Journal of Research of the National Bureau of Standards
Volume 90, Number 6, November–December 1985

# Fitting First Order Kinetic Models Quickly and Easily

**Douglas M. Bates**
University of Wisconsin-Madison, Madison, WI 53705

and

**Donald G. Watts**
Queen's University at Kingston, ON, Canada

Accepted: June 24, 1985

Kinetic models described by systems of linear differential equations can be fitted to data quickly and easily by taking advantage of the special properties of such systems. The estimation situation can be greatly improved when multiresponse data are available, since one can then automatically determine starting values and better *discriminate between rival models.*

Key words: compartment model; determinant criterion; multiresponse estimation.

## 1. Introduction

In this article we summarize the work of a series of papers [1–3][1] in which we deal with fitting first order *kinetic models to uniresponse and multiresponse data.*

**About the Authors, Paper:** Douglas M. Bates is with the Department of Statistics at the University of Wisconsin. Donald G. Watts is a member of the Department of Mathematics and Statistics, Queen's University at Kingston. The work reported on was supported by the National Science Foundation in the United States and the Natural Sciences and Engineering Research Council of Canada.

[1] Numbers in brackets indicate literature references.

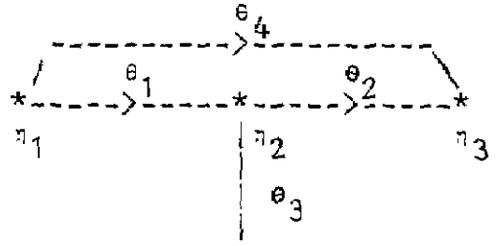
We consider systems in which the expected responses at $K$ points in the system, $\eta(t)=(\eta_1(t), \eta_2(t), \ldots \eta_K(t))'$, are described by the system of linear differential equations

$$\partial\eta/\partial t=\dot{\eta}(t)=A\,\eta(t)+\iota(t) \tag{1.1}$$

where $A$ is a $K\times K$ system *transfer matrix* depending on *rate constants* $\theta_1, \theta_2, \ldots$, and $\iota(t)$ is a vector input function to the system. We assume further that there are $K$ initial conditions $\eta o=(\eta o_1, \eta o_2, \ldots, \eta o_K)'$, some possibly unknown, and that $\tau=t-\theta o$, where $\theta o$ is a (possibly unknown) time delay. All the unknown parameters are gathered into a $P\times 1$ parameter vector $\theta$.

**Example: Oil Shale.1**

As an example of a chemical system described by a set of linear differential equations, we cite the pyrolysis of oil shale in which the model, fitted by Ziegel and Gorman [4] has the system diagram




In this system, $\eta_1$ denotes kerogen, $\eta_2$ bitumen, and $\eta_3$ oil. The model implies that kerogen decomposes to bitumen with rate constant $\theta_1$, and to oil with rate constant $\theta_4$, and bitumen produces oil with rate constant $\theta_2$, and unmeasured by-products with rate constant $\theta_3$.

## 2. Expectation Functions and Derivatives With Respect to the Parameters for First Order Kinetic Systems

Several methods are used to estimate the parameters in first order kinetic models. The most obvious method is to solve the system of differential equations corresponding to the particular compartment model and use the resulting expectation function in a standard nonlinear estimation program. A second approach is to fit a general sum of exponentials model by "peeling" [5]. A third approach is to use a standard nonlinear estimation program, using numerical integration to solve the equations. A superior approach, proposed by Jennrich and Bright [6], is to obtain the general solution to the system of equations by calculating values for the model function $\eta(t)$ and its derivatives directly, given values of $\theta$ and $t$ and $\iota(t)$.

### 2.1 The General Solution

The solution to a linear system of differential equations can be expressed in terms of convolutions using the matrix exponential [7]. The solution is

$$\eta(t)=e^{At}\eta o+e^{At}*\iota(t) \tag{2.1}$$

where the '*' denotes convolution,

$$e^{At}*\iota(t)=\int_0^t e^{A(t-\xi)}\iota(\xi)\,d\xi. \tag{2.2}$$

The integral of the vector function is evaluated componentwise. Computational methods for evaluating the convolution integral are given in Moler and Van Loan [8], when $A$ is diagonalizable, and in Bavely and Stewart [9], when $A$ is nondiagonalizable.

