SANDIA REPORT

SAND2005-6893
Unlimited Release
Printed October 2005

Confidence Region Estimation Techniques for Nonlinear Regression: Three Case Studies

Kay White Vugrin, Laura Painton Swiler, Randall M. Roberts, Nicholas J. Stucky–Mack, and Sean P. Sullivan

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online ordering:	http://www.ntis.gov/ordering.htm

SAND2005-6893
Unlimited Release
Printed October 2005

Confidence Region Estimation Techniques for Nonlinear Regression: Three Case Studies

Kay White Vugrin
Performance Assessment and Decision Analysis Department
Sandia National Laboratories
4100 National Parks Highway
Carlsbad, NM 88220
kwvugri@sandia.gov

Laura Painton Swiler
Optimization and Uncertainty Estimation Department
Sandia National Laboratories
P.O. Box 5800, MS 0370
Albuquerque, NM 87185-0370
lpswile@sandia.gov

Randall M. Roberts
Repository Performance Department
Sandia National Laboratories
4100 National Parks Highway
Carlsbad, NM 88220
rmrober@sandia.gov

Nicholas J. Stucky–Mack
Harvard University
University Hall
Cambridge, MA 02138
njstucky@fas.harvard.edu

Sean P. Sullivan
University of Texas
226 Blanton
2500 University Ave.
Austin, TX 78705-9012
spsully@mail.utexas.edu

Abstract

This work focuses on different methods to generate confidence regions for nonlinear parameter identification problems. Three methods for confidence region estimation are considered: a linear approximation method, an F–test method, and a Log–Likelihood method. Each of these methods are applied to three case studies. One case study is a problem with synthetic data, and the other two case studies identify hydraulic parameters in groundwater flow problems based on experimental well–test results. The confidence regions for each case study are analyzed and compared. Although the F–test and Log–Likelihood methods result in similar regions, there are differences between these regions and the regions generated by the linear approximation method for nonlinear problems. The differing results, capabilities, and drawbacks of all three methods are discussed.

Acknowledgments

The first three authors all work at Sandia National Laboratories, and the final two authors were summer interns at Sandia in 2005. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL8500. This work was supported by the Advanced Simulation and Computing (ASC) program of the National Nuclear Security Administration (NNSA), part of the Department of Energy.

The authors wish to thank Jon Helton and Bill Oberkampf for technical reviews. We also thank Scott Mitchell and Tom Kirchner for providing valuable background information.

Contents

Figures

Tables

Confidence Region Estimation Techniques for Nonlinear Regression: Three Case Studies

1 Introduction

Nonlinear models are frequently used to model physical phenomena and engineering applications. In this paper, we refer to a nonlinear model very broadly: the output of the model is a nonlinear function of the parameters (Draper and Smith, 1998). Thus, nonlinear models can include systems of partial differential equations (PDEs). Some examples include CFD (computational fluid dynamics) (Oberkampf and Barone, 2005), groundwater flow (Beauheim and Roberts, 2002), heat transport (Stanley, 2001), etc. Nonlinear models also include functional approximations of uncertain data via regression or response surface models. Many nonlinear models require the solution of some type of optimization problem to determine the optimal parameter settings for the model. Statisticians have long worried about how to determine the optimal parameters for nonlinear regression models (Seber and Wild, 2003). In the case of nonlinear regression, optimization methods have been used to determine the parameters which "best fit" the data, according to minimizing a least squares expression. The optimization methods may not always converge and find the true solution, although advances in the optimization methods have improved nonlinear least squares solvers.

In this paper, we are concerned about determining confidence intervals around the parameter values in a nonlinear model. The parameters may be parameters in an approximation model such as a regression model, or physics modeling parameters which are used in physical simulation models such as PDEs. We refer to data as separate from parameters: data are physical data which is input either to a regression or physical simulation, and parameters are variables which are used in the representation and solution of the nonlinear model. For example, in groundwater flow modeling, parameters include hydraulic conductivity, specific storage, etc. Data may included measured flow rate from well data. The focus of this paper is on calculating and evaluating joint confidence intervals. A joint confidence interval is one that simultaneously bounds the parameters; it is also called a simultaneous confidence interval. It is not a set of individual confidence intervals for each parameter in the problem. Individual confidence intervals on parameters usually assume independence of the parameters which may lead to large errors if one is trying to infer a region where the parameters may jointly exist, for example, with 95% confidence. A joint confidence interval for a problem with two parameters may be an ellipse, where all points within the ellipse represent potential combinations of the parameters that fall within the confidence region.

We are interested in calculating joint confidence intervals to understand the range or

potential spread in the parameter values. Given various sources of uncertainty, it is unlikely that the optimal parameters which minimize some least squares formulation are the only reasonable parameters. There are several sources of uncertainty that can contribute to difficulty in identifying optimal parameter values in nonlinear problems. The data itself may have significant uncertainties: there may be missing values, measurement error, systematic biases, etc. Nonlinear inverse problems may involve discontinuities which result in multiple values for the optimal parameters, due to complexities in the underlying physics (e.g., significant heterogeneities in material models). The parameters themselves may have significant variability as part of their inherent randomness. Finally, model form can also influence the parameter settings. For these reasons, one should not always trust the optimal parameter values obtained by a nonlinear least squares solutions. Looking at the joint confidence intervals on parameter values will give a more complete picture about the optimal values for the parameters, and their correlation.

The question of interest in this paper is: once the optimal parameters have been found, how does one determine the simultaneous confidence intervals around the parameters? Various statistical methods have been developed to do this for nonlinear regression problems. This paper examines three methods for determining joint confidence intervals in nonlinear models and compares them in case studies. These methods have been developed for the case where the regression model is a nonlinear model. In two of our cases, the nonlinear model is a groundwater flow and transport code; it is a set of PDEs. We look at the applicability of the three methods to PDE-type models, and discuss implementation issues. The first section of this paper provides background to nonlinear regression and determination of parameters in nonlinear models. The second section describes three methods which are used to determine joint confidence intervals for parameters in nonlinear models. The third section outlines the example problems used in the case studies, and the fourth section discusses the results, with a fifth section providing conclusions.

2 Nonlinear Regression

Nonlinear regression extends linear regression for use with a much larger and more general class of functions (Bates and Watts, 1988). Almost any function that can be written in closed form can be incorporated in a nonlinear regression model. Unlike linear regression, there are very few limitations on the way parameters can be used in the functional part of a nonlinear regression model. The way in which the unknown parameters in the function are estimated, however, is conceptually the same as it is in linear least squares regression. In nonlinear regression, the nonlinear model of the response y as a function of the n–dimensional inputs x is given as $y = f(\theta;x) + \varepsilon$, where f is the nonlinear model, θ is a vector of parameters, and ε is the random error term. As an example, one could have $y_i = \theta_1[1 - \exp(x_i\theta_2)]$. Note that for nonlinear functions, the derivative of f with respect to the parameters θ depends on at least one of the parameters of the vector θ. The goal of

nonlinear regression is to find the optimal values of θ to minimize the function

$$\sum_{i=1}^{n} [f(\theta; x_i) - y_i]^2.$$

Nonlinear regression requires an optimization algorithm to find the vector θ that minimizes the sum of squares. This is often difficult. Nonlinear least squares optimization algorithms have been designed to exploit the structure of a sum of the squares objective function. The objective function can be formulated as:

$$\text{Minimize } S(\theta) = \sum_{i=1}^{n} [r_i(\theta)]^2,$$

where $r_i(\theta)$ is the i^{th} least squares term (residual) $r_i(\theta) = f(\theta; x_i) - y_i$. If $S(\theta)$ is differentiated twice, terms of $r_i(\theta)$, $r_i''(\theta)$, and $[r_i'(\theta)]^2$ result. By assuming that the residuals $r_i(\theta)$ are close to zero near the solution, the Hessian matrix of second derivatives of $S(\theta)$ can be approximated using only first derivatives of $r_i(\theta)$.

