47128 \times 10^{-4}, 0.34663 \times 10^{-3})$	$(0.29662 \times 10^{-4}, 0.44033 \times 10^{-3})$	$(0.32057 \times 10^{-4}, 0.42118 \times 10^{-3})$
R, 0.42010×10^{-5} d^{-1}	$(-0.45252 \times 10^{-5}, 0.12927 \times 10^{-4})$	$(0.30538 \times 10^{-6}, 0.21895 \times 10^{-4})$	$(0.35633 \times 10^{-6}, 0.19533 \times 10^{-4})$
		Time Drawdown Data Set	
T, 692.54 ft^2/d	(172.14, 1212.9)	$(<0.83400 \times 10^{-4}$,* 1539.7)	$(<0.77192 \times 10^{-2}$,* 1661.3)
S, 0.32920×10^{-3}	$(-0.10282 \times 10^{-2}, 0.16866 \times 10^{-2})$	$(0.13281 \times 10^{-6}, 0.62542 \times 10^{-2})$	$(0.47132 \times 10^{-7}, \sim 0.71 \times 10^{-2}$†)
R, 0.19291×10^{-4} d^{-1}	$(-0.91667 \times 10^{-4}, 0.13025 \times 10^{-3})$	$(0.53841 \times 10^{-9}, \sim 0.86 \times 10^{-2}$*)	$(<0$,† $\sim 0.89 \times 10^{-2}$*)

One ft^2 = 0.093 m^3.
*Became singular for $p = 3$.
†Failed to converge in 50 iterations.

singular or is only poorly conditioned, and, if the latter is true, of approximating the unique set $\mathbf{b}_e$. The method is to change the parameters being held fixed during a solution so as to increase the size of the interval from solution to solution. If a unique grand set $\mathbf{b}_e$ is eventually obtained, then the problem is only poorly conditioned, but if it becomes apparent that the interval is increasing without bound (or, at least without a practical bound) then the problem is considered to be singular. As a practical matter, if l were greater than one or two, this method would be quite tedious. In the present case $l = 1$ and the method was applied without difficulty.

For each particular solution, the computational techniques were unchanged from the standard case. The degrees of freedom for F for the likelihood region remain $(p, n - p)$ because all p parameters can be considered to be variables in $S(\mathbf{b})$. It does not matter how $\mathbf{b}_e$ is determined as long as the set satis-

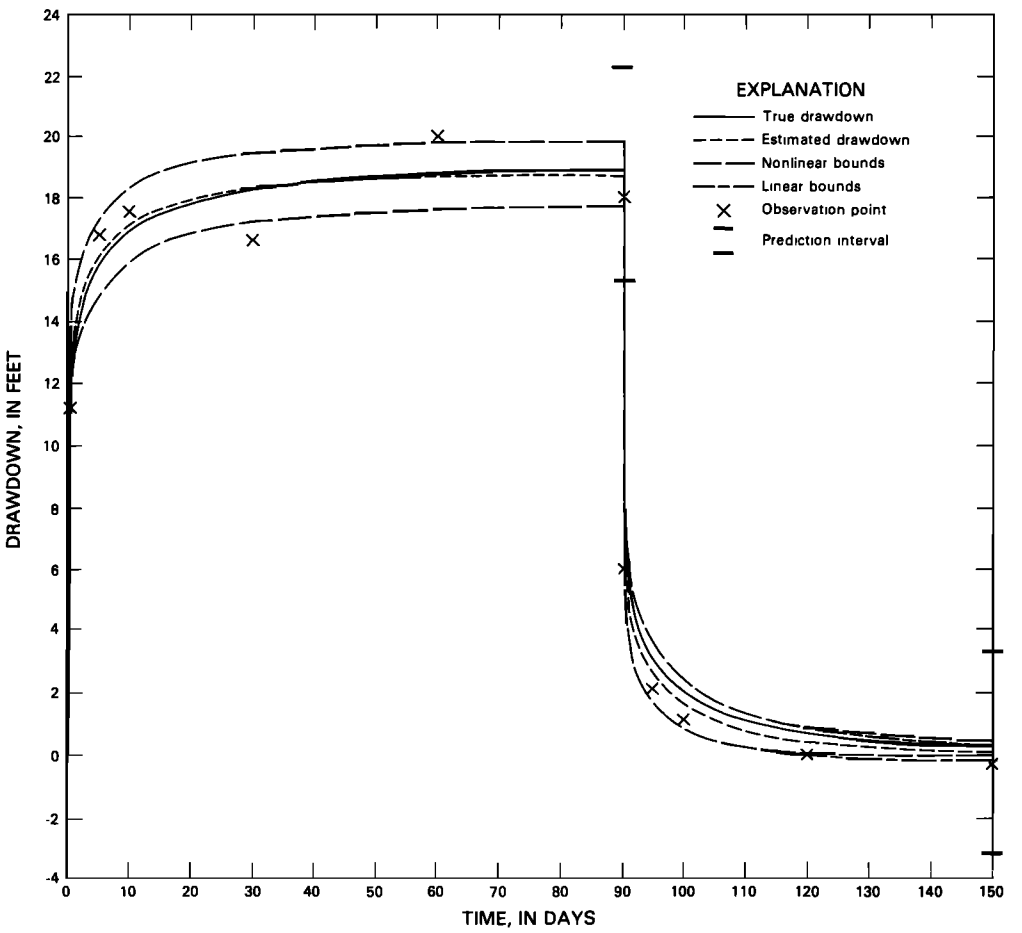

Fig. 1. Time drawdown plot using the full data set for $r = 100$ ft (1 ft = 30.48 cm). Observations are indicated by an ex, and prediction intervals are shown at $t = 90$ and $t = 150$ days.