### 2.2 Derivatives of the Expectation Function

To use a Gauss-Newton procedure to estimate the parameters we need derivatives with respect to the parameters. As shown by Jennrich and Bright [6], a great advantage to compartment models is that the derivatives can be evaluated in the same fashion as the model function itself. Instead of differentiating $\eta(t)$ directly as in Jennrich and Bright, however, we differentiate eq (1.1) and solve the resulting linear system of differential equations. This idea was discussed in another context in Smith [10] and was used by Kalbfleisch et al. [11].

To simplify notation, we use a subscript $p$ to denote differentiation with respect to the parameter $\theta_p$ so

$$\eta_p(t)=\frac{\partial\eta(t)}{\partial\theta_p}$$

for $p=1, 2, \ldots P$. The derivative of (1.1) with respect to $\theta_p$ is then

$$\dot{\eta}_p(t)=A\eta_p(t)+A_p\eta(t)+\iota_p(t)$$

for which the solution is

$$\eta_p(t)=e^{At}\eta_p(0)+e^{At}*[A_p\eta(t)+\iota_p(t)] \tag{2.3}$$

$$=e^{At}\eta_p(0)+e^{At}*\iota_p(t)+e^{At}*A_pe^{At}\eta o+e^{At}*A_pe^{At}*\iota(t).$$

Dead time, $\theta_0$, can be incorporated by modifying (1.1) to

$$\dot{\eta}(\tau)=A\eta(\tau)+\iota$$
$$\eta(0)=\eta o$$

where

$$\tau=\begin{cases} t-\theta o & t>\theta o \\ 0 & t\leqslant\theta o\,. \end{cases}$$

If $\theta o$ is known, we simply replace $t$ by $\tau$ in eqs 1.1, 2.1, and 2.3, but if $\theta o$ is unknown and depends on a parameter, the expression for the derivatives is extended to

$$\eta_p(\tau)=e^{A\tau}\eta_p(0)+e^{A\tau}*[A_p\eta(\tau)+\iota_p]$$

$$+\tau_p[A\eta(\tau)+\iota].$$

It is easy to evaluate $A_p$, $\eta_p(0)=\partial\eta o/\partial\theta_p$, and $\tau_p$ since they are constants and, in fact, usually two of the three are zero. Also, for any $t<\theta o$, $\tau_p$ is zero for all $p=1, \ldots, P$. Note that the method can be extended to higher order derivatives.

**Specifying the Model**

A simple unambiguous computer notation can be used to specify first order kinetic models in a *parameter table* consisting of three columns, the first column giving the parameter number. For a rate constant, the second column entry gives its source and the third column entry its sink, a sink with compartment number 0 denoting elimination. For initial conditions, $\eta o$, the second column entry is the number of the component in $\eta o$ and the third column entry is $-1$. For a step input, $\iota$, the second column entry is the number of the component in $\iota$ and the third column entry is $-2$. Dead time is coded as 0 in column 2 and 0 in column 3.

**Example: Oil Shale.2**

The oil shale parameter table is presented in annotated form. It is to be noted that a single parameter may represent more than one rate constant.

**Table 1.** Oil shale parameters.

| Parameter number (type) | Column 2 | | Column 3 | |
|---|---|---|---|---|
| 1 (rate constant) | 1 | (source) | 2 | (sink) |
| 2 (rate constant) | 2 | (source) | 3 | (sink) |
| 3 (rate constant) | 2 | (source) | 0 | (elimination) |
| 4 (rate constant) | 1 | (source) | 3 | (sink) |
| 5 (dead time) | 0 | | 0 | |

## 3. Multiresponse Estimation

In the multiresponse situation when the errors have unknown variances and covariances but are assumed to be temporally uncorrelated, the appropriate criterion derived via a likelihood or Bayesian approach is to minimize the $M\times M$ determinant [12].

$$|V(\theta)|=|Z'Z| \tag{3.1}$$

In eq (3.1), $Z=Y-H$ is the $N\times M$ matrix of residuals $\{z_{nk}\}=\{y_k(t_n)-\eta_k(t_n)\}$, $k=1, \ ,2 \ , \ldots, M$, $n=1, \ 2, \ldots, N$. The *expected responses* $\eta$ are assumed to depend on $P$ parameters $\theta$: except where explicitly required, we suppress the dependence on $\theta$. For first order kinetic systems with all responses measured, $M=K$.