Optimization methods are used to minimize the sum of squares term $S(\theta)$. Note that these optimization methods work both in the case where the function f is an analytical model OR in the case where f is a computational simulation model and we are trying to find the optimal value of the parameters which minimizes the differences between the model predictions and experimental data. An algorithm that is particularly well-suited to the small-residual case and the above formulation is the Gauss-Newton algorithm. This formulation and algorithm combination typically requires the user to explicitly formulate each term in the least squares (e.g., n terms for n data points) along with the gradients for each term. This may be very expensive for computationally intensive evaluations of f. The calculations necessary also will increase as the number of parameters increases. Additionally, the approximation of gradients in the presence of errors in the problem is problematic. Often, the gradient approximation has larger errors than the objective function approximation (Borggaard et al., 2002).

Because of the expense and questionable accuracy of computing gradient approximations, we choose an optimization algorithm that does not require gradients. The Shuffled Complex Evolution Method (SCEM) is a hybrid of the Nelder–Mead algorithm and evolutionary algorithms that was developed specifically for nonlinear hydraulic parameter identification problems (Duan et al., 1992, 1993). Vugrin (2005) contains a complete description of our application of the SCEM.

3 Confidence Regions

The term confidence region is often misused, but most people have the sense that they want to understand the uncertainty in a data point or in a prediction. Confidence regions are sometimes called inference regions, indicating that these are regions where one infers something about the likelihood of the parameters existing. This section will discuss two types of confidence regions: regions for the individual parameters, and simultaneous regions for the joint parameters.

3.1 Confidence Regions for Individual Parameters

One can calculate confidence regions for individual parameters using the t-test statistic. For a particular confidence level $1-\alpha$, the confidence region for an individual parameter θ_j is given by

$$\{\theta_j \pm t_{(1-\frac{\alpha}{2}),n-p}\sigma_j\}, \tag{1}$$

where σ_j is the standard deviation of parameter θ_j, n is the number of data points, and p is the number of parameters to identify (Neter et al., 1985). The term $t_{(1-\frac{\alpha}{2}),n-p}$ is "the percentage point of a t-variable with $n-p$ degrees of freedom that leaves a probability of $\alpha/2$ in the upper tail, and so $1-\alpha/2$ in the lower tail" (Draper and Smith, 1998). If one constructed confidence bounds for each parameter individually in this way, then merged them to construct a hypercube, this hypercube would be an incorrect estimate of the confidence region because it assumes that the inferences about each parameter are made independently when they are not. For example, if one constructs a 95% confidence region for θ_1 and θ_2 and defines a rectangle based on the bounds of both, this rectangle does not enclose a 95% confidence region for the joint distribution of θ_1 and θ_2. If one assumed independence in the parameters, the probability that both parameters would be inside the rectangle would be .95*.95 = .9025. However, they are not independent.

3.2 Confidence Regions for Joint Parameters

This paper explores three methods for constructing simultaneous confidence regions on parameters from nonlinear models. These methods can be found in Seber and Wild (2003). The methods are confidence regions based on linear approximation, the F-test, and the Likelihood ratio. In the various subsections below, the following notation is used:

Given nonlinear model with known functional relationship f,

$$y_i = f(\theta; x_i) + \varepsilon_i, \tag{2}$$

$E[\varepsilon_i] = 0$, $V_{\varepsilon_i} = \sigma^2$, and ε_i are identical and independently distributed (iid). The least squares estimator of the true minimum θ^* is $\hat{\theta}$, which minimizes the error sum of squares:

$$S(\theta) = \sum_{i=1}^{n} [f(\theta; x_i) - y_i]^2.$$

3.2.1 Linear Approximation Method

We implement the linear approximation method described in Rooney and Biegler (2001) and Bates and Watts (1988). The linear approximation confidence region is based on a linear approximation of f. In linear models, the sum of squares function is quadratic, and contours of constant sums of squares are ellipses or ellipsoid surfaces. In nonlinear models, if one approximates the nonlinear function with a linear Taylor series expansion about the parameter estimate $\hat{\theta}$, the sum of squares approximation is then quadratic. This results in ellipsoid contours centered at $\hat{\theta}$.

In order to implement the linear approximation method, an estimate of the Hessian matrix of $f(\theta; \mathbf{x})$ is necessary. Our estimate of the Hessian is based on the optimization termination point. The first derivative of the numerical model with respect to each parameter is approximated using the finite difference forward difference scheme. The approximated vector of first partial derivatives is denoted as $\hat{F}_\cdot = F_\cdot(\hat{\theta})$. Then, the Hessian matrix is approximated by

$$F_{\cdot\cdot}(\hat{\theta}) = \hat{F}_\cdot' \hat{F}_\cdot. \tag{3}$$

The variance of ε_i must also be approximated in order to create a linear approximation confidence region. The formula is

$$s^2 = \frac{S(\hat{\theta})}{n-p}. \tag{4}$$

The confidence region is then

$$\{\theta : (\theta - \hat{\theta})^T F_{\cdot\cdot}(\hat{\theta})(\theta - \hat{\theta}) \leq p s^2 F^{\alpha}_{p,n-p}\}, \tag{5}$$

where $F^{\alpha}_{p,n-p}$ is the upper α percentage point of the F-distribution.

For linear regression models, the linear approximation confidence regions are exact. However, for nonlinear models, this approximation may not be very accurate, especially with small data sets and/or if the model is very nonlinear with respect to one or more of the parameters. Bates and Watts (1988) state: "We hasten to warn the reader that linear approximation regions can be extremely misleading."

In addition to the concerns about nonlinearity in the model, there is another reason the linear approximation method may result in incorrect confidence regions. There are two reasons that the approximation of the Hessian may result in suspicious values. The Hessian approximation does not include any second order terms. The use of only first order terms guarantees that the Hessian approximation is positive definite and therefore nonsingular. However, the approximation may be less accurate without second order derivative terms. The residuals must be close to 0 for this approximation of the Hessian to be a good one. Also, the use of finite differences for derivative approximation in finite element models has been shown to result in very poor approximations of derivatives (Borggaard et al., 2002). Since the nonlinear regression modeling technique contains sources of errors that are largely unavoidable, residuals are often not close to 0, and finite difference derivatives should be used with caution. Clearly, errors in the Hessian approximation will impact the computed confidence region. Thus, the confidence regions computed with the linear approximation method may be compromised by highly nonlinear models and small data sets or poor Hessian approximations.