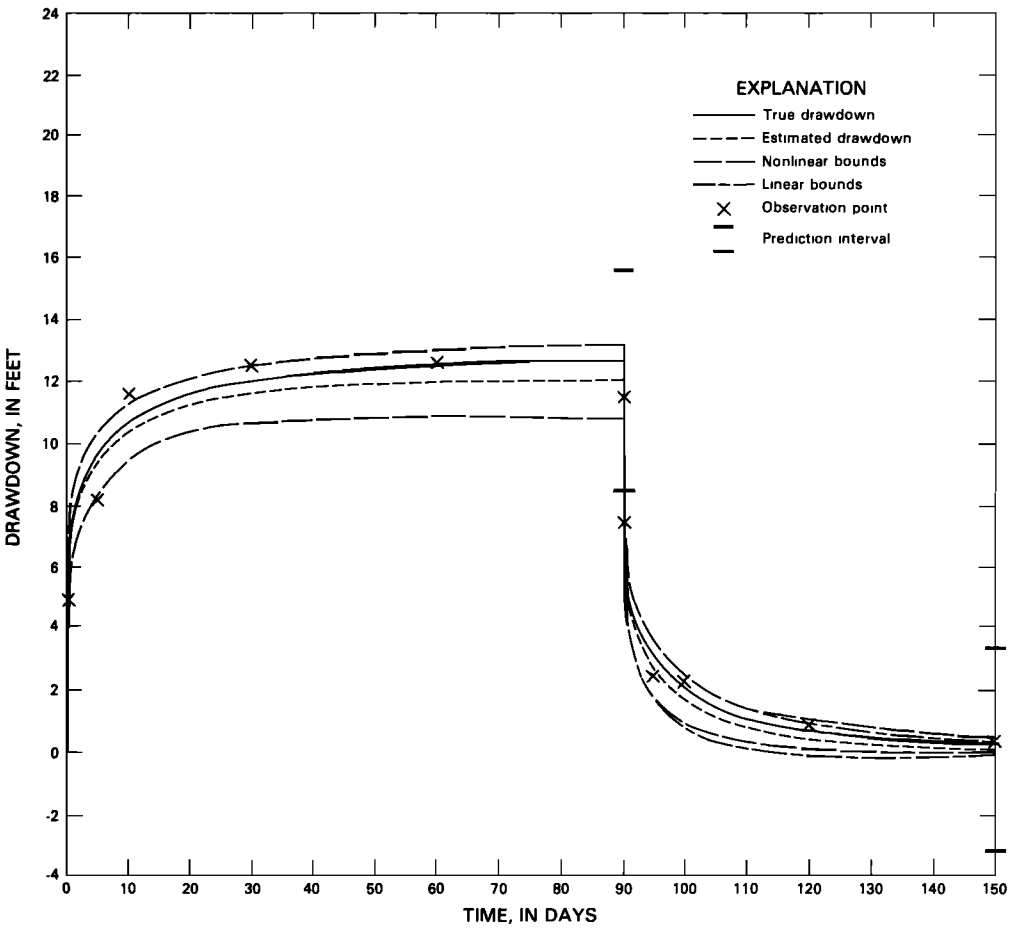

Fig. 2. Time drawdown plot using the full data set for $r = 600$ ft (1 ft = 30.48 cm). Observations are indicated by an ex, and prediction intervals are shown at $t = 90$ and $t = 150$ days.

fies (37) (or equation (62)) and is the required extreme (if one exists). However, the degrees of freedom for the exact region are based on the ranks of the quadratic forms (5) and (6) (or their counterparts for prediction intervals). If the solution is truly singular, then the rank and order of $\mathbf{X}_0{}^T\omega\mathbf{X}_0$ for the particular solution is $p - l$, and the degrees of freedom for F must be reduced to $(p - l, n - p + l)$. However, in this case the size of the interval is not unique so that the degrees of freedom used make little difference.

EXAMPLE PROBLEM

For an example of finding confidence intervals on output from nonlinear regression groundwater flow models, we will consider a hypothetical aquifer and aquitard system where the aquifer and aquitard are both homogeneous, of constant thickness, and infinite in areal extent, and where the specific storage of the aquitard is negligible. A single pumping well penetrates the aquifer, and the total simulation time was divided into two pumping periods. There are two observation wells, and 11 paired observations in time were taken at the wells. Numerical information for the problem is given in Table 3. Errors ε in observed drawdowns were generated to be $N(0, 1\ \text{ft}^2)$ ($1\ \text{ft}^2 = 0.093\ \text{m}^2$) random variables. Temporal correlations in the errors could be incorporated in the analysis through the weight matrix ω, but to simplify the example problem we assume that the errors are uncorrelated.

