Nonlinear Parameter Estimation: a Case Study Comparison

The literature abounds with the application of optimization methods for estimating model parameters in equation systems. The utility of these methods is frequently demonstrated on pathological examples using simulated data generated from a known model with a random error component and a known statistical distribution. Unfortunately, parameter estimation problems encountered in practice do not have this advantage. The true model is frequently not known. In fact, one is faced with choosing among various candidate models, all of which may be wrong. Moreover, the error structure is generally unknown and must be estimated from the data. Finally, a great deal of mathematical expertise is required to transform the model and select meaningful starting guesses before parameter estimation can be successful.

In order to demonstrate the difficulties of parameter estimation in the industrial environment and the limitations of existing methods, a parameter estimation problem formulated by the Dow Chemical Company is presented and solved. This test problem consists of a stiff differential/algebraic (DAE) model that describes complex kinetics and requires the estimation of nine parameters from batch reactor data. Here the model was inadequate to describe the data, the error structure was not specified and the starting guesses led to a nontrivial optimization problem.

The Dow parameter estimation problem was distributed in 1981 to 165 researchers as a followup to the 1980 FOCAPD conference. Of those researchers, eleven agreed to apply their methodologies and expertise to this problem. However, only five acceptable solutions were finally submitted. Here we present and compare these results. Each solution was obtained using different strategies. In most cases the form of the model was also changed to accommodate the algorithms used and to ease the solution procedure. Therefore, while this case study does not present a direct numerical comparison of algorithms, it does offer guidelines and insight towards the solution of difficult parameter estimation problems.

L. T. BIEGLER and J. J. DAMIANO

Chemical Engineering Department
Carnegie-Mellon University
Pittsburgh, PA

and

G. E. BLAU

Dow Chemical Company
Midland, MI

SCOPE

Parameter estimation arises in fitting models containing several unknown parameters to experimental data through adjustment of these parameters. Model formulation is not a unique process; many different formulations may be used to fit the data and optimize model parameters. Of particular concern are formulations that are sufficiently accurate to represent physical or chemical phenomena and also are amenable to reasonably efficient computer solutions. Here we consider a parameter-estimation problem formulated by the Dow Chemical Company in order that current parameter estimation methods can be compared on an industrial kinetic parameter estimation problem with real data.

The model consists of nonlinear differential and algebraic equations and is stiff over a wide range of parameter values. Thus one must choose a reliable model solver before parameter estimation can begin. Several methods for solving stiff ODE and DAE systems are available (see, e.g., Carnahan and Wilkes, 1981). Most of those used in this case study are based on Gear's method.

After a model is proposed and solution techniques are chosen, an objective function that determines the goodness of fit must be selected. From maximum likelihood analysis, several alternative objective function forms can be chosen, depending on our assumptions of the error structure of the data. These can range from simple least-squares functions to fairly complex nonlinear functions that incorporate general unknown covariance of the measurements and heteroscedastic errors (see e.g., Bard, 1974).

The final step in parameter estimation is to optimize the model parameters and any unknown statistical parameters in the objective function. The optimization problem formed by the model and the objective function generally requires repeated and expensive solution of the ODE or DAE system. An efficient algorithm should minimize the number of model and function evaluations and still converge easily to the solution. Generally, these algorithms require gradient information and some approximation to the second derivative matrix. For simple and weighted least-squares objective functions, the Gauss-Newton and Levenberg-Marquardt algorithms provide the second derivative information quite easily. Otherwise, quasi-Newton updating algorithms such as BFGS (Gill et al., 1981) can also be applied.

Five researchers have submitted solutions to the parameter-estimation problem described below. After the strategies that were used are outlined, the solutions are compared in terms of accuracy and efficiency.

J. J. Damiano is currently with Standard Oil of Ohio, Cleveland, Ohio.

CONCLUSIONS AND SIGNIFICANCE

A parameter-estimation problem formulated by the Dow Chemical Co. to study a batch reactor system was solved by five independent researchers. After presenting the differential/algebraic equation (DAE) model and a suggested form for the objective function, six solution strategies are described and presented. All of these solutions used implicit ODE or DAE solvers; four investigators used variations of Gear's method. Except for an unsuccessful solution presented for illustration, only Gauss-Newton-type methods and successive quadratic minimization were used for optimization.

Based on the results, it is clear that the most important consideration for parameter estimation is formulation of a simple, yet realistic process model. Guidelines for this should include the following points for more efficient solution:

1. Eliminate all dependent equations in the proposed model. This leads to a model that is more efficiently handled by the ODE solver.

2. Eliminate as many unnecessary model parameters as possible. This leads to a smaller and better behaved optimization problem.

3. If the error structure is not known, use an objective function based on maximum likelihood that allows direct application of Gauss-Newton-type methods.

4. Determine an initial set of parameter estimates through some physical insight into the model parameters. This not only leads to a closer starting point but also avoids any unnecessary stiffness problems due to starting points far from the optimum.

All of the solutions required careful attention by the researchers to several computational difficulties. None of the investigators was able to solve the problem automatically in one go, regardless of the sophistication of the software. Judging from the very wide exposure the problem received, it appears that discovery and improvement of parameter-estimation algorithms remains a very necessary and fruitful area for research and development.

Finally, the solutions presented here do not offer a direct comparison among model solving and optimization algorithms, because they relied on different problem formulations. Instead, this study illustrates the importance of model formulation and insight in tackling difficult parameter-estimation problems as well as the limitations of current methods.

PROBLEM DESCRIPTION

The parameter estimation problem is based on a kinetic model of an isothermal batch reactor system. The reactions occur in an anhydrous, homogeneous, liquid phase catalyzed by a completely dissociated species. The reacting species have been disguised for proprietary reasons. The desired reaction is given by: $HA + 2BM \rightarrow AB + MBMH$ where AB is the desired product.

The reaction is initiated by adding the catalyst QM to a batch reactor containing the two miscible reactants with reactant BM in excess. The catalyst QM is initially assumed to be 100% dissociated to Q^+ and M^- ions. The following mechanism is proposed to describe the reaction:

Slow Kinetic Reactions

$$M^- + BM \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{MBM}^-$$

$$A^- + BM \overset{k_2}{\Longrightarrow} \mathrm{ABM}^-$$

$$M^- + AB \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{ABM}^-$$

Rapid Acid-Base Reactions

$$MBMH \Leftarrow K_1 \Rightarrow MBM^- + H^+$$

$$HA \Leftarrow K_2 \Rightarrow \mathrm{A}^- + H^+$$

$$HABM \Leftarrow K_3 \Rightarrow \mathrm{ABM}^- + H^+$$

In order to devise a model to account for these reactions, it is first necessary to distinguish between the overall concentration of a species and the concentration of its neutral form. Overall concentrations are defined for three components based on neutral and ionic species.

$$[MBMH] = [(MBMH)_N] + [MBM^-]$$

$$[HA] = [(HA)_N] + [A^-]$$

$$[HABM] = [(HABM)_N] + [ABM^-]$$

where [ ] denote concentration of the species in gmol/kg.

By assuming the rapid acid-base reactions are at equilibrium, the equilibrium constants K_1, K_2, K_3 can be defined as follows:

$$K_1 = \frac{[MBM^-][H^+]}{[(MBMH)_N]}$$

$$K_2 = \frac{[A^-][H^+]}{[(HA)_N]}$$

$$K_3 = \frac{[ABM^-][H^+]}{[(HABM)_N]}$$

The anionic species may then be represented by:

$$[MBM^-] = \frac{K_1[MBMH]}{(K_1 + [H^+])} \quad \text{(a)}$$

$$[A^-] = \frac{K_2[\mathrm{HA}]}{(K_2 + [H^+])} \quad \text{(b)}$$

$$[ABM^-] = \frac{K_3[HABM]}{(K_3 + [H^+])} \quad \text{(c)}$$

Material balance equations for the three reactants in the slow kinetic reactions yield:

$$\frac{d[M^-]}{dt} = -k_1[M^-][BM] + k_{-1}[MBM^-] - k_3[M^-][AB] + k_{-3}[ABM^-] \quad \text{(d)}$$

$$\frac{d[BM]}{dt} = -k_1[M^-][BM] + k_{-1}[MBM^-] - k_2[A^-][BM] \quad \text{(e)}$$

$$\frac{d[AB]}{dt} = -k_3[M^-][AB] + k_{-3}[ABM^-] \quad \text{(f)}$$

From stoichiometry, rate expressions can also be written for the total species:

$$\frac{d[MBMH]}{dt} = k_1[M^-][BM] - k_{-1}[MBM^-] \quad \text{(g)}$$

$$\frac{d[HA]}{dt} = k_2[A^-][BM] \quad \text{(h)}$$

$$\frac{d[HABM]}{dt} = k_2[A^-][BM] + k_3[M^-][AB] - k_{-3}[ABM^-] \quad \text{(i)}$$

An electroneutrality constraint gives the hydrogen ion concentration $[H^+]$ as:

$$[H^+] + [Q^+] = [M^-] + [MBM^-] + [A^-] + [ABM^-] \quad \text{(j)}$$

The measured values came from actual kinetic data obtained from three isothermal runs in a batch reactor. The temperatures were set at 40, 67, and 100°C. The data presented for parameter estimation represent a pretreatment of the actual data. Four component concentration vs. time profiles are presented, but only three species *HA*, *HABM*, and *AB* were actually measured. These three measurements were normalized so that:

$$[HA] + [HABM] + [AB] = \text{initial } [HA]$$

Normalization adjustments resulted in concentration changes of less than ten percent. A fourth species, $[BM]$, was derived from:

$$[BM] = \text{initial } [HABM] - 2[AB]$$

However, this relationship is only true if $[AB] = [MBMH]$, an unmeasured species. From the model, this relation holds at long times. In all of the runs, the initial catalyst concentration was 0.0131 gmol/kg. The data sets are given in Appendix A.

Although many data are present, they are insufficient to estimate all of the parameters present in the proposed model and a few assumptions were made. First, it was determined from other data that the equilibrium constants did not vary with temperature over the interval 40–100°C. Also, based on the similarities of the reacting species, we assume:

$$k_3 = k_1$$
$$k_{-3} = 1/2\, k_{-1}$$

Based on these last two assumptions, three rate constants, k_1, k_2, k_{-1} must be estimated. Each of these can be fitted with two adjustable model parameters, assuming an Arrhenius temperature dependence. That is:

$$k_1 = \alpha_1 \exp(-E_1/RT)$$
$$k_{-1} = \alpha_{-1} \exp(-E_{-1}/RT)$$
$$k_2 = \alpha_2 \exp(-E_2/RT)$$

where R is the gas constant, T is the reaction temperature in Kelvins, and the parameters, α, E, represent the preexponential factor and activation energy, respectively, for the appropriate rate constant.

The model can therefore be expressed mathematically as six differential equations and four algebraic equations. The letter labels for the following equations refer to the corresponding kinetic and equilibrium expressions derived above.

$$\frac{dy_1}{dt} = -k_2 y_8 y_2 \quad \text{(1), (h)}$$

$$\frac{dy_2}{dt} = -k_1 y_6 y_2 + k_{-1} y_{10} - k_2 y_8 y_2 \quad \text{(2), (e)}$$

$$\frac{dy_3}{dt} = -k_2 y_8 y_2 + k_1 y_6 y_4 - 1/2 k_{-1} y_9 \quad \text{(3), (i)}$$

$$\frac{dy_4}{dt} = -k_1 y_6 y_4 + 1/2 k_{-1} y_9 \quad \text{(4), (f)}$$

$$\frac{dy_5}{dt} = -k_1 y_6 y_2 - k_{-1} y_{10} \quad \text{(5), (g)}$$

$$\frac{dy_6}{dt} = -k_1(y_6 y_2 + y_6 y_4) + k_{-1}(y_{10} + 1/2 y_9) \quad \text{(6), (d)}$$

$$y_7 = -[Q^+] + y_6 + y_8 + y_9 + y_{10} \quad \text{(7), (j)}$$

$$y_8 = \frac{\theta_8 y_1}{(\theta_8 + y_7)} \quad \text{(8), (b)}$$

$$y_9 = \frac{\theta_9 y_3}{(\theta_9 + y_7)} \quad \text{(9), (c)}$$

$$y_{10} = \frac{\theta_7 y_5}{(\theta_7 + y_7)} \quad \text{(10), (a)}$$

The nine parameters form the vector, θ, given by: $\theta = [\alpha_1, E_1, \alpha_2, E_2, \alpha_{-1}, E_{-1}, K_1, K_2, K_3]$ and the predicted concentrations form the vector, y, given by: $y = [HA, BM, HABM, AB, MBMH, M^-, H^+, A^-, ABM^-, MBM^-]$.