3.2.2 F-test Method

The second method we investigate is based on the F distribution. This confidence region is based on the assumption that the error terms ε_i are jointly normally distributed, or spherically normal. The confidence region is the intersection of the expectation surface (f) with a sphere centered at y (Bates and Watts, 1988). In the case of the linear model, this is an ellipse as outlined in the section above. In the nonlinear case, however, it is not. The confidence region is a set of points for which $S(\theta)$ is a constant. The formula is given by:

$$\left\{ \theta : \frac{S(\theta) - S(\hat{\theta})}{S(\hat{\theta})} \leq \frac{p}{n-p} F^{\alpha}_{p,n-p} \right\}. \tag{6}$$

Seber and Wild (2003) claim that equations (5) and (6) both yield confidence regions for θ which have the asymptotic confidence level of $100(1-\alpha)\%$. However, these regions may be different, and equation (6) is recommended. Additionally, equation (6) does not require the approximation of derivatives or Hessian matrices, so this source of error is avoided. Note that the F-test and linear approximation confidence regions are the same for linear models.

3.2.3 Log–Likelihood Method

There is a discussion in Lyman (2003) and in sections 5.3 and 5.9 in Seber and Wild (2003) about a different confidence region based on likelihood functions. In this approach, $L(\gamma)$ is the log-likelihood function for a general model with an unknown p-dimensional vector

parameter γ. In the log-likelihood method, contours of the likelihood function map out confidence regions for the parameters. A hypothesis test can be tested using the statistic LR (likelihood ratio): $LR := 2[L(\hat{\gamma}) - L(\gamma_0)]$, where $(\hat{\gamma})$ is the maximum likelihood estimate of γ which maximizes the likelihood $L(\gamma)$. The test statistic LR is approximately distributed as a χ^2_p when the null hypothesis is true. This can be used to obtain an approximate confidence interval for γ: $\{\gamma : 2[L(\hat{\gamma}) - L(\gamma) \leq \chi^2_p(\alpha)\}$. Finally, under the assumption of normally distributed residuals, a similar transformation can be used to express this confidence interval in terms of the SSE:

$$\{\theta : n[\log S(\theta) - \log S(\hat{\theta})] \leq \chi^2_p(\alpha)\}, \tag{7}$$

where $\chi^2_p(\alpha)$ is the upper α percentage point for the χ^2 distribution.

Equation (7) can include a Bartlett correction factor. We use a value of one for this investigation. See Rooney and Biegler (2001) for further discussion.

4 Test Problems

We calculate confidence regions for three test problems. The first test problem is a linear parameter identification problem with synthetic data. The other two test problems are nonlinear groundwater parameter identification problems based on experimentally collected data.

4.1 Test Problem 1: Polynomials

The first test problem is simulation of a fifth degree polynomial. The set of synthetic data points $\mathbf{d}^1$ is generated by evaluating the polynomial $p(x) = 0.15x^2 - 0.05x^5$ at 101 evenly spaced points between 0 and 1. This set of data points is perturbed by adding a noise term to each data point. The noise terms are chosen from a normal distribution with mean zero and standard deviation 0.05. Figure 1 is a plot of the original polynomial and the perturbed data set.

The noisy data set $\mathbf{d}^1$ is modeled with the fifth degree polynomial $p_s(\theta_1, \theta_2; x) = \theta_1 x^2 + \theta_2 x^5$. The optimization problem is then:

$$\text{Minimize } S(\theta_1, \theta_2) = \sum_{i=1}^{101} [p_s(\theta_1, \theta_2; x_i) - d_i^1]^2.$$

The software we use to simulate the polynomials is paCalc, a parameter identification framework developed by INTERA Engineering. The paCalc framework is used for data

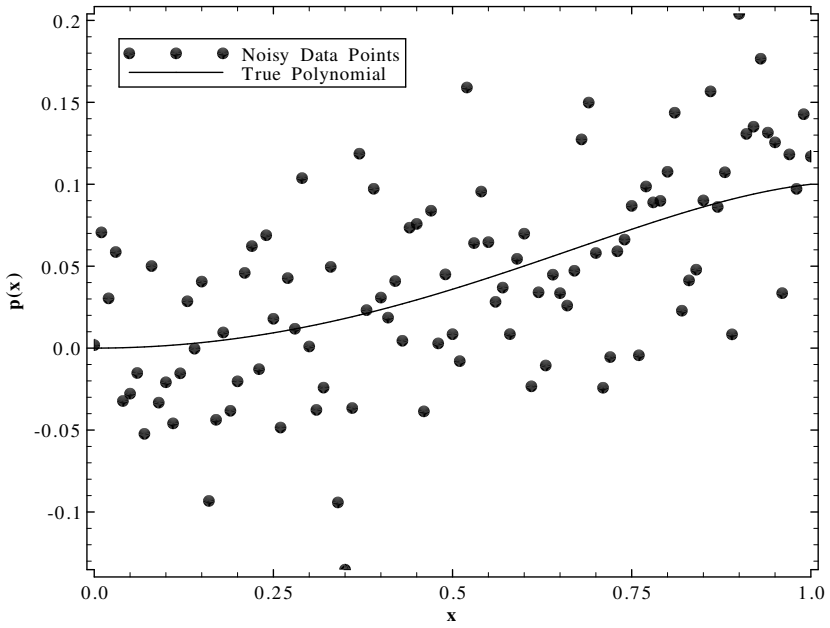

Figure 1. Test Problem 1: Underlying Polynomial and Perturbed Data Set

generation purposes, polynomial evaluations, Hessian approximation, and for application of the SCEM optimization algorithm.

Because of the construction of this test problem, we know that $\theta^* = (0.15, -0.05)$ and $S(\theta^*) = 0$ for the unperturbed problem. However, when optimization algorithms are applied to the perturbed data set, the termination point is unlikely to be equal to θ^*. In order to generate the value of $\hat{\theta}$, the SCEM is applied to the problem.

For all three test problems, we compute confidence regions at the 95% confidence level; therefore, $\alpha = 0.05$. The percentage points of the F and χ^2 distributions at $\alpha = 0.05$ are taken from Zar (1984). Table 1 is a summary of the factors that influence confidence region calculation for test problem 1. Recall that n is the number of points in the data set and p is the number of parameters. Finally, the Hessian approximation is calculated using equation (3) and s^2 is calculated using equation (4).

4.2 Groundwater Flow Model Parameters

The final two test problems are identification of hydraulic parameters for geologic formations. Well-test analysis is an inverse-modeling process in which the hydraulic parameters

Table 1. Test Problem 1: Factors Affecting Confidence Region Calculations

Variable	Value
p	2
n	101
$\hat{\theta}_1$	0.121205
$\hat{\theta}_2$	0.008131
$S(\hat{\theta})$	0.2492
s^2	0.00252
$F_{2,99}^{0.05}$	3.098
$\chi_2^2(0.05)$	5.991
Hessian Approximation:	
20.50	13.01
13.01	9.60

of an aquifer are estimated from measured transient pressure and flow rate responses to a known perturbation. During a constant-pressure (CP) test, the pressure in an isolated section (test zone) of a borehole is maintained at a nearly constant pressure that differs from the surrounding static conditions. The transient flow resulting from this imposed gradient is measured during this period. Following the CP test, the test zone is shut in and the pressure in the zone is allowed to recover to static conditions; ie. a pressure buildup (PB) test. The transient flow rate and pressure responses measured during the CP and PB tests, respectively, are then used as a data set for a regression procedure.