The analytical solution of *Hantush and Jacob* [1955], generalized to apply for more than one pumping period, forms the groundwater flow model from which drawdown, $s \equiv f(\boldsymbol{\beta})$, and sensitivities were calculated. These were used in the appropriate equations given in Table 2 to obtain the desired confidence intervals. Methods used to obtain the drawdowns and sensitivities for this leaky aquifer problem are given in the appendix.

Based on the numerical information given in Table 3, three cases were developed: (1) the full data set for which $n = 22$ and $p = 3$; (2) reduced data set constructed by using paired observations at $t = 0.5$, 10, 60, 90.5, 100, 150 days for $r = 100$ ft and $r = 600$ ft, for which $n = 12$ and $p = 3$; and (3) time drawdown data set constructed by using all time observations at $r = 100$ ft, for which $n = 11$ and $p = 3$. Cases 2 and 3 allowed determination of the effects of the two subsampling strategies on the confidence intervals. All confidence intervals are based on $\alpha = 0.05$; that is, they are 95% intervals.

Regression estimates and confidence intervals on parameters for the three cases are given in Table 4. Note that all corresponding intervals increase in size from cases 1–3. This appears to result because of the decreasing number of observations from cases 1–3 and because time drawdown data alone do not define the parameters as well as time and distance drawdown data used together. Linear intervals differ from the likelihood and exact intervals for all three cases, and the lower linear bounds on S and R become negative for all three cases. Corresponding likelihood and exact intervals are similar for the first two cases, but differ for the third case.

Wide intervals and highly nonlinear behavior characterized

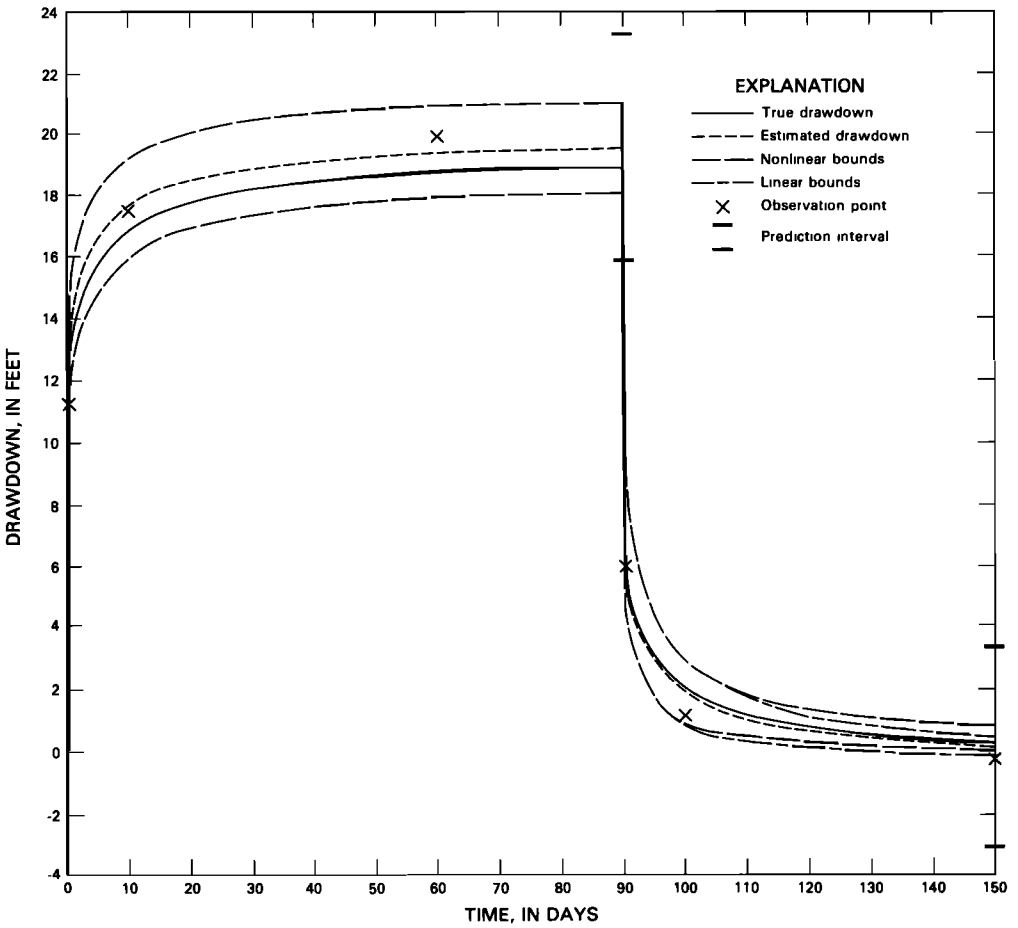

Fig. 3. Time drawdown plot using the reduced data set for $r = 100$ ft (1 ft = 30.48 cm). Observations are indicated by an ex, and prediction intervals are shown at $t = 90$ and $t = 150$ days.

the third case. The upper exact bound on S oscillated about the value of $\sim 0.71 \times 10^{-2}$, and the solution failed to converge in 50 iterations. Because the computed value of F for the last few iterations oscillated about the correct critical value, we assumed that this approximate bound is close to the correct bound. The solution also failed to converge to the lower exact bound on R because the bound is apparently less than zero, which is outside of the range of validity of the analytical solution. The solution procedure continually drove R closer to zero, but the computed value of F never became large enough to indicate that the bound had been reached. Singularity of the solution became a problem when trying to determine the lower likelihood and exact bounds on T and the upper likelihood and exact bounds on R. In both instances the singularity was caused by the production, by the search procedure, of parameter sets for which the computed drawdowns tended very quickly to steady state. In this case the storage coefficient had negligible influence on the solution, and the problem became deficient in rank by one. Therefore bounds were approximated by using the search procedure described previously. For T, for all practical purposes the bounds approach zero, and the search process was terminated at the values given in Table 4. For R, the approximate values shown in Table 4 were obtained.