The model initial conditions and the initial parameter estimates are given in Appendix A.

As a follow-up to the 1980 FOCAPD conference, the above problem was distributed to its 165 participants in 1981. Eleven of these researchers agreed to tackle this problem and five groups submitted acceptable solutions. Before describing these solutions and the methodologies behind them, we briefly summarize the algorithms that were used. We classify these in terms of solving the DAE model, choosing an appropriate objective function from maximum likelihood, and optimizing for the parameters.

SOLVING THE DAE MODEL

Numerous methods have been developed for the solution of initial value ordinary differential equations. Also, thorough analyses of these methods have led to a fairly complete classification according to stability, accuracy and performance. As a result of these comparisons, the most commonly used ODE IV methods are:

1) Explicit Runge-Kutta and linear multistep methods for nonstiff systems, and
2) Semi-implicit Runge-Kutta and the Gear methods for stiff problems.

Runge-Kutta Methods

Given a differential model of the form:

$$\frac{dy}{dt} = f(y,t) \quad (11)$$

the Runge-Kutta (R-K) techniques of integration use formulas defined by:

$$y_{n+1} = y_n + \sum_{i=1}^{r} b_i q_i$$

with

$$q_i = hf\left(t_n + c_i h, y_n + \sum_{j=1}^{r} a_{ij} q_j\right)$$

where n = step, counter
h = step size
r = order of Runge-Kutta method

The coefficients a_{ij}, b_i, c_i are determined by matching the above series with a Taylor series expansion of desired order.

When $a_{ij} = 0$ for $j \geq i$ the q_i can be calculated in order and the R-K method is called explicit. When $\alpha_{ij} = 0$ for $j > i$, and $a_{ij} \neq 0$

for $j = i$, the method is called semi-implicit. Finally, when $\alpha_{ij} \neq 0$ for $j > i$, the method is termed implicit. While implicit R-K methods have a drawback in that the q_i must be solved iteratively, semi-implicit methods can be linearized and solved with only one iteration per step. Numerous methods exploit this strategy (see, e.g., Carnahan and Wilkes, 1981). Chan et al., 1978). Prokopakis and Seider (1981), and Michelsen (1976) give accounts of semi-implicit Runge-Kutta schemes.

Linear Multistep Methods

Linear multistep methods for numerical integration of differential equations have the following form:

$$y_n = \sum_{j=1}^{l_1} a_{nj} y_{n-j} + h_n \sum_{j=0}^{l_2} b_{nj} f_{n-j} \tag{12}$$

with coefficients, a_{nj} and b_{nj}, determined by postulating the solution as an interpolating polynomial.

Gear (1971) proposed a family of formulas that are "nearly" A-stable. Hence, formulas of order r are obtained with $l_1 = r$ and $l_2 = 0$, i.e.:

$$y_n = \sum_{j=1}^{r} a_{nj} y_{n-j} + h b_{n0} f_n \tag{13}$$

These formulas, called backward difference formulas, are implemented in the popular Gear, Episode, and LSODE (Byrne, 1981) software packages. The desirable stability property they have, which has been termed stiffly stable, allows the numerical algorithm to track the model components accurately with larger step sizes after the transient or stiff region has passed. A discussion of backward difference formulas can be found in Carnahan and Wilkes (1981).

Systems of Differential/Algebraic Equations (DAE)

Models composed of nonlinear algebraic and stiff ordinary differential equations of the form:

$$\frac{dy_a}{dt} = f(y,\theta) \tag{14}$$

$$g(y,\theta) = 0$$

where y = vector of state variables
y_a = subvector of state variables to be solved from differential equations.
θ = vector of adjustable parameters

often occur in chemically reacting systems as well as in other fields such as circuit analysis or transient flowsheet simulation. Note that the system is coupled, with variables from the algebraic equations needed to compute the state variables y_a and vice versa.

Gear (1971) stated that for stiff equations, a backward difference formula should be used to discretize the differential equations. The algebraic equations could then be combined with the discretized equations and the resulting algebraic system could be solved by Newton's method. The discretized ODEs have the form of Eq. 13:

$$y_n - \sum_{j=1}^{q} a_{nj} y_{n-j} - h b_{n0} f_n = 0 \tag{15}$$

The Jacobian of this equation, needed in the Newton-Raphson iteration, is given by: $I - h b_{n0} J(y_n)$, where $J = \partial f_n / \partial y_n$.

The combined system of algebraic and differential equations can be written in the general form:

$$Ey' = F(y,\theta,t) \tag{16}$$

where F = combined f and g equations
y' = derivatives of y
θ = vector of adjustable model parameters
E = matrix partitioned as $[I/O]^T$ so that Eqs. 14 and 16 are equivalent

For DAE systems the analog of Eq. 15 has the Jacobian:

$$E - h b_{n0} \frac{\partial F}{\partial y}$$

Petzold (1982) has shown that solving DAE systems may lead to far more difficulties than solving a standard ODE model. From a practical standpoint, enforcing convergence by reducing the step size leads to an ill-conditioned Jacobian because E is singular for DAE systems.

Moreover, using linear, nonhomogeneous systems as an example, Petzold describes a number of problems with error estimates, termination criteria, and convergence failures when discontinuities in the forcing function are present. She points out that DAE systems can have many similarities to stiff systems and, in certain cases, have errors that do not vanish as the step size goes to zero. While the example problem described in this paper does not fall into this problem class, Petzold advises extreme caution in solving *nonlinear* DAE systems. Many of her suggestions for dealing with the above problems have been incorporated into her code, DASSL, which is described later.

MAXIMUM LIKELIHOOD METHOD

The choice of the objective function used in parameter estimation should incorporate the specific error structure of the experimental data if the best fit of the data is to be achieved. The likelihood function provides a general formulation for the objective function by means of which many types of error relationships can be represented.

To apply the likelihood function assume a relationship of the form:

$$z_{uj} = y_{uj}(\theta) + \epsilon_{uj}$$

where z_{uj} = measured value for component j
$y_{uj}(\theta)$ = computed value of component j from model in uth experiment
θ = vector of adjustable model parameters
ϵ_{uj} = residual error, assuming the model is correct

Here, the measured variables, z, have experimental errors associated with their values. The form of these experimental errors is given by a covariance matrix; the diagonal elements of the matrix represent the independent variances of the measured variables and off-diagonal terms represent estimates of correlated or dependent error relationships.

If it is assumed that the measured variables, z, have a Gaussian probability distribution, $p(z)$, with mean η, and covariance V, expressed as $N(\eta, V)$, then:

$$p(z) = (2\pi)^{-m/2}(V)\ \det^{-1/2}(V)\ \exp[-1/2(z-\eta)^T V^{-1}(z-\eta)] \tag{17}$$

for a single experiment in which m variables are measured with a covariance V between these measured variables. The experimental error vector, e_u, defined as the difference, $E(z-\eta)$, thus has the normal distribution, $N(0, V_u)$. If n experiments are carried out, in each of which m variables are measured with covariance V_u between the variables in experiment u, but with no correlation between measured variables in the different experiments, then:

$$p(e) = (2\pi)^{-mn/2} \sum_{u=1}^{n} \det^{-1/2}(V_u) \left(\exp[-1/2 \sum_{u=1}^{n} e_u^T V_u^{-1} e_u] \right) \tag{18}$$

If one assumes that the proposed model is correct and that the parameters θ are not far from the optimal parameters, the residual vector, ϵ_u, can be substituted into Eq. 18 as an adequate representation of the measurement errors. This leads to the likelihood function:

$$L(\theta) = (2\pi)^{-mn/2}\sum_{u=1}^{n}\det{}^{-1/2}\,V_u \exp\left(-1/2\sum_{u=1}^{n}\epsilon_u{}^T(\theta)V_u{}^{-1}\epsilon_u(\theta)\right) \quad (19)$$

The maximum likelihood method seeks those values of the adjustable model parameters θ for which the probability of obtaining values is maximized.

To solve our parameter estimation problem, the following error structure for the measured values was assumed in the Dow problem statement.

- Observation errors at different time points t_u are uncorrelated
- The errors at t_u are normally distributed with zero mean and a covariance matrix, V_u
- It is assumed the measured variables are independent so that the covariance matrix is diagonal
- It is assumed the error is heteroscedastic and the diagonal element of V_u corresponding to component j, v_{uj}, is a power transformation depending on the magnitude of the expected value of z_{uj}. Here, the model prediction, y_{ui}, for time t_u and component j is substituted for the expected value.

$$v_{uj} = \omega_i{}^2 y_{uj}{}^{\gamma_j} \quad (20)$$

where ω_j^2 = variance coefficient for component j
γ_j = 0, absolute error constant throughout experiment
γ_j = 2, relative error constant throughout experiment

Substituting this diagonal covariance matrix into the likelihood function yields:

$$\log L = -nm/2\,\log(2\pi) - 1/2\sum_{u=1}^{n}\sum_{j=1}^{m}\frac{(z_{uj}-y_{uj})^2}{y_{uj}{}^{\gamma_j}\omega^2{}_j} - 1/2\sum_{u=1}^{n}\sum_{j=1}^{m}[2\log\omega_i + \log(y_{uj})^{\gamma_j}] \quad (21)$$

where n = total number of measurements for a component
m = number of measured components
y_{uj} = model prediction which is a functin of θ

Note that the knowledge of the error structure of the data is instrumental in the derivation of the final model and objective function equations. For some functions the unknown statistical parameters are not present, and in other instances they can be removed by simple transformation at the optimum. In this particular problem, the likelihood function is to be maximized for the θ, ω, and γ parameters. At this maximum,

$$\frac{\partial \log L}{\partial \omega_i} = 0$$

Using this relation results in solving explicitly for ω_j:

$$\omega^2_j = 1/n\sum_{u=1}^{n}\frac{(\epsilon_{uj})^2}{y_{uj}{}^{\gamma_j}} \quad (22)$$

where ϵ_{uj} is the model error, $z_{uj} - y_{uj}$. Substituting Eq. 22 into Eq. 21 yields:

$$\log L = -nm/2[\log(2\pi) + 1] - 1/2\sum_{j=1}^{m} n\,\log\left[1/n\sum_{u=1}^{n}\frac{(\epsilon_{uj})^2}{y_{uj}{}^{\gamma_j}}\right] - 1/2\sum_{u=1}^{n}\sum_{j=1}^{m}\gamma_i\log y_{ui} \quad (23)$$

This objective function was suggested by Reilly et al. (1977). In this case study, several participants modified this objective function even further. If we assume that the errors in each experiment are Gaussian and distributed with the same diagonal covariance matrix but are not heteroscedastic, then all γ_j's are set to zero and Eq. 23 simply becomes:

$$\log L(\theta) = mn/2[\log(n/2\pi) - 1] - n/2\sum_{a=1}^{m}\log\sum_{u=1}^{n}\epsilon_{ua}{}^2(\theta) \quad (24)$$

Thus, one must maximize Eq. 24 for parameters θ and estimate v_a from Eq. 22 using the optimal parameters, θ (Note that by maximizing log L the residuals should be smaller than the actual errors, since the parameters θ were chosen to make the residuals as small as possible. Because of this, some bias is introduced. Assuming a general unknown covariance matrix leads to an expression similar to Eq. 24 (Bard, 1974, p. 66).

Finally, if we assume that each measured value has Gaussian errors with the same known variance, then the likelihood function reduces to a simple residual sum of squares which can be minimized to fit the observed data, i.e.:

$$\min \Phi = \sum_{u=1}^{n}\epsilon^2{}_u(\theta)$$

where ϵ_u = residual vector for uth experiment

It should be noted that the above derivations apply only to experimental runs with no missing data. This is not true in the example considered here and different approaches for missing entries were used in the solutions described below. Analyses of this problem can also be found in Bard (1974) and Stewart and Sorensen (1981).