Values of the hydraulic parameters are estimated using the numerical well-test analysis code, nSIGHTS (n-dimensional Statistical Inverse Graphical Hydraulic Test Simulator), developed by Sandia National Laboratories. The nSIGHTS code integrates the ability to simulate the solution to the numerical model and the SCEM optimization capability. Although the numerical model is quite complicated, there are two primary governing equations. The first governing equation is the "generalized radial flow equation" from Barker (1988):

$$S_s \frac{\partial h(r,t)}{\partial t} = \frac{K}{r^{n_f-1}} \frac{\partial}{\partial r} \left(r^{n_f-1} \frac{\partial h(r,t)}{\partial r} \right), \tag{8}$$

where

S_s = specific storage (1/m),
h = hydraulic head (m),
t = time (s),
K = hydraulic conductivity (m/s),
r = radial distance from borehole (m), and
n_f = flow dimension (dimensionless).

This model is based on conservation of mass and Darcy's law (Freeze and Cherry, 1979):

$$Q(r) = -K\frac{\partial h}{\partial r}A(r), \tag{9}$$

where Q is the flow rate (m^3/s) and $A(r)$ is the flow area (m^2) at a distance r from the borehole. The relationship between hydraulic head and pressure is

$$p = \rho g h, \tag{10}$$

where p (kPa) is pressure, ρ is the density of water (kg/m^3), and g (m/sec^2) is the acceleration due to gravity.

The initial condition is

$$p(r,0) = p_0,$$

where p_0 is the static formation pressure. The boundary condition at the well is

$$p(0,t) = p^m(t),$$

where $p^m(t)$ is the measured value of the pressure at the well at time t. The boundary condition at the external boundary is

$$p(r_b,t) = p(r_b) = p_0,$$

where r_b is the radial distance (m) from the well to the boundary (Avis, 1996). A complete description of the nSIGHTS governing equations can be found in Pickens et al. (1987), which discusses the well-test analysis code GTFM (Graph Theoretic Field Model), the DOS-based precursor to nSIGHTS.

nSIGHTS applies a Graph Theoretic Field Model to generate an approximate solution the system of equations. The numerical solution is essentially a finite volume approximation to the governing equations. See Pickens et al. (1987) and Savage and Kesavan (1979) for a complete discussion of the Graph Theoretic Field Model. nSIGHTS's use of the model is discussed in Avis (1996).

nSIGHTS approximates the Hessian matrix in the same way as paCalc. The forward difference finite difference scheme is used to approximate the first derivatives of the numerical model with respect to each parameter. Then, equation (3) is applied in order to approximate the Hessian matrix.

4.3 Test Problem 2: Groundwater Parameters

Test problem 2 is based on a well–test performed as part of Ontario Power Generation's Moderately Fractured Rock (MFR) experiment. The well-test used for test problem 2 consists of a combination of a constant-pressure (CP) withdrawal test and pressure-buildup (PB) test. The MRF experiment was conducted within a 100,000 m^3 volume of fractured crystalline plutonic rock at the 240-m level of Atomic Energy of Canada's Underground Research Laboratory. A 30-minute CP test was performed in zone seven of borehole MF16 and was followed by a PB test (Roberts, 2002). We estimate hydraulic parameters by matching the measured transient flow rates $\mathbf{d}^2$ shown in Figure 2. This data set contains noise, which is especially visible after $t = 100$.

The two hydraulic parameters that we identify for test problem 2 are conductivity K and specific storage S_s. Let $y_n(K, S_s; p^m)$ be the nSIGHTS solution to the numerical model of the problem. The optimization problem is then

$$\text{Minimize } S(K, S_s) = \sum_{i=1}^{566} \left[y_n(K, S_s; p^m(t_i)) - d_i^2 \right]^2 .$$

The SCEM optimization algorithm is used to identify $\hat{K}$, $\hat{S}_s$, and $S(\hat{K}, \hat{S}_s)$. Table 2 is a summary of the values used in confidence region calculations.

4.4 Test Problem 3: Groundwater Parameters

The third test problem is based on data collected as part of the Swedish Nuclear Fuel and Waste Management Company's (SKB) Tracer Retention Understanding Experiments (TRUE) at the Äspö Hard Rock Laboratory (Winberg, 1996). Borehole KXTT2 was one of five boreholes in the TRUE-1 borehole array drilled in fractured crystalline rock. The testing sequence consisted of a 30-minute CP withdrawal test followed by a 30-minute PB

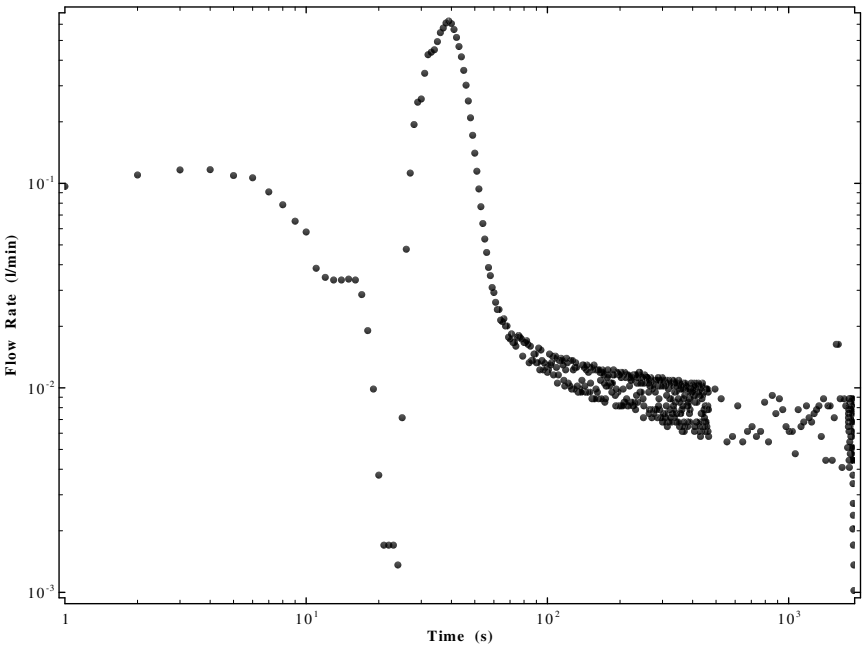

Figure 2. Test Problem 2: Matching Points

test. Hydraulic parameter estimates were obtained by matching the CP flow rates, the PB pressure data, and the derivative of the PB pressure data. Figure 3 is a plot of the measured flow rate and Figure 4 is a plot of the measured pressure and its derivative. The vector $\mathbf{d}^3$ represents the data points from all three plots.

The two hydraulic parameters that we identify for test problem 3 are again conductivity K and specific storage S_s. Let $y_n(K,S_s;t)$ be the nSIGHTS solution to the numerical model of the problem. The optimization problem is then

$$\text{Minimize } S(K,S_s) = \sum_{i=1}^{317} \left[y_n(K,S_s;t_i) - d_i^3\right]^2.$$

Again, the SCEM is used to identify $\hat{K}$, $\hat{S}_s$, and $S(\hat{K},\hat{S}_s)$. Table 3 is a summary of the values used in confidence region calculation.