Time drawdown plots for the three cases are illustrated in Figures 1 through 5. In spite of any differences in bounds of exact and likelihood confidence intervals on parameters, the exact and likelihood bounds on drawdown are virtually identical in all three cases. Furthermore, for cases 1 and 2 the linear bounds are almost identical to the nonlinear (exact and likelihood) bounds except for times after 115 days for $r = 100$ ft and after 95 days for $r = 600$ ft.

The effects of nonlinearity are more pronounced for case 3, but even in this case they are significant only from 0 to 10 days and after 95 days, where the nonlinearity is especially significant. After 95 days, the lower linear bound is negative, and after 120 days the linear interval does not even contain the true drawdown. After 95 days the solutions for the lower nonlinear bounds become singular. The bounds, which are virtually zero, reflect a system that tends to steady state very quickly, and, as in the instance of the singular parameter bounds, the solutions are deficient in rank by one.

Carrera and Neuman [1986] give a two-parameter numerical example of a groundwater flow model in which nonlinear effects in the likelihood function are reduced by log transforming the parameters. It is possible that a similar reparameterization of the present model, such as log transforming T, R, and, perhaps, S, might induce closer agreement of the nonlinear and linear intervals. However, we consider exploration of reparameterization methods to be beyond the scope of the present paper.

For cases 1 and 2, four simultaneous prediction intervals (at $t = 90$ days and $t = 150$ days for $r = 100$ ft and $r = 600$ ft) were constructed. To construct these intervals, critical values $M_\alpha(p, n, m)$ had to be obtained, as explained previously. In all cases 5000 Monte Carlo trials were used. For case 1, $M_{0.05}(3,$

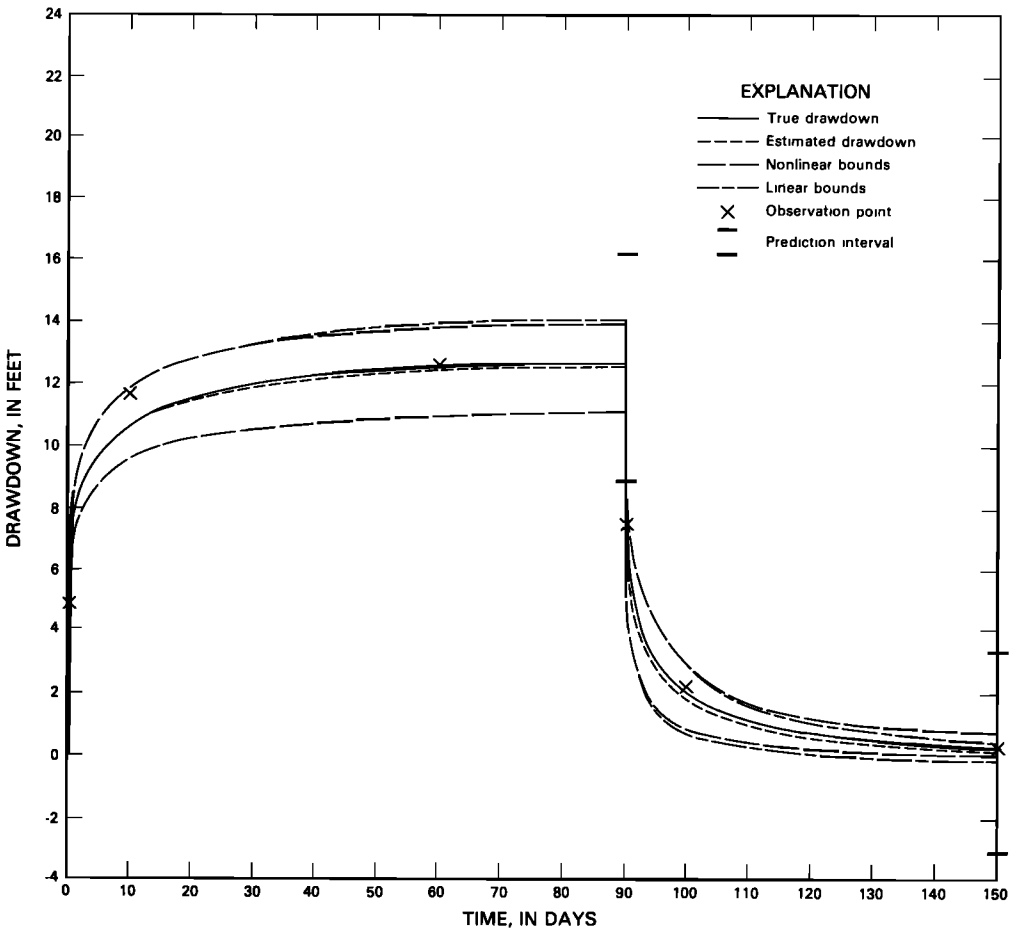

Fig. 4. Time drawdown plot using the reduced data set for $r = 600$ ft (1 ft = 30.48 cm). Observations are indicated by an ex, and prediction intervals are shown at $t = 90$ and $t = 150$ days.