OPTIMIZATION ALGORITHMS FOR PARAMETER ESTIMATION

A number of gradient-based algorithms have been developed which exploit least-squares structure and significantly reduce the computation necessary to find the optimal parameters. The important steps of these algorithms are the calculation of a search direction and determination of the step length to take along this direction. The iteration step for the optimization variables, θ, can be expressed as:

$$\theta_{k+1} = \theta_k + \delta(H)^{-1}g$$

where δ = step length
H = Hessian matrix of the objective function or its approximation
g = gradient of objective function

One can approximate Hessian information by exploiting the form of the objective function. In particular, Gauss-Newton and Marquardt methods take advantage of the structure of several functions derived from maximum likelihood. For example, the simple least-squares function is given by:

$$\Phi(\theta) = \sum_{u=1}^{n}(z_u - y_u(\theta))^T(z_u - y_u(\theta))$$

or

$$\Phi(\theta) = \sum_{u=1}^{n} \epsilon_u(\theta)^T \epsilon_u(\theta)$$

where y_u = vector of model predictions, y_{ui}, at time u
z_u = vector of measured values at time u

with the gradient represented by:

$$g(\theta) = 2\sum_{u=1}^{n} \epsilon_u{}^T \frac{\partial \epsilon_u}{\partial \theta}$$

or

$$g(\theta) = -2\sum_{u=1}^{n} \epsilon_u{}^T \frac{\partial y_u}{\partial \theta}$$

and the Hessian obtained by:

$$H(\theta) = -2\sum_{u=1}^{n} \epsilon_u{}^T \frac{\partial^2 y_u}{\partial \theta^2} + 2\sum_{u=1}^{n} J^T J$$

with

$$J = -\frac{\partial y_u}{\partial \theta}$$

Note that the first term in the Hessian formulation contains the residual vector, ϵ_u. Assuming that the residual is small, the Hessian can be approximated only by the first derivatives contained in the second term of the Hessian. Using this Hessian to find a search direction, s_k, leads to the Gauss-Newton method:

$$\sum_{u=1}^{n} J^T J\, s_k = -\sum_{u=1}^{n} J^T \epsilon$$

However, problems can occur when solving the least squares problem if the matrix, J, does not have full rank.

The Levenberg-Marquardt method proposes a nonnegative addition to the Gauss-Newton Hessian approximation to insure a positive definite Hessian and a descent search direction. The alternative Hessian is proposed as:

$$H = \left(\sum_{u=1}^{n} J^T{}_k J_k + \mu I\right)$$

and the search direction is chosen to satisfy:

$$\left(\sum_{u=1}^{n} J^T{}_k J_k + \mu I\right) s_k = -\sum_{u=1}^{n} J^T \epsilon$$

The choice of μ is crucial in order to insure a descent direction. Bard (1974) summarizes Marquardt's original algorithm for the selection of μ.

Another way of insuring descent directions with Gauss-Newton type methods is through rotational discrimination (Fariss and Law, 1979), Here the H matrix undergoes a spectral decomposition to:

$$H = R\,\Lambda\,R^T$$

where Λ is a diagonal matrix. Defining a new coordinate transformation:

$$\tilde{s} = R^T s$$
$$\tilde{g} = R^T g$$

leads to each element of s given by:

$$\tilde{s}_i = -\nu_i \tilde{g}_i$$

$\nu_i = \lambda_i^{-1}$ if $\lambda > \zeta$, or

$\nu_i = 0$, otherwise

Here ζ is a small positive tolerance. Other methods for handling nondescent directions after spectral decomposition are given in Bard (1974).

Finally, several quasi-Newton methods are available for unconstrained minimization (see Gill et al., 1981). These methods assume nothing about the structure of the objective function, but instead approximate second derivatives from past changes to the gradient vector. While these methods ensure positive definiteness of H, they may become unstable if the problem is ill-conditioned. These methods were only tried in the attempted solution and yielded unfavorable results.

Computing the Gradient

In parameter estimation, gradient calculations needed in the optimization algorithm usually make up the most expensive step. Here, the adjustable model parameters, θ, do not appear explicitly in the least-squares or likelihood function. Thus, the calculation of the gradients of the objective function with respect to the model parameters cannot be calculated analytically if the model predictions, y, are not solved explicitly from the model. Two alternative approaches are available to calculate the gradients.

The first alternative is to use either forward or central difference formulas. In either case, the choice of the finite-difference perturbation size is important. Its value must be chosen to minimize the effects of truncation error and computation error. Truncation error is due to the neglected higher-order terms in the Taylor series expansion and is proportional to the perturbation size for the forward difference formula, and perturbation size squared for the central difference formula. The computation error is due to errors in computing the function values used in the gradient approximation. This type of error is proportional to the inverse of the perturbation size for both finite-difference formulas. Therefore, it is important that the finite-difference interval be kept small to control the truncation errors, but not too small to lead to roundoff errors. Since forward differences require half the number of function evaluations needed for central difference approximations, the former are more efficient but less accurate. For instance, forward differences should not be used if the change in perturbed function values is small for a given step size, since computation errors will overshadow the gradient calculation. Thus, as the magnitude of the gradient approaches zero at the solution of an unconstrained problem, a switch to central differences may be required.

The second alternative for gradient calculation is through sensitivity analysis. Here, one needs to obtain $dy/d\theta$, where the values, y, are the model predictions used in the objective function. We differentiate both sides of Eq. 16 with respect to θ, and using the chain rule we have:

$$E\frac{\partial}{\partial \theta}\left(\frac{dy}{dt}\right) = \frac{\partial F}{\partial \theta} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial \theta}$$

Interchanging the order of differentiation yields:

$$E\frac{d}{dt}\left(\frac{\partial y}{\partial \theta}\right) = \frac{\partial F}{\partial \theta} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial \theta} \qquad (25)$$

In order to obtain $\partial y/\partial \theta$ one must solve the DAE model together with the set of simultaneous differential and algebraic equations given by Eq. 25, with the quantities $\partial F/\partial \theta$ and $\partial F/\partial y$ determined by simple differentiation. Equations 25 are called sensitivity equations and the values $\partial y/\partial \theta$ are called sensitivity coefficients. Bard (1974) illustrates the use of these equations to calculate the gradient for a nonlinear differential equation model.

SOLUTIONS OF PARAMETER-ESTIMATION PROBLEM

Five research groups provided solutions to the parameter-estimation problem formulated in the first section. All of the solu-

tions were obtained by reformulating the model or the parameters. In three cases, different initial parameter estimates were used for the optimization. Needless to say, these modifications preclude a straightforward comparison of the model solution and optimization algorithms. However, some conclusions can be drawn from this study and much can be gained from the problem-solving techniques used for this parameter estimation problem.

The investigators are listed with their affiliations in Table 1. A summary of results is given in Table 2. However, since only three solutions regressed on all four measurements, only these could be compared with a common objective function. Instead, a transformation of the final parameter vectors for each solution is given in Table 3. The explanation for this transformation will be discussed after describing the solutions. Due to the different nature of each solution, the most straightforward comparison can be made simply by plotting the model, solved with each set of final parameter values, against the three data sets. However, for the sake of brevity we have chosen to present only the first (and largest) data set in Figures 2 through 8, with a plot of the model with initial parameter estimates in Figure 1. It should be cautioned that systematic errors may be observed in the fits of this data set due to the need to compensate for opposing errors in the other sets. Hence, while Figures 2 through 8 give a qualitative comparison of the five solutions, they may not be totally objective.

Finally, the optimal parameter values and their confidence intervals are listed in Appendix B. Due to the different nature of the solutions and for the sake of brevity, we refrain from reporting and interpreting final covariance matrices on a consistent basis. Instead we describe the characteristics of the solutions and model more qualitatively. More specific details about each solution are available on request from the first author.

Attempted Solution

To provide a measure of the difficulty of this problem, a solution was attempted by the second author using the model, objective function, and starting point stated above. The intention of this exercise was to see how easily the problem could be solved with available software and no "coaxing" on the researcher's part. A brief account of the computational difficulties encountered is given below.

The original model consists of a DAE system with six ODE's and four algebraic equations and is stiff for the initial set of parameters. Very few numerical integration packages have the built-in capability of solving DAE's simultaneously. One of these is the DASSL software package developed at the Sandia National Laboratory in Livermore, California. This code uses the stable Gear backward difference corrector formulas to convert the differential equations to algebraic equations. The algebraic and differential equation variables are then solved using a modified Newton method. With this approach, convergence problems can occur in solving the model for certain choices of the model parameters picked by the optimization scheme. It was observed, however, that convergence failures occurred less often if tighter tolerances are imposed for each step of the DAE solver.

Two optimization strategies, a modified Complex (MC) routine (Box, 1965) and a quasi-Newton (QN) gradient-based strategy were chosen for parameter adjustment. Both algorithms handle parameter bounds easily but do not take advantage of the structure of the objective function.

The two different optimization techniques required different model tolerances to handle the limitations of the model solving routine. With the MC routine, convergence problems could be handled by defining a constraint function that was violated whenever the model failed to converge. This prevented the Complex algorithm from using an incorrect value of the objective function but required the algorithm to choose more points. Because the model convergence problems could not be eliminated completely by tightening the error tolerance, a loose error tolerance was used to evaluate more points for a fixed CPU time. For the gradient-based algorithm, however, it was essential for the model to converge as often as possible. Thus, tight relative and absolute error tolerances were used in order to compute gradients and search directions. If model convergence failure occurred during a forward difference perturbation then the gradient with respect to the perturbed variable was set to zero. This heuristic helped to avoid grossly inaccurate search directions as a result of failed gradient calculations.

Before beginning, a few minor changes were made in the formulation of the objective function. To prevent the second term in Eq. 21 from getting too large, the lower bound for the value of the model prediction, y_{uj}, was set to 1.0×10^{-5}. Also, since the experimental values of component *BM* are derived and the other

Table 1. Summary of Submitted Solutions*

Investigator	Objective Function	No. of Components Regressed	No. of Parameters Optimized	Final Model Formulation
(1) M. Caracotsios, W. E. Stewart & J. Sorensen, University of Wisconsin, Madison, WI	Unknown General V_{ij} Unknown γ_i	2	12	6 ODE's 4 Algebraic Eqs.
(2) R. H. Fariss, Monsanto Plastics and Resins, Indian Orchard, MA	Unknown Diagonal V_{ij} Unknown γ_i	4	17	3 ODE's 4 Algebraic Eqs.
(3) B. S. Ahn, Korean Advanced Institute of Science and Technology, Seoul, Korea	Unknown Diagonal $V_{ij,\gamma i} = 1$	4	9	3 ODE's
(4) R. Klaus, T. Rimensberger & D. W. T. Rippin, Eidgenoessische Technische Hochschule, Zurich, Switzerland	Unknown Diagonal $V_{ij,\gamma i} = 0$	4	9	4 ODE's
(5) B. Sarup, M. Michelsen & J. Villadsen, Danmarks Tekniske Hojskole, Lyngby, Denmark	Unit Matrix V_{ij} (Least Squares)	3	8	3 ODE's

*See Appendix B for initial and final parameter estimates of each investigator.

TABLE 2. SUMMARY OF OTIMIZATION RESULTS

	Participants (Refer to Table 1)					
	(1)	(2)	(3)	(4)	(5)	(*)
Optimization algorithm	Quadratic Min.	Rot. Discrim.	Marq.	Marq.	Marq.	QN (MC)
Final objective function value	−1,009.9	753.84	561.72	362.80	0.233	−199.9 (−295.0)
Equivalent value, Eq. 23	——	753.84	561.72	656.94	——	−199.9 (−295.0)
No. of iterations	19	24	17	47	20	55 (115)
No. of function evaluations	38**	577	104	144	20**	637 (290)
CPU time, min	21.53	38.35	14.30	***	0.17	459.0 (55.27)
Computer system	VAX 11/780	IBM 3081	Cyber 174/18	CDC 6500	IBM 3033	DEC-20

*Attempted solution (Damiano).