To determine the minimum of $|Z'Z|$, it is advantageous to calculate the gradient and Hessian and exploit Gauss-Newton optimization techniques. Efficient numerical procedures for computing the gradient and approximate Hessian are given in [2], and an algorithm which performs the calculations in [1].

Alternative expressions for the components of the gradient $\gamma$ and the Hessian $\Gamma$ are

$$\gamma_p=\partial|V|/\partial\theta_p=2|V|tr[V^{-1}Z'Z_p], \quad p=1,2,\ldots,P, \tag{3.2}$$

and

$$\Gamma_{pq}=\partial^2|V|/\partial\theta_p\partial\theta_q= \tag{3.3}$$

$$=\gamma_p\gamma_q+2|V|tr[V^{-1}Z'Z_qV^{-1}Z'Z_p] +2|V|tr[V^{-1}Z'Z_qV^{-1}Z_p'Z]$$

$$+2|V|tr[V^{-1}Z_p'Z_q]+2|V|tr[V^{-1}Z'Z_{pq}], \quad p,q=1,2,\ldots,P.$$

The second derivative terms $Z_{pq}$ in eq (3.3) are ignored to produce an approximate Hessian.

## 4. Practical Aspects

**Linear Constraints**

Sometimes the data matrix $Y$ involves dependencies as a result of imputation of responses or mass-balance calculations. If these dependencies also occur in the expected responses, then important modifications to the multiresponse estimation procedure must be made so as to avoid convergence to spurious optima [13,14]. It is therefore necessary to examine the residual matrix $Z(\theta)$ for singularities, which can be done by arranging the rounding units in the columns of $Y$ to be approximately equal and taking a singular value decomposition of $Z$ [15]. As explained there, singular values on the order of the rounding unit indicate singularity and should prompt the analyst to search for constraints in the data. Such examinations should be done at the beginning of the analysis using the initial parameter values and at the end of the analysis using the converged values. To aid convergence, logarithms of parameters are used during estimation.

Linear constraints can be dealt with easily by combining the linear constraint vectors into a matrix, performing a $QR$ decomposition of that matrix, and letting the rotation matrix $W$ be the columns of $Q$ which are orthogonal to the constraint vectors. We then simply minimize $|(ZW)'(ZW)|$, where $ZW=YW-HW$. Clearly, the gradient and Hessian of this determinant

are obtained from eqs 3.2 and 3.3 by replacing $Z$ by $ZW$ and $Z_p$ by $Z_pW$.

### Constraints on the Number of Parameters, Responses and Observations

The determinant criterion implies two constraints on the number of observations [2,3]. First, $N$ must be at least equal to $M$ since otherwise the determinant is identically zero. Second, $N$ must exceed $P$ otherwise the criterion can be made zero by fitting any one response perfectly, which can generate up to $M$ distinct minima. Thus the residual matrix has effectively $N-P$ degree of freedom. It may seem that there should be more degrees of freedom since there are $NM$ separate observations, but the criterion can be locally controlled by any one response so the effective number of observations is $N$ rather than $NM$.

### Starting Values

An important part of fitting nonlinear models is determining good starting values. For uniresponse data, an effective method is to use peeling in which we plot the logarithm of the response versus time and fit a straight line to the segment at large $t$ values. The slope of the line gives an estimate of the smallest eigenvalue of the $A$ matrix. Using the fitted line to generate residuals and plotting the logarithm of the residuals versus $t$ should again reveal a straight line portion at large $t$ values, so the process is repeated, thereby obtaining estimates for the eigenvalues. As mentioned in section 2, this process is often used for parameter estimation, but we do not recommend it.