22, 4) was computed to be 0.764, and for case 2, $M_{0.05}(3, 12, 4)$ was computed as 2.028. Because only one observation well was used for case 3, only two simultaneous prediction intervals at $t = 90$ days and $t = 150$ days were constructed for this case. The critical value $M_{0.05}(3, 11, 2)$ was computed as 2.175. In Figures 1 through 5 corresponding linear and nonlinear prediction intervals could not be distinguished for all three cases, and the prediction intervals are substantially larger than the confidence intervals. The negligible difference between the linear and nonlinear prediction intervals, as well as their large width, are probably due to the conservative nature of the intervals, the assumption of uncorrelated errors, and to the rather large assumed error variance.

SUMMARY AND CONCLUSIONS

1. Three types of simultaneous confidence and prediction intervals (exact, likelihood, and linearized) on output from nonlinear regression models such as groundwater flow models can be computed by using a straightforward Lagrangian optimization scheme. These three types of intervals are based on the classical exact, likelihood, and linearized confidence regions on parameters. The same Lagrangian optimization scheme can be used to obtain Scheffè-type confidence intervals on individual regression parameters based on the exact or likelihood regions.

2. The basic distributional assumption is that the errors in the regression model are distributed normally. Correlations among the errors or differences among their variances are assumed to be expressed in terms of a known weight matrix. Hence critical values for the bounds on the confidence intervals are obtained from the standard F distribution. Critical values for simultaneous prediction intervals can be obtained through use of a simple, efficient Monte Carlo scheme.

3. Application of the method to a hypothetical example problem based on the *Hantush and Jacob* [1955] solution for flow to a well penetrating a leaky aquifer showed that sampling a full data set (consisting of 22 paired observations made at 11 time points and 2 distances from the pumped well) by eliminating every other time observation had little effect on either confidence intervals on parameters or confidence and prediction intervals on drawdown. For both cases (the full and reduced data sets) the linear intervals on parameters were somewhat different from the exact and likelihood intervals, which were very similar, with the biggest difference being that the lower linear bounds for the storage coefficient and leakance were negative. The linear confidence intervals on drawdown were generally different from the nonlinear confidence intervals only during the latter parts of the simulations after the well discharge ceased; corresponding exact, likelihood, and linear prediction intervals were virtually identical. Results of a third case, obtained by eliminating one of the observation wells to produce time drawdown data only, differed greatly from results of the first two cases. Confidence intervals on parameters were very large, and linear and nonlinear intervals did not correspond well. Also, solutions to obtain the bounds became singular in several instances. In these instances, ap-