**Sensitivity equations evaluated once per iteration.

*** Not reported due to excessive I/O at execution.

three measurements were normalized, only two of the four "experimental" components, *HA* and *HABM*, were used. Logarithmic transformations of the kinetic parameters were made for manipulation by the optimization algorithms. Finally, the heteroscedasticity parameters, γ_i, were held constant at 1.0 throughout the optimization runs.

The results of the optimizations using *MC* and *QN* are displayed in Figures 2 and 3, respectively. As mentioned above, much tighter model tolerances were chosen for *QN*, with the result that it required much more computational effort than *MC*. Both algorithms terminated far from the optimum after exhausting too much CPU time. The DAE solver had failure rates of 28 and 12% for *MC* and *QN*, respectively. A summary of the results is given in Table 2.

First Solution

The first group of investigators also used DASSL to solve the DAE model. To avoid the convergence problems discussed above, after some trial runs the error tolerances were adjusted automatically over the course of the optimization. Also, sensitivity equations given by Eq. 25 were used to calculate the gradient of the objective function with respect to all of the model and statistical parameters.

Several modifications were made to the maximum likelihood objective function. First, because of the data dependencies mentioned above, at most two observations were used (*HABM* and *AB*) from each time point t_u. This number was reduced to one whenever a concentration was missing; [AB] was chosen for deletion when [HA] was missing. Secondly, the diagonal variance in Eq. 20 was replaced by a full covariance matrix with elements: $V_{uij} = \omega_{uij} y_{ui}^{\gamma_i/2} y_{uj}^{\gamma_j/2}$. Here ω_{uij} is normally equal to the parameter ω_{ij}, but is replaced by δ_{ij} if either y_{ui} or y_{uj} is missing. This generalizes the objective function to:

$$\Phi = \sum_{u=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}(V_u^{-1})_{ij}\tilde{e}_{ui}\tilde{e}_{uj} + \sum_{u=1}^{n}\log(\det(\omega_u)) + \sum_{u=1}^{n}\sum_{i=1}^{m}\gamma_i \log y_{ui}$$

where $\tilde{e}_{ui} = e_{ui}/y_{ui}^{\gamma_i/2}$. The model parameters were transformed as follows:

$$\log k_1 = p_1 - (1/T - 1/T_b)p_7$$
$$\log k_2 = p_2 - (1/T - 1/T_b)p_8$$
$$\log(k_1/k_{-1}) = -p_3/T_b - (1/T - 1/T_b)p_9$$
$$\log(K_1) = -p_4/T_b$$
$$\log(K_2) = -p_5/T_b$$
$$\log(K_3) = -p_6/T_b$$
$$T_b = 342.15\ K$$

This transformation was chosen to prevent the parameters from varying over large orders of magnitude and to reduce the intercorrelation of the parameters in the rate constants. Finally, the elements of the measurement covariance matrix were added as regression parameters.

The successive quadratic minimization algorithm described in Stewart and Sorensen (1981) was applied here. This approach is a generalization of the Gauss-Newton method that allows for missing data, linear constraints on the parameters, and general unknown covariance. Solution of this problem proceeded in two stages. After determining that the *HABM* and *AB* measurements were probably not independent, the problem was formulated without heteroscedasticity. After 15 iterations with Marquardt's method, the problem converged but the normal equations were ill-conditioned. Closer inspection of the Hessian showed that parameters p_3 and p_5 could not be estimated simultaneously.

The second stage proceeded from this solution with the "heteroscedastic" objective function given above and p_5 fixed. Convergence with successive quadratic minimization was obtained after four additional iterations and the objective was found to be

TABLE 3. SUMMARY OF OPTIMAL SOLUTION VECTORS

	Participants				
Transformed Parameters	(1)	(2)	(3)	(4)	(5)
$k_1(342)$	1.88	1.84	1.94	2.04	2.21
$k_2(342)$	2.73	2.89	2.38	2.68	2.78
$E_1(\times 10^{-3})$	18.74	18.48	18.84	18.26	17.84
$E_2(\times 10^{-3})$	18.88	19.07	17.87	18.41	18.85
$E_3(\times 10^{-3})$	25.67	26.05	25.15	21.90	25.20
K_1/K_3	1.44	1.44	1.42	1.44	1.43
$k_{-1}(342)K_3/K_2$	9.84	10.92	8.52	9.98	10.14

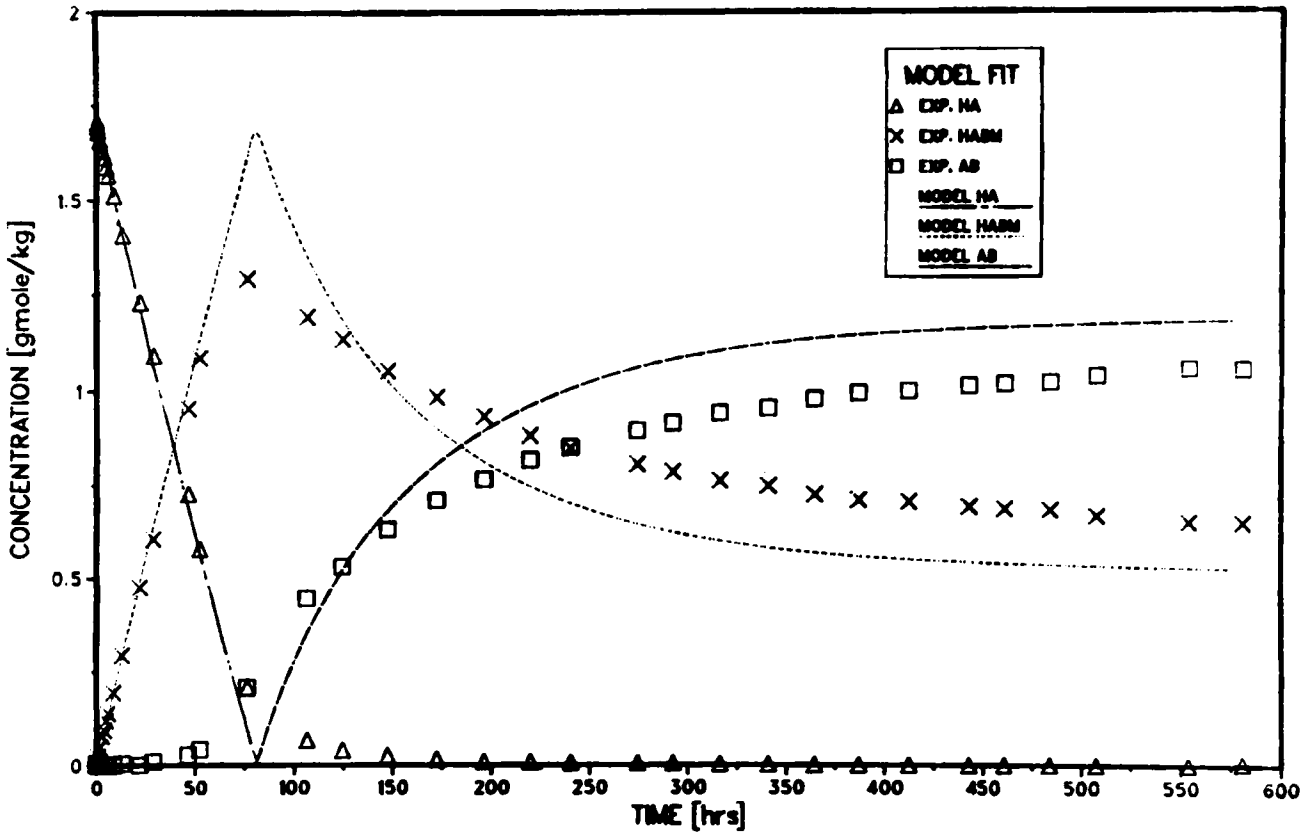

Figure 1. Model with initial parameter estimates plotted with first data set.

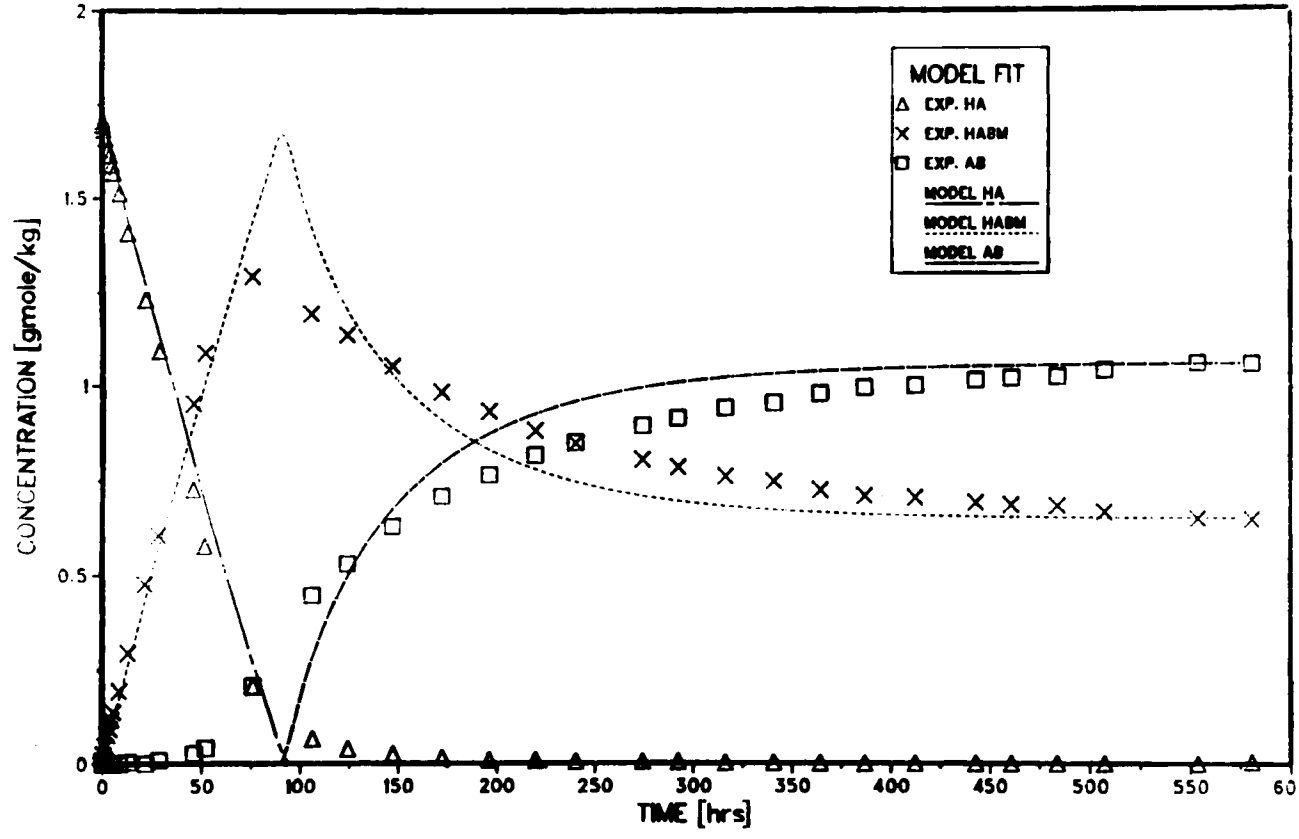

Figure 2. Model with optimal parameter estimates from MC attempted solution.