In the case of multiresponse data for first order kinetics, the problem is easily solved using linear least squares by exploiting the linear relation between the rates and the responses! As noted in [3], if we could measure the rates $\dot{\boldsymbol{\gamma}}$ and the responses $\boldsymbol{\gamma}$ at a particular time $\tau$, then using $\dot{\boldsymbol{\gamma}}(\tau)=A\boldsymbol{\gamma}(\tau)$ produces a linear relation between the "dependent" variable $\mathbf{y}=\dot{\boldsymbol{\gamma}}$ and the "independent" variables $x_p = A_p\boldsymbol{\gamma}$ in the form $\mathbf{y} = X\boldsymbol{\theta}$. We can thus solve for $\hat{\boldsymbol{\theta}}$ by using linear least squares. A simple procedure for obtaining starting values, then, is to use approximate rates from finite differences of the responses at successive time points and $x_p$ values from the corresponding averages. Alternatively, one could smooth the data for each response by fitting splines so as to obtain better rate and response values, and then use these in a linear least squares routine.

### Example: α-pinene.1

Data on the thermal isomerization of α-pinene at 189.5° were reported by Fuguitt and Hawkins [16], and a thorough multi-response analysis was presented in [13]. The fitted model was described by the system

$$\dot{\boldsymbol{\gamma}} = \begin{bmatrix}\dot{\gamma}_1\\ \dot{\gamma}_2\\ \dot{\gamma}_3\\ \dot{\gamma}_4\\ \dot{\gamma}_5\end{bmatrix} = \begin{bmatrix}-\theta_1\gamma_1-\theta_2\gamma_1 & & & & \\ \theta_1\gamma_1 & & & & \\ & \theta_2\gamma_1-\theta_3\gamma_3-\theta_4\gamma_3+\theta_5\gamma_5 & & & \\ & & \theta_3\gamma_3 & & \\ & & & \theta_4\gamma_3-\theta_5\gamma_5 & \end{bmatrix} = A\boldsymbol{\gamma}$$

which can also be written $\dot{\boldsymbol{\gamma}}=X\boldsymbol{\theta}$, where

$$X = \begin{bmatrix}-\gamma_1 & -\gamma_1 & 0 & 0 & 0\\ \gamma_1 & 0 & 0 & 0 & 0\\ 0_1 & \gamma_3 & -\gamma_3 & -\gamma_5 & \gamma\\ 0 & 0 & \gamma_3 & 0 & 0\\ 0 & 0 & 0 & \gamma & -\gamma_5\end{bmatrix}$$

Substituting estimated rates for each time and joining them into a single vector $\mathbf{y}$, and calculating $X$ matrices for each time and joining them into a single $X$ matrix, allows us to use linear regression to estimate starting values for $\boldsymbol{\theta}$. The starting values obtained are listed in column 2, table 2. Note their closeness to the converged values, column 4.

**Table 2.** Estimation for α-pinene at 189.5°.

| Parameter | Start | Box et al. | Bates and Watts Full Model | Bates and Watts Reduced Model |
|---|---|---|---|---|
| 1 | 5.84 | 5.95 | 5.94 | 5.95 |
| 2 | 2.65 | 2.85 | 2.84 | 2.82 |
| 3 | 1.63 | 0.50 | 0.45 | - |
| 4 | 27.77 | 31.5 | 31.21 | 30.75 |
| 5 | 4.61 | 5.89 | 5.79 | 5.72 |
| Determinant | 600 | | 28.4 | 29.0 |

Parameter estimates are listed in table 2 in minutes$^{-1}\times 10^{-5}$; that is a table value of 5.84 is actually a rate constant of $5.84\times 10^{-5}$ minutes$^{-1}$.

When some of the responses are not measured, it is still possible to use approximate rates provided other information, such as a mass-balance, is substituted.

## 5. Two Examples

### 5.1 Oil Shale

The model and data presented by Ziegel and Gorman [4] were fitted using the procedures described here. In this problem the concentration of $\eta_1$ was not measured, which introduces some complexity in determining start-

ing values. Second, the elimination compartment, corresponding to coal and gas, was not measured. Third, there was a time delay caused by the shale having to reach the reaction temperature. The observed responses are $y_2$ and $y_3$, measured in percent of the initial kerogen $\eta_1$, so $K=3$, $M=2$, with $N=14$.