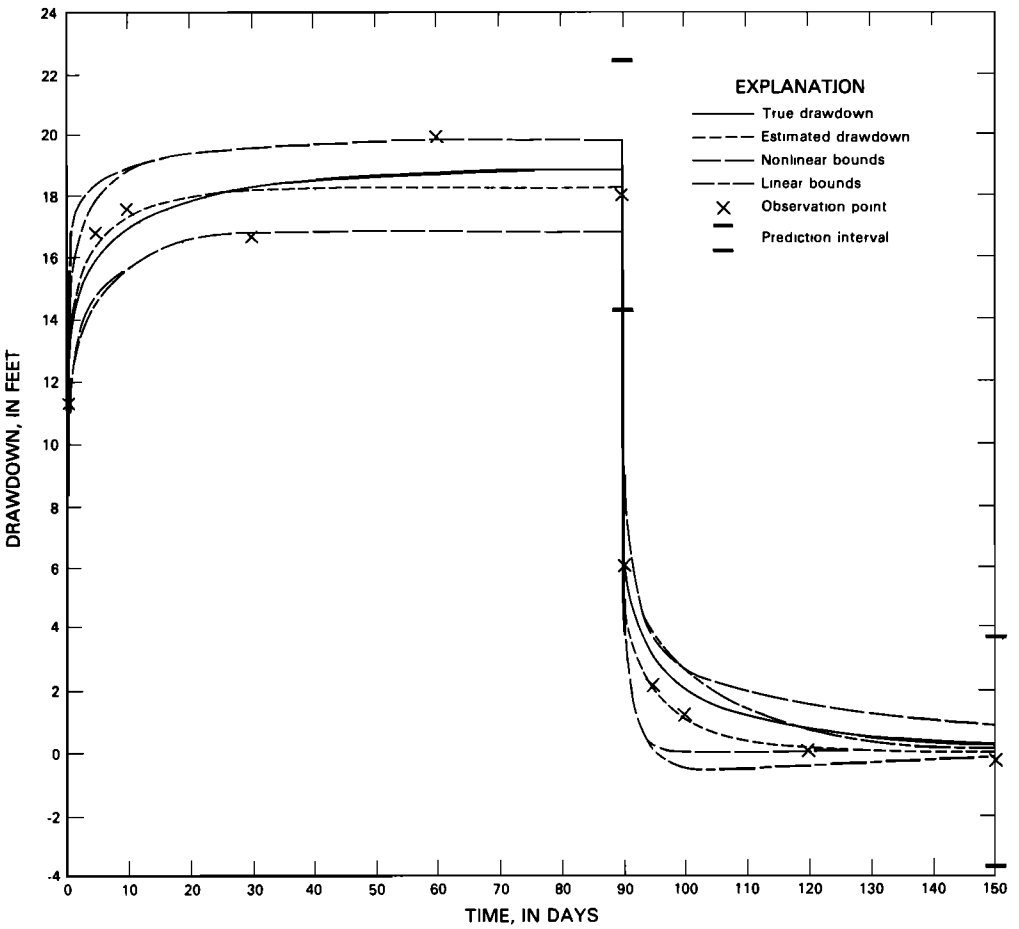

Fig. 5. Time drawdown plot using the time drawdown data set for $r = 100$ ft (1 ft = 30.48 cm). Observations are indicated by an ex, and prediction intervals are shown at $t = 90$ and $t = 150$ days.

proximate bounds were obtained with use of a trial and error search technique based on a reduced parameter set. Linear and nonlinear confidence intervals on drawdown corresponded more closely, and the effects of nonlinearity were significant only during the latter part of the simulation after the well discharge ceased. During this time, the linear intervals were a poor approximation of the nonlinear intervals. Linear and nonlinear prediction intervals on drawdown for the third case were again virtually identical.

APPENDIX

Hantush and Jacob [1955] gave the following equivalent, alternative solutions for flow to a well pumping at a constant rate in an infinite, homogeneous, leaky aquifer:

$$s = \frac{Q}{4\pi T}\int_u^\infty e^{-v^2/y-y}\frac{dy}{y} \tag{A1}$$

$$s = \frac{Q}{4\pi T}\left[2K_0(2v) - \int_w^\infty e^{-v^2/y-y}\frac{dy}{y}\right] \tag{A2}$$

where

$$u = \frac{r^2S}{4Tt} \tag{A3}$$

$$v = \left(\frac{r^2K'}{4Tb'}\right)^{1/2} \tag{A4}$$

$$w = \frac{K't}{Sb'} \tag{A5}$$

s is drawdown (L); r is distance from the pumped well (L); t is time since pumping commenced (T); Q is volumetric pumping rate (L^3T^{-1}); T is transmissivity (L^2T^{-1}); S is the storage coefficient; K' is the vertical hydraulic conductivity of the aquitard (LT^{-1}); b' is the thickness of the aquitard (L); and $K_0(\ )$ is the modified Bessel function of the second kind and zero order. *Hantush and Jacob* [1955, p. 97] also gave a series expansion of the integrals in (A1) and (A2), which may be written in the form

$$\int_x^\infty e^{-v^2/y-y}\frac{dy}{y} = e^{-x}\sum_{n=1}^{\infty}\frac{v^{2n}}{(n!)^2}\sum_{m=1}^{n}\frac{(-1)^m(m-1)!}{x^m} + W(x)I_0(2v) \tag{A6}$$

where

$$W(x) = \int_x^\infty e^{-y}\frac{dy}{y} \tag{A7}$$

and $I_0(\ )$ is the modified Bessel function of the first kind and zero order. Substitution of (A6) into (A1) and (A2) gives two alternative solutions, one efficient for early time ($w < u$) and one efficient for late time ($u \leq w$). For $w < u$,

$$s = \frac{Q}{4\pi T}\left[e^{-u}\sum_{n=1}^{\infty}\frac{v^{2n}}{(n!)^2}\sum_{m=1}^{n}\frac{(-1)^m(m-1)!}{u^m} + W(u)I_0(2v)\right] \tag{A8}$$

For $u \le w$,

$$s = \frac{Q}{4\pi T}\left[2K_0(2v) - e^{-w} \cdot \sum_{n=1}^{\infty} \frac{v^{2n}}{(n!)^2} \sum_{m=1}^{n} \frac{(-1)^m(m-1)!}{w^m} - W(w)I_0(2v)\right] \quad \text{(A9)}$$

Sensitivities are obtained by differentiating the appropriate solutions with respect to the respective parameters. Differentiation of (A8) and (A9) with respect to T and collection of terms yields the following. For $w < u$,

$$\frac{\partial s}{\partial T} = -\frac{1}{T}\left\{s + \frac{Q}{4\pi T}\left[e^{-u}\left(\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2} \cdot \sum_{m=1}^{n} \frac{(-1)^m(m-1)!}{u^m} - e^{-w}\right) + W(u)\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2}\right]\right\} \quad \text{(A10)}$$

For $u \le w$,

$$\frac{\partial s}{\partial T} = -\frac{1}{T}\left\{s + \frac{Q}{4\pi T}\left[D(2v) - e^{-w}\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2} \cdot \sum_{m=1}^{n} \frac{(-1)^m(m-1)!}{w^m} - W(w)\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2}\right]\right\} \quad \text{(A11)}$$

where

$$D(2v) = v\frac{\partial}{\partial v}K_0(2v) \quad \text{(A12)}$$

Differentiation of (A1) with respect to S yields, for all u and w,

$$\frac{\partial s}{\partial S} = -\frac{Q}{4\pi TS}e^{-u-w} \quad \text{(A13)}$$

Finally, differentiation of (A8) and (A9) with respect to K'/b' and collection of terms gives the following. For $w < u$,

$$\frac{\partial s}{\partial R} = \frac{Q}{4\pi TR}\left[e^{-u}\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2} \sum_{m=1}^{n} \frac{(-1)^m(m-1)!}{u^m} + W(u)\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2}\right] \quad \text{(A14)}$$

For $u \le w$,

$$\frac{\partial s}{\partial R} = \frac{Q}{4\pi TR}\left[D(2v) - e^{-w}\left(\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2} \cdot \sum_{m=1}^{n} \frac{(-1)^m(m-1)!}{w^m} - e^{-u}\right) - W(w)\sum_{n=1}^{\infty} \frac{nv^{2n}}{(n!)^2}\right] \quad \text{(A15)}$$

where $R = K'/b'$.