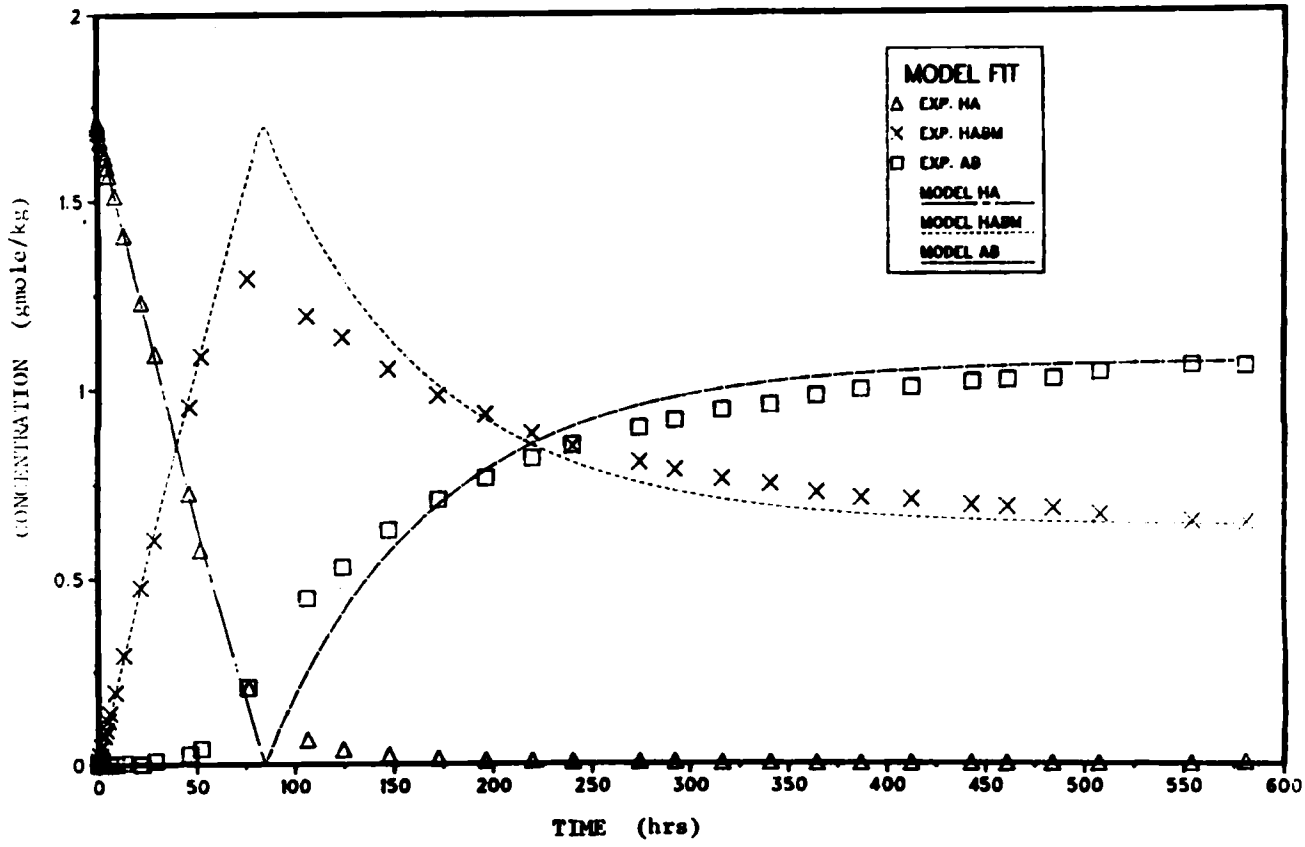

Figure 3. Model with optimal parameter estimates from QN attempted solution.

independent of p_6 as well. The parameter γ_{HABM} converged to its lower bound of zero, while the final γ_{AB} was 0.19 ± 0.12. Figure 4 shows how this solution fits the experimental data.

Second Solution

Results from the second solution are shown in Figure 5. Here the DAE model was transformed by noting that three of the six ODE's are dependent and can be reduced to algebraic equations. The revised model consisted of Eqs. 1, 4, and 6, with predictions for y_2, y_3, and y_5 calculated by material balances using y_1, y_4, and y_6. Problems were observed in the numerical integration scheme with y_6 becoming negative. To eliminate these problems, Eqs. 1 and 6 were divided by y_1 and y_6, respectively. This left the left hand sides of Eqs. 1 and 6 as $d(\log y_1)/dt$ and $d(\log y_6)/dt$, respectively. To avoid additional problems with $\log y_1$, Eq. 1 was further modified to force the derivative $d(\log y_1)/dt$ to zero smoothly for small values of y_1. The differential equations were first integrated using the Harwell stiff integrating package, DCO1AD, which uses Gear's backward difference formulas. Then for fixed values of y_1 to y_6, the original algebraic equations were solved using y_7 and Eq. 7 as a tear set and subsequently solving for y_8, y_9, and y_{10}. Here, a tight tolerance was required for the algebraic system to avoid numerical problems with the integration scheme. Also, the model parameters were transformed to:

$$\log k_1^\circ, \log k_2^\circ, \log k_{-1}^\circ, E_1, E_2, E_{-1}, \log K_1, \log K_2, \log K_3$$

The $\log k^\circ$ parameters were calculated relative to a base temperature of 67°C, according to the transformation:

$$\log k = \log k^\circ + (E / 0.67558)\frac{T - 67}{T + 273}$$

This participant opted for the above transformation because α values tend to vary over orders of magnitude and their interaction with the activation energy leads to an ill-conditioned problem. In addition, the objective function, Eq. 21 was modified by replacing the model prediction y_{uj}, by $1/2\ (z_{uj} + y_{uj})$. Since a diagonal covariance matrix was assumed, missing data were handled simply by using partial summations in the objective function. The rotational discrimination technique of Fariss and Law (1979) was used for optimization. Gradients were calculated using a central difference formula requiring twice as many function evaluations as gradient values. Finally, this investigator also optimized the statistical parameters γ and ω and noted that the solution is not unique because two linear combinations of the model parameters can be identified. More will be said later about dependency in the model parameters.

The solution was found by executing six passes of the optimization algorithm. In each pass a subset of the kinetic and statistical parameters were allowed to vary, with all parameters varying in the last pass as the optimum is approached. The solution was thus found by careful monitoring and appropriate intervention by the investigator over the course of the convergence history. No confidence intervals were reported by this investigator because he felt the assumed error structure was inappropriate for this problem. This is confirmed by his results in Appendix B. Here V_{22} is about the same value as the other variances but the parameters γ_2 and ω_2 that determine it are drastically different.

Third Solution

The third investigator reduced the DAE model to three independent ODE's that provide values for y_1, y_3, and y_5. The IMSL routine DGEAR was used to integrate the ODE's and Eq. 23 was

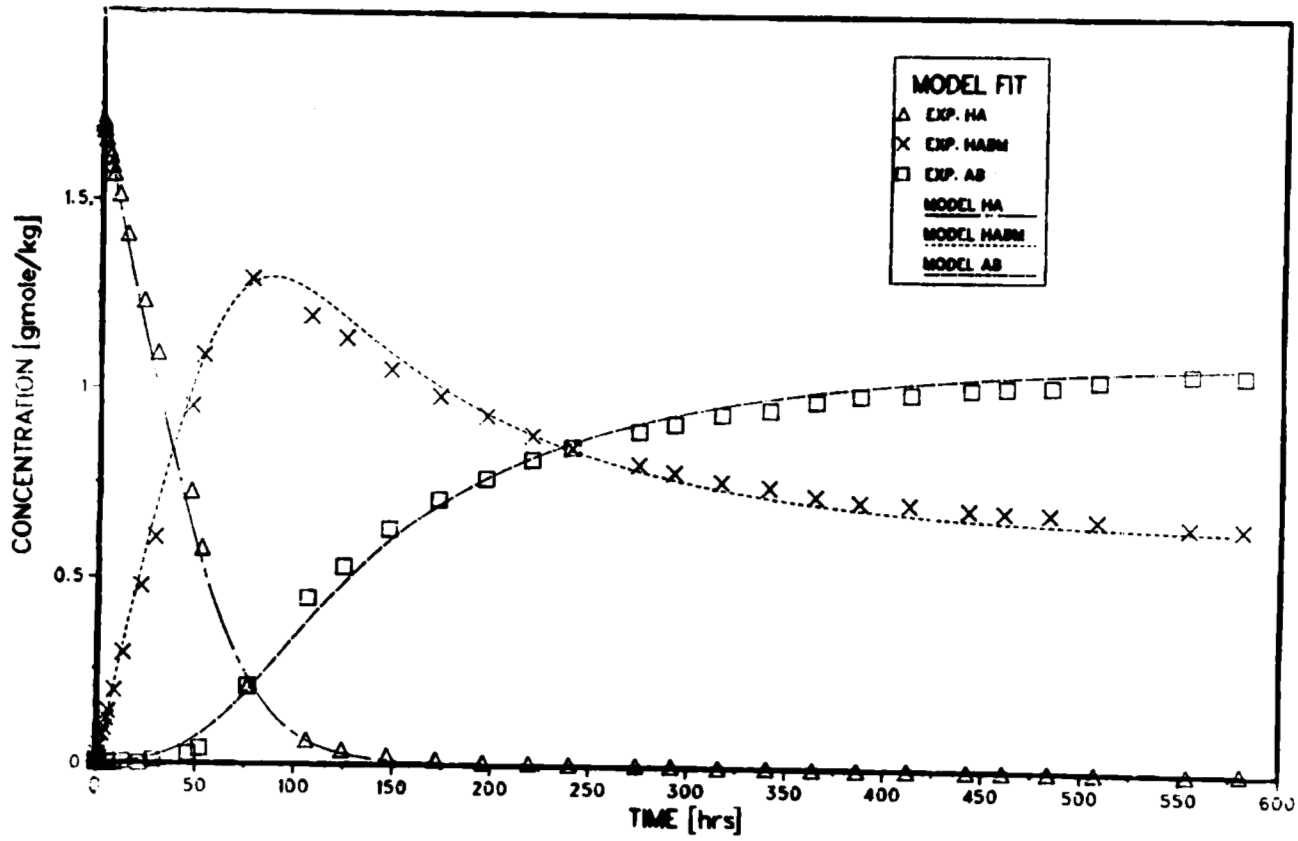

Figure 4. Model with optimal parameter estimates from solution (1).

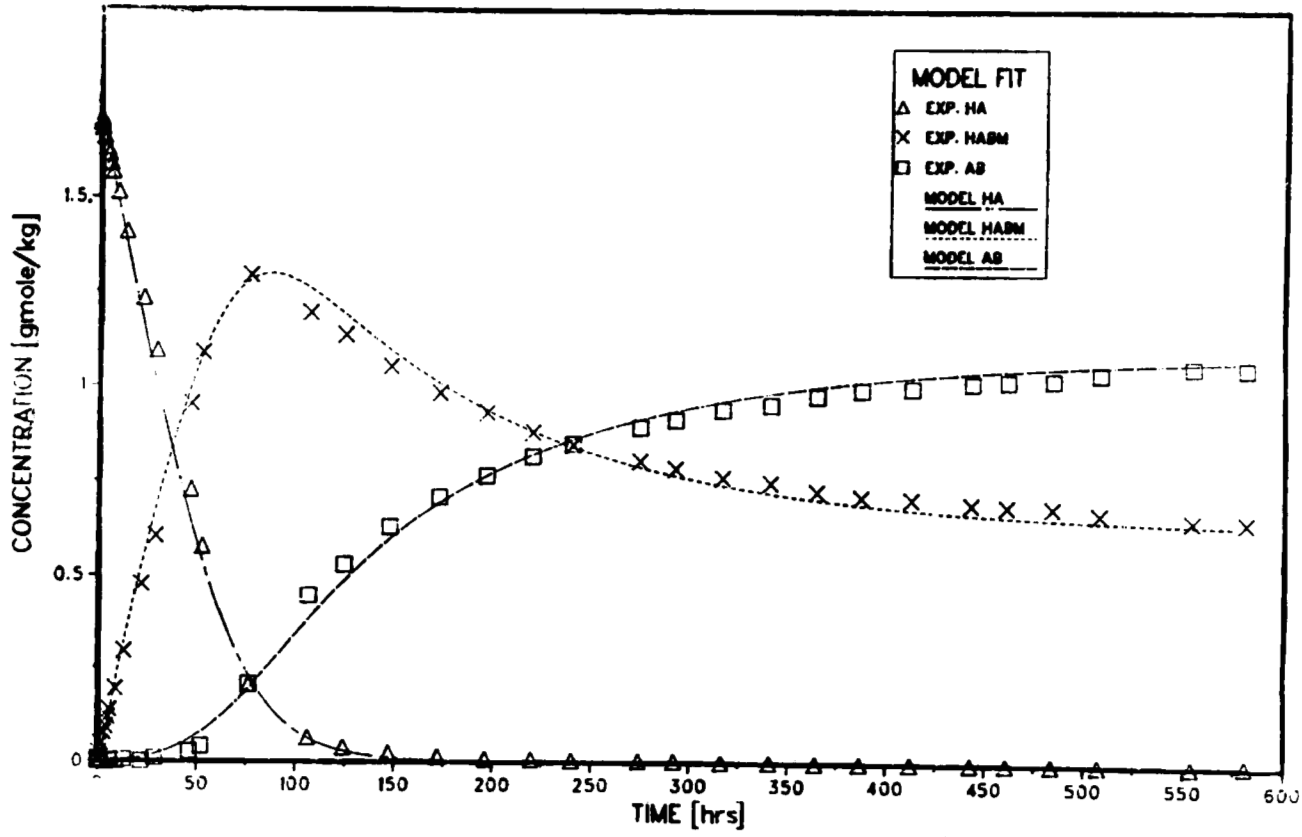

Figure 5. Model with optimal parameter estimates from solution (2).

changed so that the variance was proportional to the measured value raised to a power, rather than the value of the model prediction. This substitution set the last term and the denominator of the second term in Eq. 23 equal to a constant, which allowed him to optimize a simpler objective function, and still provide a value for the original objective function. The heteroscedasticity parameters, γ, were set to unity for the optimization runs and the Levenberg-Marquardt routine from the IMSL library was used to optimize the objective function. Gradients were calculated by finite difference as described in the IMSL manual. A partial summation was used in the objective function to account for the missing data.