Table 2. Test Problem 2: Factors Affecting Confidence Region Calculations

Variable	Value
p	2
n	566
$\hat{K}$	2.93E-12
$\hat{S}_s$	1.96E-6
$S(\hat{K},\hat{S}_s)$	17.482
s^2	0.031
$F^{0.05}_{2,564}$	3.01
$\chi^2_2(0.05)$	5.991
Hessian Approximation:	
7.52E21	8.92E15
8.92E15	1.19E10

Table 3. Test Problem 3: Values used for Confidence Region Calculations

Variable	Value
p	2
n	317
$\hat{K}$	4.09E-10
$\hat{S}_s$	1.39E-4
$S(\hat{K},\hat{S}_s)$	37.528
s^2	0.119
$F^{0.05}_{2,315}$	3.024
$\chi^2_2(0.05)$	5.991
Hessian Approximation:	
1.23E21	-8.68E14
-8.68E14	7.58E8

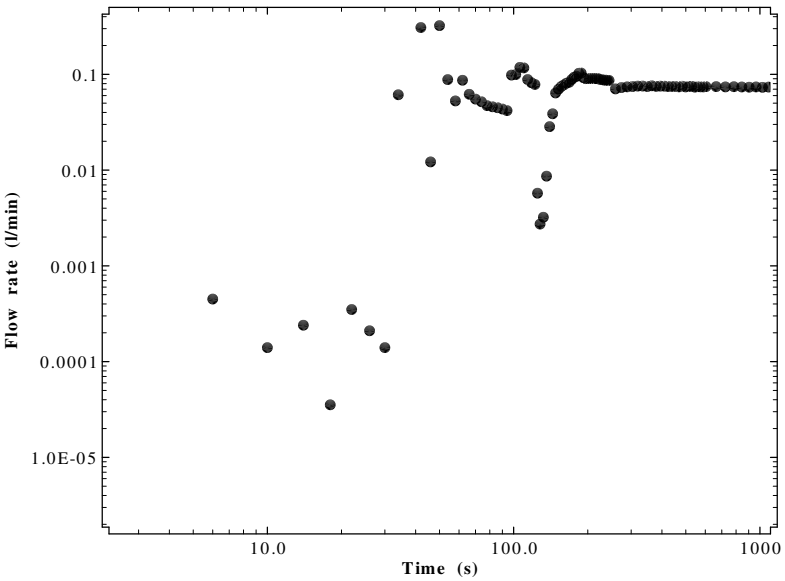

Figure 3. Test Problem 3: Matching Flow Rate Points

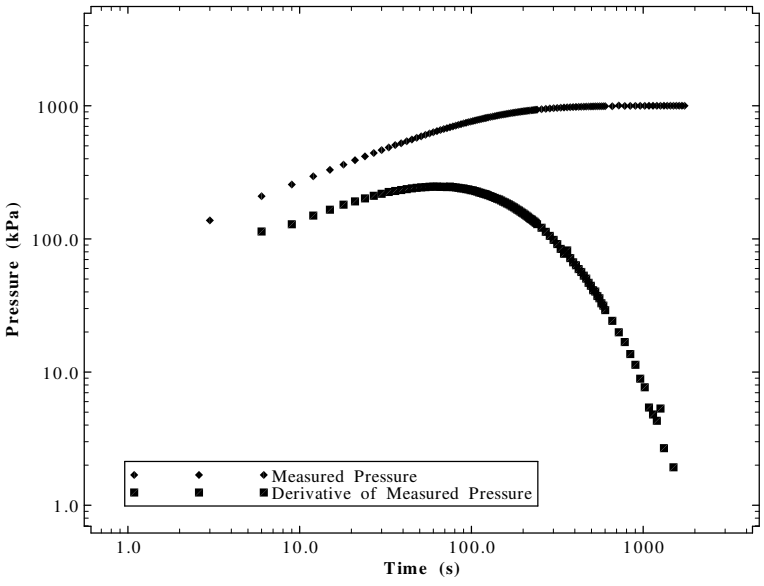

Figure 4. Test Problem 3: Matching Pressure Points

5 Results and Discussion

5.1 Visualization of Confidence Regions

For each test problem, equations (5), (6), and (7) are used to compute limits that define each type of confidence region. To be within the confidence region for linear approximation method, the parameter values must satisfy (5). To be within the confidence region for the F-test and Log–Likelihood methods, the parameter values must have a fit value lower the value identified by equations (6) and (7), respectively. Note that it is not a trivial problem to find the set of parameter values that lie exactly on the boundary of the confidence region. We did not use any computational geometry techniques, but instead enumerated various combinations of parameters until we obtained combinations that were very close to the boundary. The boundary points of the plots of each confidence region depicted in this section are all less than the identified boundary value and within 0.035% of the target boundary value. Thus, the visualized confidence regions are slightly smaller than the true regions. However, every point on the boundary and interior of the visualized confidence regions is within the calculated confidence region.

In order to identify the boundary points for the visualized confidence intervals, perturbations of $\hat{\theta}$ were evaluated until parameter values that were within 0.035% of the identified boundary were found. Once an acceptable point was found (generally by trial and error), the immediate neighborhood was searched for more points. Once a few boundary points were established, patterns arose. As more boundary points were discovered, the shape of the confidence interval could more clearly be seen by graphing, and estimations of boundary points became even easier as visual approximations of points based on the slope of the graph were possible. The tips of the confidence intervals were established especially rigorously, with many data points tested. As the number of parameters increases, the identification and visualization of the joint confidence region becomes a much harder problem.

5.2 Test Problem 1

Figure 5 is a summary of the results for confidence region calculation for test problem 1. The minimum of the unperturbed problem θ^* and the SCEM termination point $\hat{\theta}$ of the perturbed problem are both included, along with results from the three different techniques for estimating confidence regions. For this linear test problem, the three techniques produced confidence regions that are almost identical. Although it may be difficult to tell by visual inspection, the confidence region produced by the linear approximation method is slightly larger than the other two regions, and the confidence region produced by the F–test method is slightly larger than the confidence region produced by the Log–Likelihood method.