In order to improve the efficiency and numerical performance of calculating drawdowns and sensitivities, $W(\ )$ and $K_0(\ )$ were approximated with equations given by *Gautschi and Cahill* [1965, p. 231] and *Olver* [1965, p. 379]. For $0 \le x \le 1$,

$$W(x) = -\ln(x) - 0.57721566 + 0.99999193x - 0.24991055x^2 + 0.05519988x^3 - 0.00976004x^4 + 0.00107857x^5 + \varepsilon(x) \quad \text{(A16)}$$

where $|\varepsilon(x)| < 2 \times 10^{-7}$. For $1 \le x < \infty$,

$$W(x) = \frac{e^{-x}}{x}\left[\frac{x^4 + a_1x^3 + a_2x^2 + a_3x + a_4}{x^4 + b_1x^3 + b_2x^2 + b_3x + b_4}\right] + \frac{e^{-x}}{x}\varepsilon(x) \quad \text{(A17)}$$

where $|\varepsilon(x)| < 2 \times 10^{-8}$ and

$$a_1 = 8.5733287401$$
$$a_2 = 18.0590169730$$
$$a_3 = 8.6347608925$$
$$a_4 = 0.2677737343$$
$$b_1 = 9.5733223454$$
$$b_2 = 25.6329561486$$
$$b_3 = 21.0996530827$$
$$b_4 = 3.9584969228$$

For $0 < x \le 2$,

$$K_0(x) = -\ln(x/2)I_0(x) - 0.57721566 + 0.42278420(x/2)^2 + 0.23069756(x/2)^4 + 0.03488590(x/2)^6 + 0.00262698(x/2)^8 + 0.00010750(x/2)^{10} + 0.00000740(x/2)^{12} + \varepsilon(x) \quad \text{(A18)}$$

where $|\varepsilon(x)| < 1 \times 10^{-8}$.

For $2 \le x < \infty$,

$$K_0(x) = \frac{e^{-x}}{\sqrt{x}}[1.25331414 - 0.07832358(2/x) + 0.02189568(2/x)^2 - 0.01062446(2/x)^3 + 0.00587872(2/x)^4 - 0.00251540(2/x)^5 + 0.00053208(2/x)^6] + \frac{e^{-x}}{\sqrt{x}}\varepsilon(x) \quad \text{(A19)}$$

where $|\varepsilon(x)| < 1.9 \times 10^{-7}$. The term $D(2v)$ was obtained by differentiating (A18) or (A19), depending on the value of $2v$.

Although the series solutions for drawdowns and sensitivities given above are valid for all values of v, evaluation of (A6) and the similar forms in the equations for sensitivities involves subtraction of two nearly equal large numbers for values of v greater than about 2. Hence for $v \ge 2$, (A1) and (A2) were numerically integrated with a scheme explained further on. Equation (A1) was used for $w < u$ and (A2) was used for $u \le w$. Sensitivities to T and R derived from (A1) and (A2) were also numerically integrated and are as follows.

For $w < u$,

$$\frac{\partial s}{\partial T} = -\frac{1}{T}\left\{s - \frac{Q}{4\pi T}\left[v^2\int_u^{\infty} e^{-v^2/y-y}\frac{dy}{y^2} + e^{-u-w}\right]\right\} \quad \text{(A20)}$$

For $u \le w$,

$$\frac{\partial s}{\partial T} = -\frac{1}{T}\left\{s + \frac{Q}{4\pi T}\left[D(2v) + v^2\int_w^{\infty} e^{-v^2/y-y}\frac{dy}{y^2}\right]\right\} \quad \text{(A21)}$$

For $w < u$,

$$\frac{\partial s}{\partial R} = -\frac{Q}{4\pi TR}v^2\int_u^{\infty} e^{-v^2/y-y}\frac{dy}{y^2} \quad \text{(A22)}$$

For $u \le w$,

$$\frac{\partial s}{\partial R} = \frac{Q}{4\pi TR}\left[D(2v) + v^2\int_w^{\infty} e^{-v^2/y-y}\frac{dy}{y^2} + e^{-u-w}\right] \quad \text{(A23)}$$