Moreover, the starting point was altered by changing the initial value of K_2 from 10^{-11} to 10^{-15}. Even so, the problem required 104 function evaluations before it converged. Figure 6 displays the solution found by this investigator.

Fourth Solution

This participant reduced the six ODE's and four algebraic equations to four ODE's by substituting for the equilibrium relationships and dependent state variables. The model was solved using a semiimplicit third-order Runge-Kutta method by Michelsen (1976). The optimization routine was developed by Klaus and Rippin (1979). It contains a modified Marquardt algorithm with revisions by Fletcher (1971), a gradient projection technique for handling bounds, and several strategies for treating numerical difficulties. All gradients were calculated by finite difference. Zeroes were substituted for missing data in calculating the objective function.

Some problems were encountered in the integration step with initial values of the parameter estimates, but these were resolved by changing the initial estimate of K_2 from 1.0×10^{-11} to 1.0×10^{-17}. An optimal solution could not be reported with the original objective function. However, this investigator reported good results using an unknown diagonal covariance matrix without heteroscedasticity. The objective function takes the general form given by the second term in Eq. 24. The best results reported by this investigator are shown in Table 2 and Figure 7.

Fifth Solution

These participants elected to reduce the system of ODE's to only three by writing total mass balances. Further, using the three equilibrium relations and assuming almost undissociated species and weak acids, they reduced the DAE model to three coupled ODE's. These researchers refrained from using the likelihood function based on Eq. 20, stating that the dependencies in the data do not make this choice suitable. Instead, they used a simple least-squares function and excluded the derived *BM* measurements from the analysis; a partial summation of this function was used to account for the missing data. Sensitivity equations were formulated and solved simultaneously with the model equations to compute the gradients of the residuals with respect to the model parameters. The IMSL library subroutine, DGEAR, was used for this task, after modification to handle the sensitivity equations. The Harwell library routine (VBOlAD), based on the Levenberg-Marquardt method, was used to minimize the least-squares problem. Also, the rate constant parameters were reformulated using a transformation similar to that used in the second solution. Log k_i° and (E_i/RT_{ref}) $(T_{ref} = 340$ K) were the adjustable kinetic parameters and the parameters log (K_1/K_3) and log (K_3/K_2) were given starting guesses of zero. More will be said

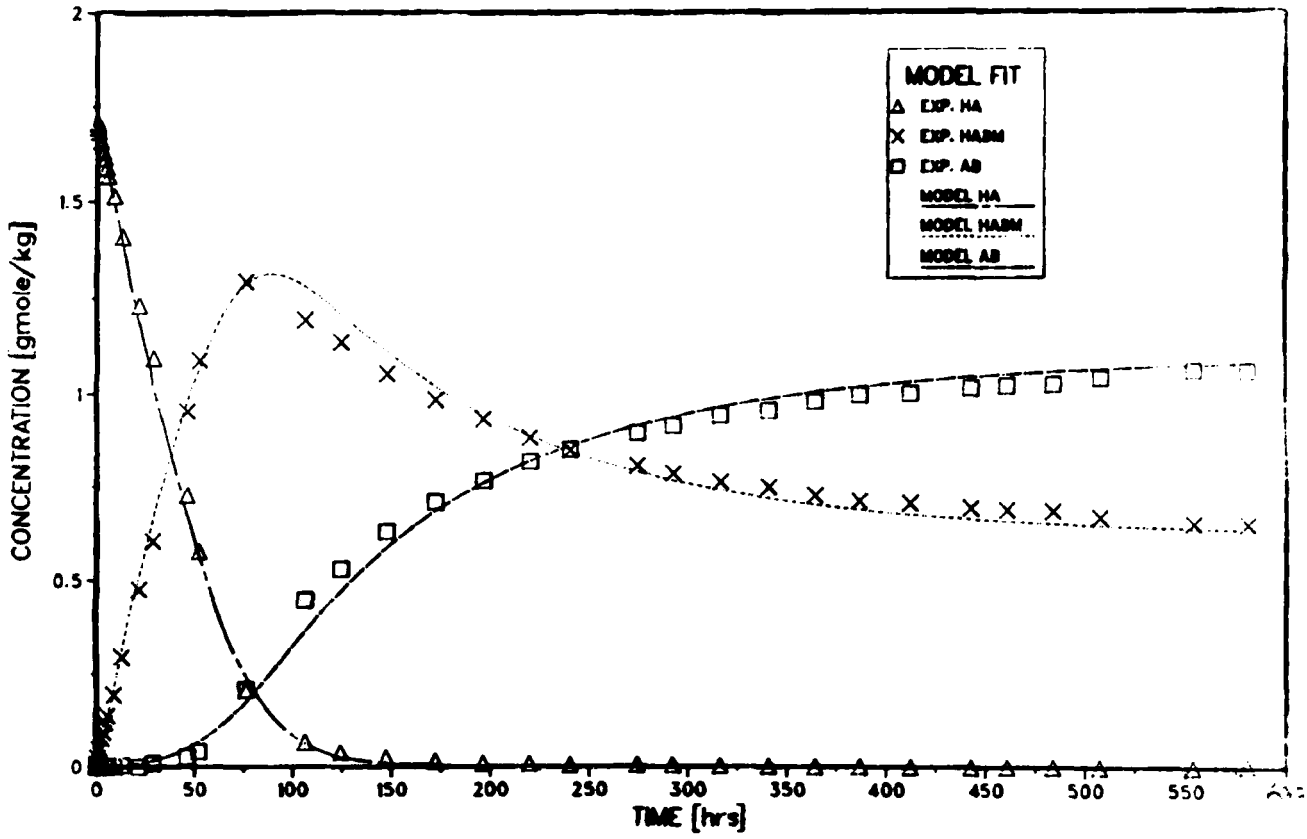

Figure 6. Model with optimal parameter estimates from solution (3).

about this transformation in the following sections. Results for this solution are presented in Figure 8.

SUITABILITY OF THE KINETIC MODEL

From the solutions described above, a number of observations can be made about the kinetic model and how it should be treated for parameter estimation. With their solutions, several investigators submitted very detailed analyses about the suitability of the model. A summary of their comments along with a few other points is given below.

First, it is readily seen from the kinetic mechanism that [*HA*] is predicted by the model to vanish at steady state. The [*HA*] measurements, however, level off at about $3 - 4 \times 10^{-3}$ gmol/kg. Thus the errors at long times will be biased by the model and render the error assumptions invalid. The second investigator attempted to remedy this situation by making the second reaction reversible and introducing new parameters for k_{-2}. His solution to the modified model yielded an improved maximum likelihood optimum and led to final [*HA*] of about 10^{-3} gmol/kg. He also found the modified model much easier to solve. Clearly the addition of the reverse step helps to remedy the model deficiency, although this step is not apparent from prior knowledge of this reaction.

Second, several investigators noted dependencies among the kinetic parameters. This not only leads to nonunique optimal solutions but also yields an H matrix that becomes singular. Consequently, one also encounters convergence problems. Through analysis of the Hessian matrix after spectral decomposition, the second investigator noted a "two-dimensional dependency" among the parameters k_{-1}, K_1, K_2, and K_3. In Table 3 we present values for K_1/K_3, $k^\circ_{-1}K_3/K_2$ and the remaining parameters for the five solutions. Although the original optimal parameter values vary a great deal for these solutions (see Appendix B), the transformed parameters in Table 3 are very similar.

The model can also be reduced to deal with a smaller parameter set through physical arguments. The fifth investigator performed part of this reduction by assuming that the small equilibrium constants would lead to $[H^+] \approx 0$. This immediately eliminates the equilibrium relations and reduces the model to eight parameters and three ODE's. After solving the regression problem, this assumption was also consistent with the results.

Finally, we note that three species were measured as experimental data and normalized to meet a mass balance constraint. The fourth component was derived from a relation on [*BM*] that generally does not hold. Consequently, these data dependencies and the model deficiency mentioned above clearly violate the assumed-error structure for which the maximum likelihood function was derived.

The first and fourth investigators observed that these measurements were not independent even if only two components were used in the objective function. Thus the unknown diagonal covariance assumption covariance should be replaced with an unknown general covariance, as was done by the first investigator.

Also, while the justification for the heteroscedasticity assumption is weak, the complexity of the objective function greatly increases. This becomes especially serious when [*HA*] goes to zero and the objective function becomes unbounded. Thus, during the optimization, γ_1 is forced to zero regardless of the actual error structure. To avoid this problem, the second and third investigators substituted either $z_{uj}^{\gamma_1}$ or $1/2(z_{uj} + y_{uj})^{\gamma_1}$ for $y_{uj}^{\gamma_1}$ in the objective function. The first investigators used $[\max(10^{-10}, y_{uj})]^{\gamma_1}$ in their calculations.

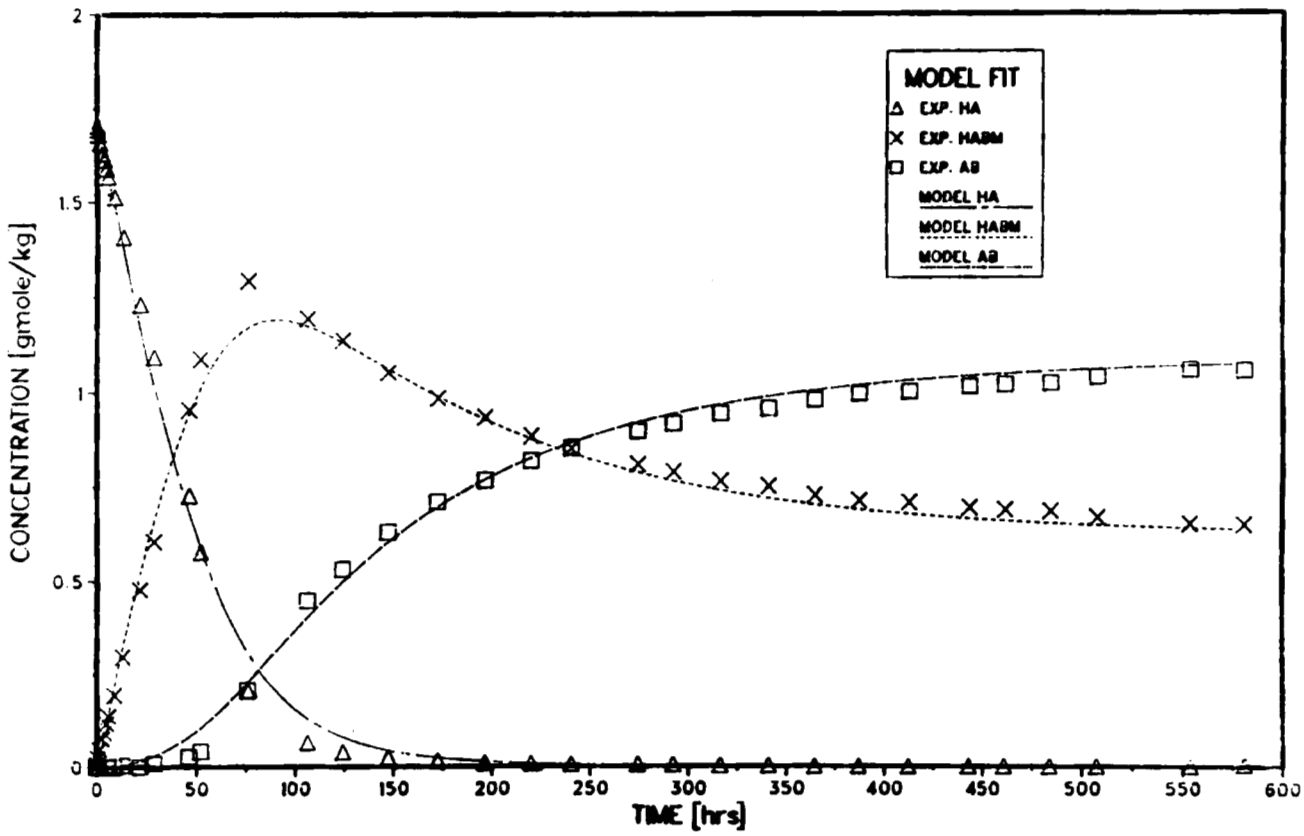

Figure 7. Model with optimal parameter estimates from solution (4).