All three confidence regions contain $\hat{\theta}$, which is guaranteed by the construction methods in equations (5), (6), and (7). All three regions are also exactly ellipses, since test problem

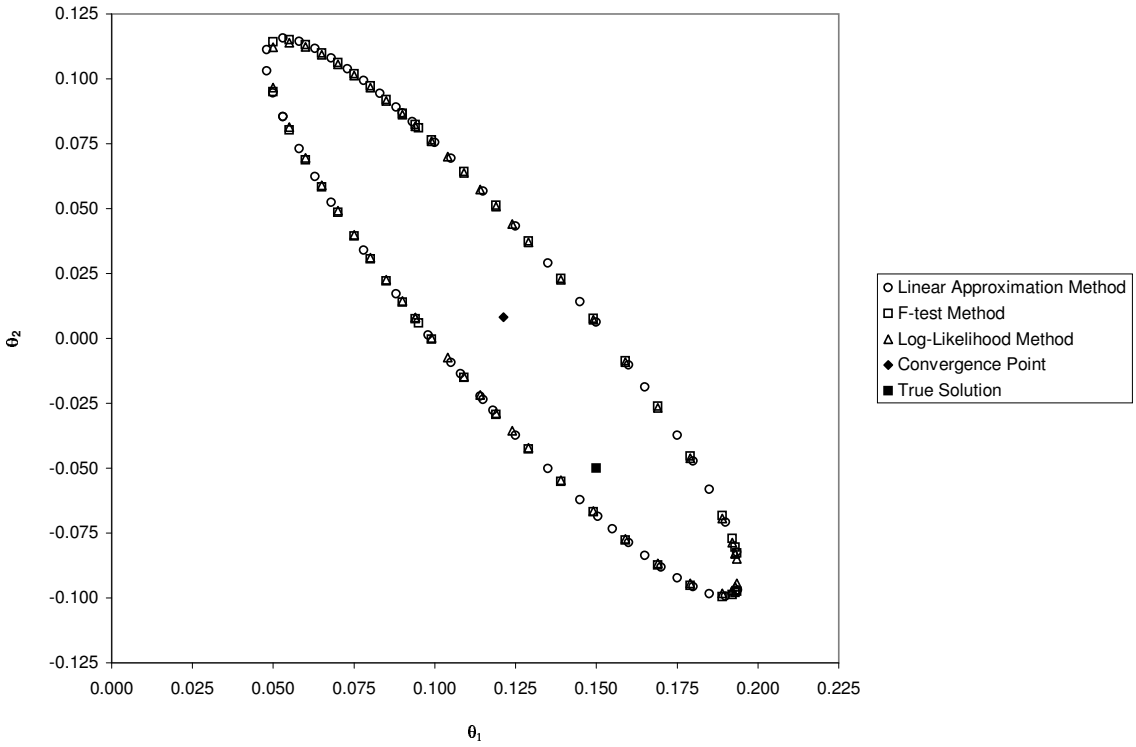

Figure 5. Test Problem 1: Confidence Regions

1 is linear. All three methods also contain θ^*, the true solution to the unperturbed problem. This is a welcome result, since confidence regions are used to identify a range of parameters that might be good solutions to the underlying (and generally unknown) unperturbed problem. Clearly, θ^* is one such parameter value.

5.3 Test Problem 2

Figures 6 and 7 are summaries of the results for confidence region calculation for test problem 2. The value for $\hat{\theta}$ is included in the plots. Figure 6 is a plot of the confidence regions produced by the F–test and Log–Likelihood methods. The two regions are almost identical, although the F–test region is slightly larger than the Log–Likelihood region. Both confidence regions are composed of positive values for each parameter. Although the true values for K^* and S_s^* are unknown, the physics of groundwater flow require that each of these parameters take on a positive value.

Figure 7 is a plot of all three confidence regions for test problem 2. The difference in scale between Figures 6 and 7 is very large. The linear approximation confidence region is so large that the other two confidence regions are not readily visible when the regions are plotted on the same scale. The linear approximation region contains negative values for both conductivity and specific storage. In fact, most of the points in the linear approximation region have negative values for conductivity or specific storage. The immense size of

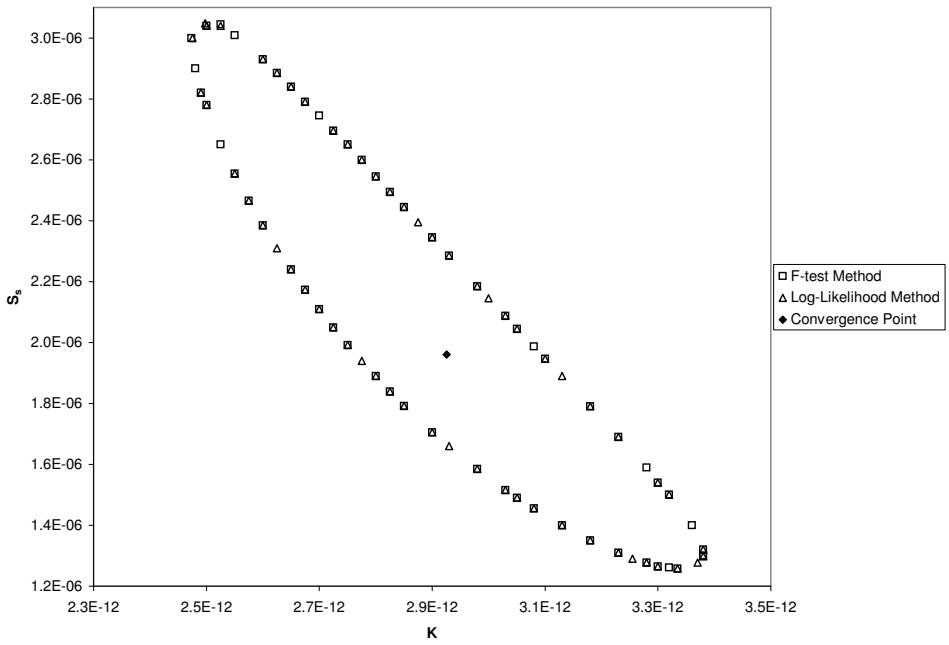

Figure 6. Test Problem 2: F–test Method and Log–Likelihood Confidence Regions

the region and the inclusion of negative parameter values strongly suggests that the linear approximation method overestimates the confidence region.