Numerical evaluation of the integrals in (A1), (A2), and (A20) through (A23) was obtained by using Simpson's rule [*McCracken and Dorn*, 1964, pp. 172–173] in the following form:

$$I(x) = \frac{1}{6}\left(T_0 + 4\sum_{i=1}^{J} T_{ai} + 2\sum_{i=1}^{J-1} T_{bi} + T_f\right) \tag{A24}$$

where $I(x)$ is the integral being evaluated; x is either u or w; T_0, T_{ai}, T_{bi}, and T_f represent the integrand of $I(x)$ at different positions, y; T_{ai} and T_{ai+1} (or T_{bi} and T_{bi+1}) are spaced $2\Delta y$ apart; $\Delta y = 1/2$; T_0 is located at x; T_{a1} is located at $x + \Delta y$; T_{b1} is located at $x + 2\Delta y$, and T_f is located at $x + 2J\Delta y$. The point J is determined as the point where

$$T_{aJ} \Big/ \sum_{i=1}^{J} T_{ai} < 1 \times 10^{-5} \tag{A25}$$

To allow nonconstant pumping rate with time, subdivide the total simulation time into $k = 1, 2, \cdots, P$ pumping periods, so that during pumping period t_{k-1} to t_k the well is pumping at constant rate Q_k. Then, by using the principal of superposition, write the solution for total drawdown in the form

$$s = \frac{1}{4\pi T}\sum_{k=1}^{P}(Q_k - Q_{k-1})L(x_{k-1}, v^2) \tag{A26}$$

where $L(x_k, v^2)$ is the integral in (A1) or equivalent terms in (A2), (A8), or (A9); x_k is either u_k or w_k, as appropriate;

$$u_k = \frac{r^2 S}{4T(t - t_k)} \qquad t > t_k$$
$$u_k = \infty \qquad t \le t_k \tag{A27}$$

$$w_k = \frac{K'(t - t_k)}{Sb'} \qquad t > t_k$$
$$w_k = 0 \qquad t \le t_k \tag{A28}$$

$Q_0 = 0$ and $t_0 = 0$. Sensitivities are obtained by superimposing appropriate derivatives in the same manner.

Acknowledgments. The authors thank S. Gorelick, D. Knopman, B. Troutman, and the referees for their many helpful comments.

REFERENCES

Carrera, J., Estimation of aquifer parameters under transient and steady state conditions, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1984.

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.

Cooley, R. L., and R. L. Naff, Regression modeling of groundwater flow, *U.S. Geol. Surv. Open File Rep.*, *85-180*, 450 pp., 1985.

Cooley, R. L., and A. V. Vecchia, Calculation of nonlinear confidence and prediction intervals for groundwater flow models, *Water Resour. Bull.*, in press, 1987.

Donaldson, J. R., and R. B. Schnabel, Computational experience with confidence regions and confidence intervals for nonlinear least squares, *Technometrics*, *29*, 67–82, 1987.

Draper, N., and H. Smith, *Applied Regression Analysis*, 2nd ed., 709 pp., John Wiley, New York 1981.

Gallant, A. R., Confidence regions for the parameters of a nonlinear regression model, *Mimeogr. Ser. 1077*, 31 pp., Inst. of Stat., N.C. State Univ., Raleigh, 1976.

Gautschi, W., and W. F. Cahill, Exponential integral and related functions, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, pp. 227–251, Dover, Mineola, N. Y., 1965.

Graybill, F. A., *Theory and Application of the Linear Model*, 704 pp., Duxbury Press, North Scituate, Mass., 1976.

Hahn, G. J., Simultaneous prediction intervals for a regression model, *Technometrics*, *14*, 203–214, 1972.

Halperin, M., Confidence interval estimation in nonlinear regression, *J. R. Stat. Soc., Ser. B*, *25*, 330–333, 1963.

Hantush, M. S., and C. E. Jacob, Nonsteady radial flow in an infinite leaky aquifer, *Eos Trans. AGU*, *36*(1), 95–100, 1955.

Hartley, H. O., Exact confidence regions for the parameters in nonlinear regression laws, *Biometrika*, *51*, 347–353, 1964.

Lieberman, G. J., Prediction regions for several predictions from a single regression line, *Technometrics*, *3*, 21–27, 1961.

Marquardt, D. W., An algorithm for least squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, *11*, 431–441, 1963.

McCracken, D. D., and W. S. Dorn, *Numerical Methods and Fortran Programming*, 457 pp., John Wiley, New York, 1964.

Olver, F. W. J., Bessel functions of integer order, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, pp. 355–433, Dover, Mineola, N. Y., 1965.

Rao, C. R., *Linear Statistical Inference and Its Applications*, 625 pp., John Wiley, New York, 1973.

Stewart, G. W., *Introduction to Matrix Computations*, 441 pp., Academic, Orlando, Fla., 1973.

Sundararaj, N., A method for confidence regions for nonlinear models, *Aust. J. Stat.*, *20*, 270–274, 1978.

Wallace, J. R., and J. L. Grant, A least squares method for computing statistical tolerance limits, *Water Resour. Res.*, *13*(5), 819–823, 1977.

Williams, E. J., Exact fiducial limits in nonlinear estimation, *J. R. Stat. Soc., Ser. B*, *24*, 125–139, 1962.

Yeh, W. W-G., Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*(1), 95–108, 1986.

R. L. Cooley, U.S. Geological Survey, Mail Stop 413, Federal Center, Lakewood, CO 80225.

A. V. Vecchia, Department of Mathematics, Colorado School of Mines, Golden, CO 80401.

(Received September 2, 1986;
revised March 20, 1987;
accepted March 25, 1987.)