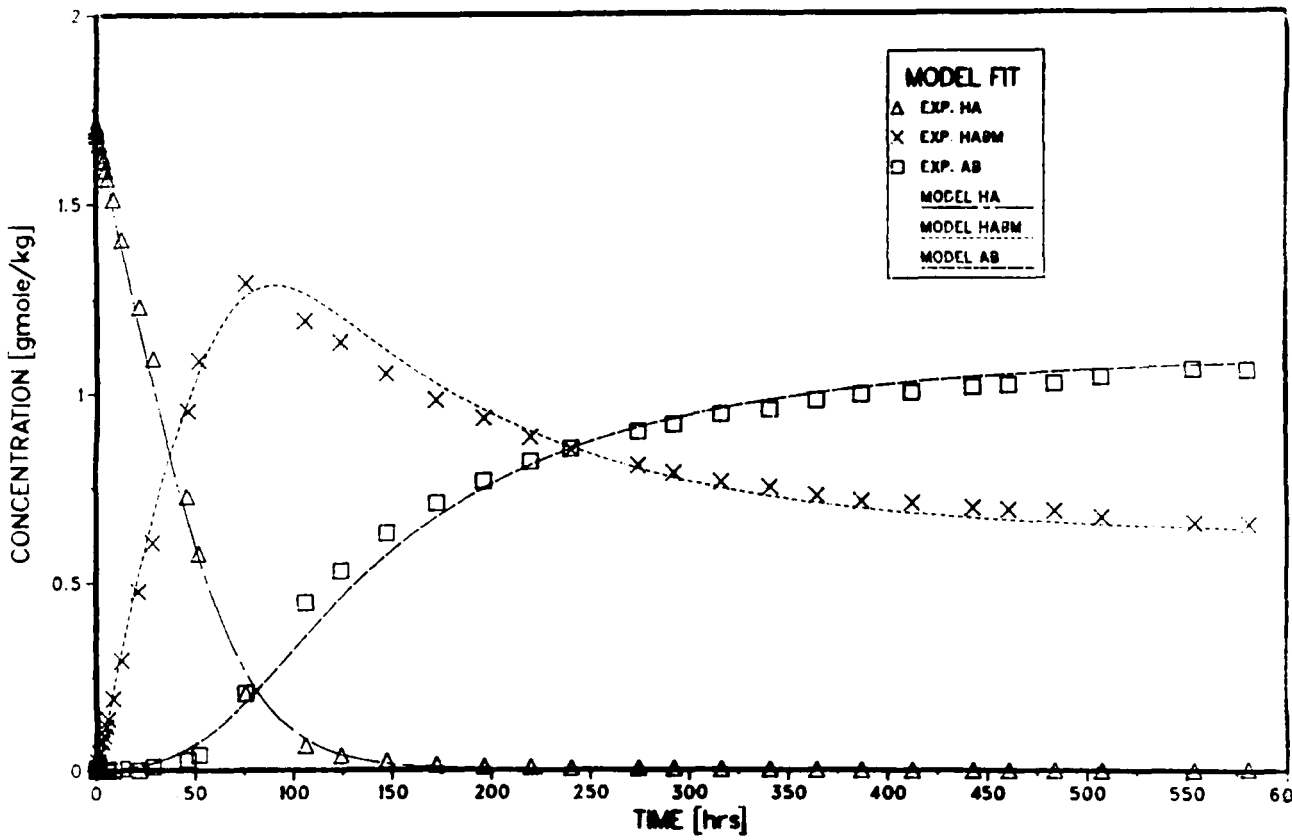

Figure 8. Model with optimal parameter estimates from solution (5).

GUIDELINES FOR PARAMETER ESTIMATION

The following conclusions serve as guidelines for tackling dynamic parameter-estimation problems.

First, we need to consider transformation of kinetic parameters. Here, the parameters should be relatively uncorrelated and scaled so that they do not vary over many orders of magnitude. All but the third and fourth investigators used log transformations of the equilibrium constants and kinetic parameters centered about a mean base temperature. This helps to improve the conditioning, and hence the convergence of the problem.

Second, a number of problems can also be avoided by transforming the model. For the original model, DASSL was used to solve the DAE system. Here, numerous difficulties were encountered especially for stiff sets of parameter values. For the attempted solution these led to frequent convergence failures and poor performance. Most investigators reduced the model to a system of ODE's. This led to fewer convergence problems and more efficient solutions.

Finally, the last three investigators used different starting guesses in solving this problem. The only explanation for this was offered by the fifth investigators who noted that, with the initial guesses in the Dow problem statement, complete conversion of *HA* to *HABM* occurs before any *AB* is formed. This is easily seen in Figure 1. To obtain lower, more reasonable peak values of *HABM*, a starting guess of $K_1/K_2 = 1$ and $K_3/K_1 = 1$ was selected.

The use of different starting points suggests a more general approach that appears in all of the solutions. Investigators one and two used multiple restarts to converge from the original starting point. Moreover, their preliminary runs dealt with simpler optimization problems which may have given better progress than Eq. 23. Although it was not necessary for obtaining a solution, the first investigators began with a likelihood function without heteroscedasticity, while the second kept the statistical parameters constant until he was close to the solution. This suggests that an interactive approach and a great deal of experience may still be required to solve difficult parameter-estimation problems. Also, from the above solutions it appears that efficient and user-friendly software is not yet available that automatically handles all of the difficulties encountered.

ACKNOWLEDGMENT

The authors wish to express their gratitude to G. Agin, L. Kirkby, and M. Marks of the Dow Chemical Company who assisted in the problem formulation. Special thanks go to the five participants whose submitted solutions, helpful comments, and interest made this paper possible. Finally, acknowledgment is made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this work.

NOTATION

a_{ni}, b_{ni}, b_i, c_i = coefficients in ODE solver
E = partitioned matrix in DAE system
E_i = activation energies
e_u = experimental error vector for experiment u
f = righthand side of ODE system
F = righthand side of DAE system
g = algebraic equations; also gradient vector
h = step size for ODE solver
H = Hessian matrix

k = iteration counter
K_i = equilibrium constants
k_i = rate constants
$L(\theta)$ = maximum likelihood function
m = number of measured variables
n = number of experiments
p_i = transformed parameter vector
q_i = function values in ODE solver
R = eigenvector matrix for rotational discrimination
s = search direction for θ
T = temperature
t_u = time at which experiment u was taken
V = covariance matrix
y = model prediction
z = measured variables

Greek Letters

α_i = preexponential factor in rate constants
γ_i = heteroscedasticity parameters
δ = step length
ϵ_u = residual vector for experiment u
ζ = tolerance for rotational discrimination
θ = parameter vector
η = mean of measured variables
λ = eigenvalues of H
μ = Levenberg-Marquardt parameter
ν = transformed search direction
ω = variance coefficient for heteroscedasticity

LITERATURE CITED

Bard, Y. A., *Nonlinear Parameter Estimation*, Academic Press. New York (1974).

Box, M. J., "A New Method of Constrained Optimization and a Comparison with Other Methods." *Computer J.*, 8 (1), 42 (1965).

Byrne, G. D., "Some Software for Solving Ordinary Differential Equations." *Foundations of Computer-Aided Process Design*, Engineering Foundation, New York, 1, 403 (1981).

Carnahan, B., and J. O. Wilkes. "Numerical Solution of Differential Equations—An Overview, *Foundations of Computer-Aided Chemical Process Design*, Engineering Foundation, New York, 1, 225 (1981).

Chan, Y. N., I. Birnbaum, and L. Lapidus, "Solution of Stiff Differential Equations and the Use of Embedding Techniques" *I & EC Fund.* 17, 133, (1978).

Fariss, R. H., and V. H. Law, "An Efficient Computational Technique for Generalized Application of Maximum Likelihood to Improve Correlation of Experimental Data," *Comp. and Chem. Eng.*, 3, 95 (1979).

Fletcher, R., "A Modified Marquardt Subroutine for Nonlinear Least Squares," *AERE Report. No. 6799*, Harwell, England (1971).

Gear, C. W., "Simultaneous Numerical Solution of Differential-Algebraic Equations," *IEEE Trans. on Circuit Theory*, 18 (1), 89 (Jan. 1971).

Gill, P. E., W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London (1981).

Klaus, R., and D. W. T. Rippin "A New Flexible and Easy-to-Use General-Purpose Regression Program," *Comp. and Chem. Eng.*, 105 (1979).

Michelsen, M. L., "An Efficient General-Purpose Method for the Integration of Stiff Ordinary Differential Equations," *AIChE J.*, 22, 594 (1976).

Petzold, L., "Differential/Algebraic Equations Are Not ODE's," *Siam J. Sci. Stat. Comput.*, 3 (3), 367 (1982).

Prokopakis, G. J., and W. D. Seider, "Adaptive Semi-Implicit Runge-Kutta Method for Solution of Stiff Ordinary Differential Equations," *Ind. Eng. Chem. Fund.*, 20, 255 (1981).

Reilly, P. M., et. al., "Guidelines for the Optimal Design of Experiments to Estimate Parameters in First-Order Kinetic Models," *Can. J. Chem. Eng.*, 55, 614, (1977).

Stewart, W. E., and J. Sorensen, "Bayesian Estimation of Common Parameters from Multiresponse Data with Missing Observations," *Technometrics*, 23, 131 (1981).

Manuscript received June 5, 1984, and revision received Jan. 22, 1985.