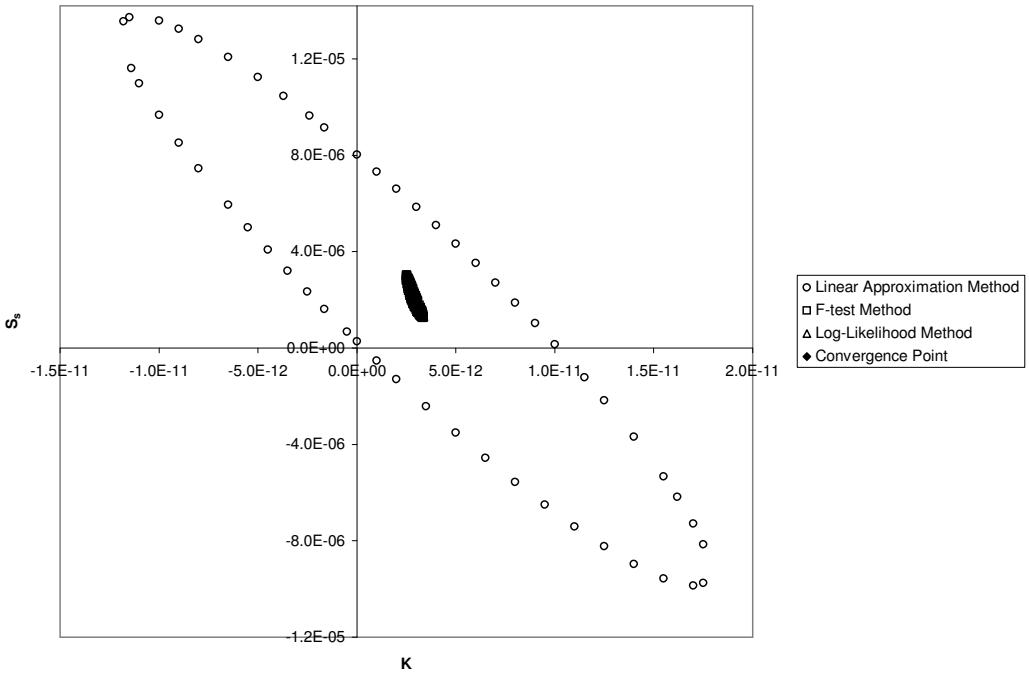

Figure 7. Test Problem 2: Confidence Regions

5.4 Test Problem 3

Figures 8 and 9 are summaries of the results for confidence region calculation for test problem 3. Again, $\hat{\theta}$ is included in the plots. Figure 8 is a plot of the confidence regions produced by the F–test and Log–Likelihood methods. Again, the two regions are almost identical, although the F–test region is slightly larger than the Log–Likelihood region. Both confidence regions are composed of positive values for each parameter. For both methods, the confidence region for test problem 3 is disjoint. The confidence regions are actually composed of the union of two separate areas. The confidence regions for these two methods are composed of parameter values with fit values below the thresholds computed in equations (6) and (7). Test problem 3 has at least two local minima in the plot of fit values. Each of these two minima are surrounded by an area of parameters with fit values below the computed thresholds. However, there are parameters in between the two local minima that have fit values that are well above the threshold values. These parameters are not included in the confidence regions, thus producing confidence regions that are composed of two disjoint pieces.

Figure 9 includes the confidence regions calculated by all three methods. The region created by the linear approximation method is not disjoint. This method is not able to create disjoint confidence regions, since members of the confidence region are included only based on their parameter values and not on their fit values. Although the linear approximation confidence region overlaps with the other two confidence regions, there are obvious differences between the linear approximation region and the other two regions. The regions are all on the same scale, and none of the regions contain negative values for either of the two parameters. The regions produced by all three methods seem reasonable.

6 Conclusions

For all three test problems, the F–test and Log–Likelihood Methods have very similar results. Although we cannot draw the conclusion that these two methods will always produce similar confidence regions, we are unable to observe differences between these two methods based on our case studies. The linear approximation method, however, produces regions that are noticeably different from the other two methods for the two nonlinear test problems. In test problem 2, the linear approximation region is very large. Since it includes negative values for the hydraulic parameters, we know that this region includes values that are physically meaningless. Thus, in this case, the region is too big.

There could be several reasons the linear approximation method fails on test problem 2 and yet produces a reasonable region on test problem 3. One difference between test problems 2 and 3 became apparent when we looked at three–dimensional plots of the objective functions versus the parameters. The objective function for test problem 2 increases

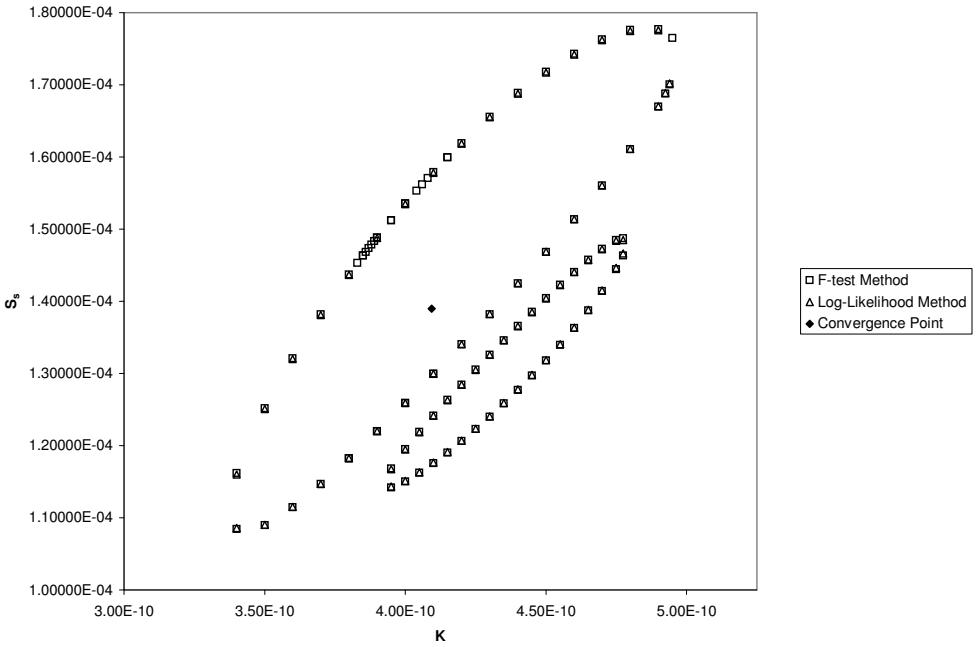

Figure 8. Test Problem 3: F–test Method and Log–Likelihood Confidence Regions

very steeply with distance from the minimum. The F-test and Log-Likelihood methods produce confidence regions that follow the contours of the objective function, and these regions are small for test problem 2. Since the linear approximation region is not based on objective function values, perhaps it is not surprising that the region computed with this method is much larger than the other two methods. The objective function for test problem 3, on the other hand, does not increase quite so steeply in the neighborhoods around the local minima. The contours of the objective function are one possible explanation for poor performance of the linear approximation region relative to the other two methods.

There are several open questions in the field of confidence region estimation. One is the determination of which confidence region method to use for general nonlinear models that are "black box" simulation models. We first thought that the linear approximation method might be sufficient for a large number of situations, but after performing the case studies, we believe the F-test and Log-likelihood methods produce more correct results because they do not rely on a calculation of the Hessian matrix, they are more reflective of the objective function, and they are able to capture disjoint confidence regions. Note that the efficacy of any of the methods is not known for very high dimensional problems. The case studies we are aware of only examine joint confidence regions for a few parameters. However, most realistic computational models involves dozens to hundreds of parameters. We easily see the potential for very ill-conditioned Hessian matrices in application of the linear approximation method. For the F-test and Log-likelihood method, the problem of finding combinations of parameters which satisfy the constraints in equations (6) and (7) could become extremely difficult: this is a one–to–many inverse problem which can be very challenging to solve. The brute-force enumeration approach that we used may not be com-