To determine starting values in this case, we plotted the data and obtained a rough estimate of $\theta o=6$ min. Estimating the initial slopes of $\eta_2$ and $\eta_3$ from the graph gave values of 0.029 $\text{min}^{-1}$ for $\theta_1$ and 0.013 $\text{min}^{-1}$ for $\theta_4$. This delay estimate and the rate constants were used to obtain starting estimates of $\theta_2$ and $\theta_3$ as shown in column 2, table 3, following the procedure of section 4. The final results, together with those of Ziegel and Gorman, are shown in columns 3 and 4 of table 3.

Table 3. Parameter estimates for oil shale data.

| Parameter | Start | Bates-Watts | Ziegel-Gorman |
|---|---|---|---|
| 1 | 0.029 | 0.0172 | 0.0173 ($k_1$) |
| 2 | 0.012 | 0.0090 | 0.0092 ($k_2f_2$) |
| 3 | 0.028 | 0.0205 | 0.0201 ($k_2(t-f_2)$) |
| 4 | 0.013 | 0.0104 | 0.0104 ($k_3$) |
| $\theta o$ | 6.0 | 7.8 | 7.7 |
| Determinant | | | |
| | 67980 | 429 | |

### 5.2 α-pinene

In their analysis of the α-pinene data, Box et al. [13] noted that response 4 was imputed using $y_4=0.03(100-y_1)$, and that the data set was subject to a mass-balance constraint, $y_1+y_2+y_3+y_4+y_5=100$. To avoid convergence to spurious optimal parameter values, they recommended that these data dependencies be taken into account by using observation vectors consisting of linear combinations of $y_1$, $y_2$, $y_3$, $y_4$ and $y_5$ which are orthogonal to the space defined by the vectors $(0.03,\ 0,\ 0,\ 1,\ 0)'$ and $(1,\ 1,\ 1,\ 1,\ 1)'$. We therefore treat $y_4$ as an unmeasured component for estimation purposes, and use linear combinations of the responses which are orthogonal to the vectors $\mathbf{a}_1'=(0,\ 0,\ 0,\ 1,\ 0)$ and $\mathbf{a}_2'=(1,\ 1,\ 1,\ 1,\ 1)$. The rotation matrix $M$ and the modified responses can therefore be determined by performing a $QR$ decomposition on the matrix $(\mathbf{a}_1,\ \mathbf{a}_2)$ and using the last 3 columns of $Q$ coupled with all the responses. In this case, $K=5$, $M=3$, with $N=8$.

Approximate 95% confidence limits for $\ln\theta_3$ were very wide, suggesting that $\theta_3$ was badly estimated and could be zero. We therefore fitted a reduced model in which there was no path from $\eta_3$ to $\eta_4$, see column 5. The change in the determinant is 0.6 on 1 degree of freedom, which, when compared with the scaling factor $s^2=28.4/3=9.46$ on 3 degrees of freedom, is clearly small, verifying that the reduced model is adequate for this data set.

To further substantiate the adequacy of the reduced model, we fitted both models to a second set of data taken 204.5° [16]. The results of this fitting procedure are presented in table 4. The reduced model appears to be adequate for both data sets.

Table 4. Estimation for α-pinene at 204.5°.

| Parameter | Start | Full Model | Reduced Model |
|---|---|---|---|
| 1 | 23.0 | 22.6 | 22.6 |
| 2 | 13.2 | 12.6 | 12.6 |
| 3 | 8.3 | 0.02 | - |
| 4 | 76.9 | 72.8 | 72.8 |
| 5 | 16.0 | 15.6 | 15.6 |
| Determinant | | | |
| | 116 | 0.55 | 0.55 |

## 6. Conclusions

Several advantages of the direct multiresponse estimation approach for systems of differential equations are apparent. First, the model can be specified directly from the network diagram. Second, there is no need to obtain the analytic solution to the differential equations describing the reactions. Third, there is no need to code the model functions in a nonlinear estimation routine. Fourth, the bothersome and error-prone step of obtaining and coding derivatives of the expected responses with respect to the parameters is eliminated. Fifth, excellent starting values can be determined automatically.