APPENDIX A: EXPERIMENTAL DATA AND INITIAL CONDITIONS

The initial parameter estimates are:

$$\alpha_1 = 2.0 \times 10^{13}$$
$$E_1 = 2.0 \times 10^{4}$$
$$\alpha_2 = 2.0 \times 10^{13}$$
$$E_2 = 2.0 \times 10^{4}$$
$$\alpha_{-1} = 4.3 \times 10^{15}$$
$$E_{-1} = 2.0 \times 10^{4}$$
$$K_1 = 1.0 \times 10^{-17}$$
$$K_2 = 1.0 \times 10^{-11}$$
$$K_3 = 1.0 \times 10^{-17}$$
$$\gamma_1 = 1.0$$
$$\gamma_2 = 1.0$$
$$\gamma_3 = 1.0$$
$$\gamma_4 = 1.0$$

The initial model conditions in addition to those given in the data sets are:

$$y_5 = 0$$
$$y_6 = 0.0131$$
$$y_7 = 1/2\{-K_2 + [K_2^2 + 4K_2 y_1(0)]^{0.5}\}$$
$$y_8 = y_7$$
$$y_9 = 0$$
$$y_{10} = 0$$

Table A1. Run 1 Concentration vs. Time Data at 40°C

Time h	Concentration (gmol/kg) HA	BM	HABM	AB
0.00	1.7066	8.3200	0.0000	0.0000
0.08	1.6960	8.3065	0.0077	0.0029
0.58	1.6826	8.2954	0.0234	0.0006
1.58	1.6596	8.2730	0.0470	—
2.75	1.6305	8.2437	0.0763	—
3.75	1.6143	8.2277	0.0923	—
4.75	1.5892	8.2026	0.1174	—
5.75	1.5673	8.1781	0.1371	0.0024
8.75	1.5133	8.1265	0.1935	—
13.05	1.4075	8.0167	0.2949	0.0042
21.75	1.2308	7.8440	0.4760	—
28.75	1.0931	7.6977	0.6047	0.0088
46.25	0.7268	7.3134	0.9530	0.0268
52.25	0.5773	7.1495	1.0881	0.0412
76.25	0.2065	6.6123	1.2929	0.2074
106.25	0.0650	6.2309	1.1941	0.4475
124.25	0.0391	6.1220	1.1370	0.5305
147.25	0.0244	6.0084	1.0528	0.6294
172.25	0.0145	5.9193	0.9835	0.7086
196.25	0.0083	5.8556	0.9326	0.7659
219.75	0.0074	5.8037	0.8821	0.8171
240.25	0.0050	5.7680	0.8492	0.8514
274.25	0.0047	5.7222	0.8064	0.8957
292.25	0.0042	5.7021	0.7869	0.9155
316.25	0.0015	5.6722	0.7628	0.9425
340.75	0.0017	5.6593	0.7495	0.9556
364.25	—	5.6351	0.7263	0.9793
386.75	—	5.6176	0.7112	0.9956
412.25	—	5.6131	0.7063	1.0003
442.75	—	5.5991	0.6927	1.0141
460.75	—	5.5939	0.6871	1.0195
483.75	—	5.5905	0.6837	1.0229
507.25	—	5.5736	0.6672	1.0396
553.75	—	5.5558	0.6494	1.0574
580.75	0.0046	5.5631	0.6467	1.0551
651.25	—	5.5472	0.6408	1.0660
673.25	—	5.5516	0.6452	1.0616
842.75	—	5.5465	0.6397	1.0669

Table A2. Run 2 Concentration vs. Time Data at 67°C

Time h	Concentration (gmol/kg) HA	BM	HABM	AB
0.00	1.6497	8.2262	0.0104	0.0017
0.08	1.6400	8.2158	0.0186	0.0028
1.08	1.4068	7.9826	0.2522	0.0026
2.33	1.0895	7.6584	0.5656	0.0095
3.33	0.8389	7.3861	0.7915	0.0312
4.33	0.6485	7.1231	0.9097	0.1036
5.33	0.4604	6.8643	1.0267	0.1745
12.83	0.1068	6.0636	0.9332	0.6216
23.33	0.0322	5.7212	0.7396	0.8896
27.83	0.0171	5.6208	0.6690	0.9751
30.83	0.0298	5.6680	0.6916	0.9402
51.67	0.0049	5.5139	0.5877	1.0692
83.33	0.0030	5.4859	0.5627	1.0957

Table A2. Run 2 Concentration vs. Time Data at 67°C (cont'd)

Time h	Concentration (gmol/kg) HA	BM	HABM	AB
93.33	0.0022	5.4796	0.5580	1.1012
102.42	0.0022	5.4799	0.5587	1.1007
124.83	0.0026	5.4792	0.5572	1.1018
148.08	0.0012	5.4790	0.5602	1.1004
171.83	0.0036	5.4857	0.5617	1.0963
197.33	0.0027	5.4843	0.5617	1.0970
228.33	0.0024	5.4862	0.5646	1.0946
270.33	0.0026	5.4858	0.5638	1.0952
293.33	0.0032	5.5069	0.5441	1.0945

Table A3. Run 3 Concentration vs. Time Data at 100°C

Time h	Concentration (gmol/kg) HA	BM	HABM	AB
0.00	1.5608	8.3546	0.0082	0.0086
0.08	1.5316	8.3325	0.0445	0.0015
0.42	0.7016	7.3364	0.7088	0.1674
0.75	0.3763	6.7519	0.7745	0.4268
1.17	0.2229	6.3849	0.7147	0.6402
1.50	0.1793	6.2899	0.7065	0.6918
2.00	0.1336	6.1287	0.6363	0.8075
2.50	0.0894	5.9985	0.5953	0.8931
3.00	0.0752	5.9541	0.5785	0.9237
3.50	0.0626	5.9186	0.5686	0.9464
4.00	0.0518	5.8818	0.5542	0.9720
4.50	0.0598	5.8878	0.5434	0.9744
5.00	0.0320	5.8285	0.5397	1.0059
5.50	0.0228	5.8086	0.5382	1.0166
6.50	0.0225	5.8048	0.5354	1.0199
7.00	0.0180	5.7944	0.5336	1.0260
7.50	0.0155	5.7870	0.5320	1.0305
8.00	0.0135	5.7833	0.5315	1.0326
8.50	0.0118	5.7775	0.5295	1.0365
9.00	0.0110	5.7774	0.5302	1.0362
9.50	0.0096	5.7735	0.5303	1.0381
10.00	0.0085	5.7706	0.5288	1.0403
10.50	0.0081	5.7704	0.5294	1.0401
11.00	0.0200	5.7912	0.5268	1.0310
11.50	0.0183	5.7875	0.5261	1.0332
12.50	0.0060	5.7644	0.5276	1.0440
13.50	0.0192	5.7874	0.5246	1.0340
14.50	0.0059	5.7641	0.5271	1.0444
16.50	0.0115	5.7680	0.5202	1.0459
21.75	0.0059	5.7596	0.5234	1.0485
29.50	0.0049	5.7515	0.5173	1.0556
53.00	0.0039	5.7468	0.5138	1.0597

APPENDIX B: INITIAL AND FINAL PARAMETER ESTIMATES WITH 95% CONFIDENCE INTERVALS

TABLE B1. INVESTIGATOR (1), CARACOTSIOS, STEWART, AND SORENSEN, ESTIMATES

Parameter	Initial Estimate	Final Estimate
p_1	1.1938	0.7965 ± 0.0615
p_2	1.1938	1.1658 ± 0.0558
p_3	1,837.0	12,392 ± 50
p_4	13,393.0	19,885 ± 10
p_5	8,666.0	8,200**
p_6	13,393.0	20,000**
p_7	10,070.0	9,493 ± 211
p_8	10,070.0	9,457 ± 233
p_9	0	−3,230 ± 579
p_{10}	*	$(1.08 \pm 0.32)\cdot10^{-3}$
p_{11}	*	$(-2.39 \pm 1.95)\cdot10^{-4}$
p_{12}	*	$(5.72 \pm 2.06)\cdot10^{-4}$
p_{13}	*	0**
p_{14}	*	0.190 ± 0.118

*Not specified.
**95% confidence intervals are not estimated.

TABLE B2. INVESTIGATOR (2), FARISS, ESTIMATES

Parameter	Initial Estimate	Final Estimate
α_1	2.0×10^{13}	1.3708×10^{12}
E_1	2.0×10^{4}	1.8476×10^{4}
α_2	2.0×10^{13}	5.2282×10^{12}
E_2	2.0×10^{4}	1.9075×10^{4}
α_{-1}	4.3×10^{15}	1.6215×10^{20}
E_{-1}	2.0×10^{4}	2.6046×10^{4}
K_1	1.0×10^{-17}	2.575×10^{-16}
K_2	1.0×10^{-11}	4.876×10^{-14}
K_3	1.0×10^{-17}	1.7884×10^{-16}
γ_1	1.0	0.8149
γ_2	1.0	7.2271
γ_3	1.0	−0.2972
γ_4	1.0	0.1352
ω_1	0.065	0.06393
ω_2	0.143	$4.792\cdot10^{-3}$
ω_3	0.057	0.03105
ω_4	0.0525	0.02473

*95% confidence intervals not reported.

TABLE B3. INVESTIGATOR (3), AHN, ESTIMATES

Parameter	Initial Estimate	Final Estimate
$\ln\alpha_1$	30.627	28.533 ± 1.264
E_1	$2.0\cdot10^{4}$	$(1.8835 \pm 0.0943)\cdot10^{4}$
$\ln\alpha_2$	30.627	27.313 ± 0.391
E_2	$2.0\cdot10^{4}$	$(1.7875 \pm 0.0254)\cdot10^{4}$
$\ln\alpha_{-1}$	35.997	44.594 ± 10.173
E_{-1}	$2.0\cdot10^{4}$	$(2.5150 \pm 0.0473)\cdot10^{4}$
$\ln K_1$	−39.144	−38.667 ± 5.822
$\ln K_2$	−34.388	−33.779 ± 9.166
$\ln K_3$	−39.144	−39.019 ± 5.821

TABLE B4. INVESTIGATOR (4), RIPPIN, ESTIMATES

Parameter	Initial Estimate	Final Estimate
α_1	2.0×10^{13}	$(1.0955 \pm 0.00186)\cdot10^{12}$
E_1	2.0×10^{4}	$(1.8255 \pm 0.00036)\cdot10^{4}$
α_2	2.0×10^{13}	$(1.8165 \pm 0.00515)\cdot10^{12}$
E_2	2.0×10^{4}	$(1.8412 \pm 0.00024)\cdot10^{4}$
α_{-1}	4.3×10^{15}	$(2.8397 \pm 0.00897)\cdot10^{16}$
E_{-1}	2.0×10^{4}	$(2.1901 \pm 0.00076)\cdot10^{4}$
K_1	1.0×10^{-17}	$(1.2263 \pm 0.00266)\cdot10^{-18}$
K_2	1.0×10^{-17}	$(2.0404 \pm 0.0132)\cdot10^{-17}$
K_3	1.0×10^{-17}	$(8.4829 \pm 0.0196)\cdot10^{-19}$

TABLE B5. INVESTIGATOR (5), SARUP, MICHELSEN, AND VILLADSEN, ESTIMATES

Parameter	Initial Estimate	Final Estimate
$\ln k_1^o$	1.022	0.78 ± 0.164
E_1	20,000	17,837 ± 1000
$\ln k_2^o$	1.022	1.01 ± 0.033
E_2	20,000	18,850 ± 222.3
$\ln k_{-1}^o$	6.393	5.0 ± 0.82
E_{-1}	20,000	25,210 ± 553.9
$\ln(K_1/K_3)$	0	0.3570 ± 0.0181
$\ln(K_3/K_2)$	0	−2.700 ± 0.5

T_o = 340 K

TABLE B6. THIS STUDY, DAMIANO, MC ESTIMATES

Parameter	Initial Estimate	Final Estimate
α_1	2.0×10^{13}	1.8745×10^{13}
E_1	2.0×10^{4}	1.960×10^{4}
α_2	2.0×10^{13}	1.8911×10^{13}
E_2	2.0×10^{4}	2.0046×10^{4}
α_{-1}	4.3×10^{15}	3.8003×10^{15}
E_{-1}	2.0×10^{4}	2.0110×10^{4}
K_1	1.0×10^{-17}	1.0296×10^{-17}
K_2	1.0×10^{-11}	1.3969×10^{-11}
K_3	1.0×10^{-17}	6.3845×10^{-18}

*95% confidence intervals not estimated.

TABLE B7. THIS STUDY, DAMIANO, QN ESTIMATES

Parameter	Initial Estimate	Final Estimate
α_1	2.0×10^{13}	7.8961×10^{13}
E_1	2.0×10^{4}	1.7486×10^{4}
α_2	2.0×10^{13}	7.1852×10^{11}
E_2	2.0×10^{4}	1.7958×10^{4}
α_{-1}	4.3×10^{15}	5.1796×10^{15}
E_{-1}	2.0×10^{4}	2.2052×10^{4}
K_1	1.0×10^{-17}	1.1944×10^{-17}
K_2	1.0×10^{-11}	1.9016×10^{-11}
K_3	1.0×10^{-17}	7.8221×10^{-18}

*95% confidence intervals not estimated.