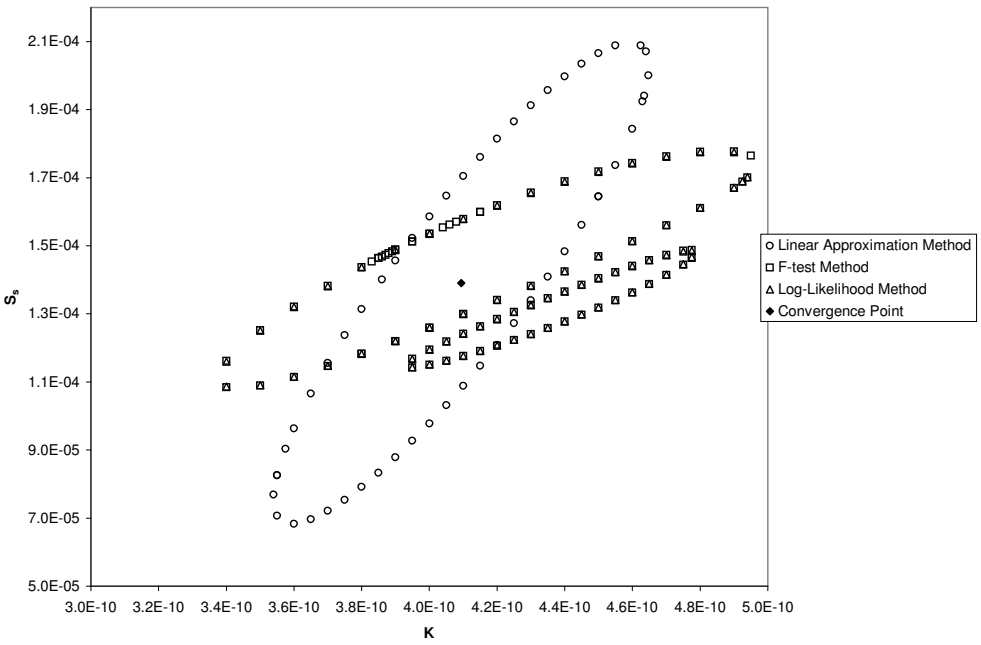

Figure 9. Test Problem 3: Confidence Regions

putationally feasible for high-dimensional problems. Another issue that complicates the visualization of confidence regions occurs when more than two parameters are identified. Although a three–dimensional graph is possible, the question of how to visualize relationships among more than three parameters is likely to arise. Although a simple calculation reveals whether a particular parameter combination is inside or outside the confidence region for all three methods, total representation of the confidence region is a significant challenge.

The need for calculating joint confidence intervals around parameters in nonlinear models is common for many scientists and engineers running simulation models. The state of the art is limited at this point. While some general approaches are available, we do not see that they can scale up to realistic problem sizes and be computationally feasible to solve for the interval boundaries. Approximation methods or better ways of "partitioning" the parameter space so that only subsets of parameters are examined simultaneously may be paths forward. One approach is to evaluate a conditional likelihood function for a pair of parameters while holding the others fixed at their least squares estimates. Another approach is to evaluate what Bates and Watts (1988) call the profile likelihood function, which involves finding the minimum sum of squares over all other parameters for pair of parameters plotted on a 2-D grid. This approach may become very expensive computationally since evaluating the profile likelihood function requires solving a $P-2$ dimensional nonlinear least squares problem for each of the points on $P(P-1)/2$ grids. Bates and Watts propose the idea of profile traces and profile pair sketches. These approaches also look at pairs of parameters conditional on the remaining parameters. However, they use some efficient interpolation schemes to plot the contours defining confidence regions for each pair of parameters. We have not seen widespread use of the profile trace or profile sketch, but they

warrant further investigation for parameter spaces of higher dimension.

References

J. Avis. GTFM–User documentation: Functional description, theoretical development, and software architecture. Sandia National Laboratories, 1996. (Copy on file in the Sandia WIPP Records Center, Carlsbad, NM as ERMS # 240244).

J. R. Barker. A generalized radial flow model for hydraulic tests in fractured rock. *Water Resour. Res.*, 24(10):1796–1804, 1988.

D. B. Bates and D. G. Watts. *Nonlinear Regression Analysis and Its Applications*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1988.

R. L. Beauheim and R. M. Roberts. Hydrology and hydraulic properties of a bedded evaporite formation. *J. Hydrol.*, 259:66–88, 2002.

J. Borggaard, D. Pelletier, and K. Vugrin. On sensitivity analysis for problems with numerical noise. In *Proc. 9th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 2002. AIAA Paper 2002-5553.

N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1998.

Q. Y. Duan, V. K. Gupta, and S. Sorooshian. Shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.*, 76(3):501–521, 1993.

Q. Y. Duan, S. Sorooshian, and V. K. Gupta. Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.*, 28(4):1015–1031, 1992.

R. A. Freeze and J. A. Cherry. *Groundwater*. Prentice–Hall, Englewood Cliffs, 1979.

G. J. Lyman. Rock fracture mean trace length estimation and confidence interval calculation using maximum likelihood methods. *Int. J. Rock Mech. Min. Sci.*, 40:825–832, 2003.

J. Neter, W. Wasserman, and M. H. Kutner. *Applied Linear Statistical Models*. Irwin, Homewood, IL, 1985.

W. L. Oberkampf and M. F. Barone. Measures of agreement between computation and experiment: Validation metrics. Technical Report 2005-4302, Sandia National Laboratories, 2005. SAND Report.

J. F. Pickens, G. E. Grisak, J. D. Avis, and D. W. Belanger. Analysis and interpretation of borehole hydraulic tests in deep boreholes: Principles, model development, and applications. *Water Resour. Res.*, 23(7):1341–1375, 1987.

R. Roberts. Moderately fractured rock experiment: Well test analysis using nSIGHTS. Technical Report 06819–REP-01300–10062–R00, Ontario Power Generation, Nuclear Waste Management Division, 2002. Prepared by Sandia National Laboratories.

W. C. Rooney and L. T. Biegler. Design for model parameter uncertainty using nonlinear confidence regions. *Process Systems Engineering*, 47:1794–1804, 2001.

G. J. Savage and H. K. Kesavan. The Graph–Theoretic Field Model–I. Modelling and formulations. *Journal of the Franklin Institute*, 307(2):107–147, 1979.

G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. Wiley Series In Probability and Statistics. John Wiley & Sons, Hoboken, 2003.

L. Stanley. Shape sensitivities for optimal design: A case study on the use of continuous sensitivity equation methods. In D. Gao et al., editor, *Nonsmooth/Nonconvex Mechanics*. Kluwer Academic Publishers, 2001.

K. E. Vugrin. *On the Effects of Noise on Parameter Identification Optimization Problems*. PhD thesis, Virginia Tech, Blacksburg, VA, April 2005.

A. Winberg. Tracer retention understanding experiments (TRUE): Test plan for the TRUE block scale experiment. Technical report, Swedish Nuclear Fuel and Waste Management Company, 1996. International Cooperation Report 97-02.

J. H. Zar. *Biostatistical Analysis*. Prentice–Hall, Inc., Englewood Cliffs, 2nd edition, 1984.

DISTRIBUTION:

1 MS 0370
Brian Adams, 01411

1 MS 0370
Michael Eldred, 01411

1 MS 0370
Scott Mitchell, 01411

3 MS 0370
Laura P. Swiler, 01411

1 MS 0370
Timothy G. Trucano, 01411

1 MS 0828
Kevin Dowding, 01533

1 MS 0828
Anthony Giunta, 01533

1 MS 0828
Jon Helton, 01533

1 MS 0828
Richard G. Hills, 01533

1 MS 0828
William Oberkampf, 01533

1 MS 0828
Martin Pilch, 01533

1 MS 0828
Vicente Romero, 01533

1 MS 1395
David Kessel, 6821

5 MS 1395
Kay W. Vugrin, 6821

1 MS 1395
Mario Chavez, 6820

2 MS 1395
Randall Roberts, 6822

1 MS 1395
Richard Beauheim, 6822

1 MS 1395
Eric Vugrin, 6821

1 MS 1395
Tom Kirchner, 6821

1 MS 9018
Central Technical File, 08945-1

2 MS 0899
Technical Library, 09616

2 MS 1395
WIPP Library, 6034