## References


[1] Bates, D. M., and D. G. Watts, A multi-response Gauss-Newton algorithm, Commun. Statist.—Simul. Comput. **B(13)**, 705–715 (1984).

[2] Bates, D. M., and D. G. Watts, A generalized Gauss-Newton procedure for multi-response parameter estimation, SIAM Journal on Scientific and Statistical Computing, to appear.

[3] Bates, D. M., and D. G. Watts, Multiresponse estimation with special application to systems of linear differential equations, Technometrics **27**, 190–201 (1985).

[4] Ziegel, E. R., and J. W. Gorman, Kinetic modelling with multiresponse data, Technometrics **22**, 139–151 (1980).

[5] Anderson, D. H., Compartmental modeling and tracer kinetics, Springer-Verlag (1983).

[6] Jennrich, R. I., and P. R. Bright, Fitting systems of linear differential equations using computer generated exact derivatives, Technometrics **18**, 385–392 (1976).

[7] Noble, B., Applied linear algebra, Prentice Hall (1969).

[8] Moler, C., and C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, SIAM Review **20** (1978).

[9] Bavely, C. A., and G. W. Stewart, An algorithm for computing reducing subspaces by block diagonalization, SIAM Journal of Numerical Analysis 16, 359–367 (1979).

[10] Smith, W. R., *Parameter estimation in nonlinear models of biological systems, Fisheries and Marine Services Technical Report 889, Fisheries and Environment Canada (1979).*

[11] Kalbfleisch, J. D.; J. F. Lawless and W. M. Vollmer, Estimation in Markov models from aggregate data, *Biometries* 39, 907–919 (1983).

[12] Box, G. E. P., and N. R. Draper, The Bayesian estimation of common parameters from several responses, Biometrika 52, 355–365 (1965).

[13] Box, G. E. P.; W. G. Hunter, J. F. MacGregor, and J. Erjavec, Some problems associated with the analysis of multiresponse data, Technometrics 15, 33–51 (1973).

[14] McLean, D. D.; D. J. Pritchard, D. W. Bacon, and J. Downie, Singularities in multiresponse modelling, Technometrics 21, 291–298 (1979).

[15] Dongarra, J. J.; J. R. Bunch, C. B. Moler, and G. W. Stewart, Linpack User's Guide, Chap. 11, Philadelphia: S.I.A.M. (1979).

[16] Fuguitt, R. E., and J. E. Hawkins, Rate of thermal isomerization of α-pinene in the liquid phase, J. Amer. Chem. Soc. 69, 319–312 (1947).

# DISCUSSION

## of the Bates-Watts paper, Multiresponse Estimation With Special Applications to First Order Kinetics.

**Michael Frenklach**

Department of Materials Science and Engineering
Pennsylvania State University [1]

The authors presented an interesting approach to parameter estimation for first order kinetic systems. The method is user oriented and particularly suited for computer implementation as a "canned" program. Indeed, present chemical kinetic codes input reaction mechanism in a natural chemical language, that is, specifying reactions (usually in unformatted READ routines) as they are conventionally written on the paper. This information is automatically converted to a so-called reaction matrix and, based on it, to differential equations describing the kinetics of reaction species. The reaction matrix, which contains all the stoichiometry of the system, can conveniently provide the required input information.

Another important feature, from the user's point of view, is that the presented method is applicable to multiresponse data. It should be realized that modern problems of interest to chemical kinetics get tougher, as for example, formation of pollutants in hydrocarbon combustion. The experimental answer to the growing complexity of the systems is the employment of multiple diagnostics for simultaneous monitoring of various process variables. However, interpretation of the experimental results cannot be fully realized without reliable and convenient multiresponse methods.

The following are some of my thoughts on the needs in this area:

1) Often, kineticists exhibit a philosophical resistance to a multiparameter approach to experimation for automatic coding of the method of Bates and Watts.

[1] Michael Frenklach's contribution to the subject stems from work performed in the Department of Chemical Engineering, Louisiana State University.