Parameter Estimation from Multiresponse Data

Warren E. Stewart, Mike Caracotsios, and Jan P. Sørensen
Dept. of Chemical Engineering, University of Wisconsin, Madison, WI 53706

Multiresponse data are familiar outputs of experiments and processes involving multicomponent mixtures, multiple streams or multiple methods of observation. The resulting arrays of data exhibit various structures, including rectangular (no missing values), block-rectangular, and irregular (missing values with no simple pattern). Methods for modeling such data are described, emphasizing parameter estimation strategy and available software. Experiences in multiresponse modeling are then reviewed for several chemical engineering problems.

Introduction

Chemical engineers deal regularly with multiresponse data in experimental studies of chemical reactions, mixtures and processes. This review surveys modern methods and software for fitting models to such data and optimally estimating their significant parameters. The preferred methods are based on a famous theorem by Bayes (1763) on inference of causes from observed effects.

The method of least squares is not well suited for dealing with multiresponse data, though a weighted version is applicable when the various responses have independent, unbiased Normal distributions of error. Nonlinear models with multiple responses were treated under these assumptions by Gauss (1809) and Deming (1943). Aitken (1935) generalized linear weighted least squares to multiple responses with a specified error covariance matrix; his method was extended to nonlinear parameter estimation by Bard (1968, 1974) and by Klaus and Rippin (1979). Software programs for nonlinear least squares are widely available now and are adaptable by expert users to multiresponse estimation with a given covariance matrix. This approach to multiresponse estimation is rather subjective, however, since the covariance matrix is seldom precisely known.

Bayes' theorem (Bayes, 1763; Box and Tiao, 1973) permits estimation of the error covariance matrix Σ from a multiresponse data set, along with the parameter vector θ of a predictive model. An objective version of the theorem is available from the work of Jeffreys (1961). In this article, we summarize the resulting estimation formulas for problems with Normally distributed errors and describe some experiences in their use.

Jeffreys (1961) made a major advance in Bayesian estimation theory by giving an objective prior probability density for the parameters in any sufficiently differentiable model. He gave a modified formula for use when some parameters are unrelated *a priori*. The naturalness of Jeffreys' treatment led to renewed interest in the Bayes approach.

Box and Draper (1965) took another major step, by deriving a posterior density function $p(\theta|y)$ for estimation of a parameter vector θ from a full matrix y of multiresponse observations. The errors in the observations were assumed to be Normally distributed, with an unknown dense covariance matrix Σ. Box and Draper (1972) gave a corresponding density function for treating a full or partial data matrix y of independent blocks of responses, and applied this function to sequential design of experiments. These seminal articles and the book by Box and Tiao (1973) have inspired many further developments.

The posterior density function found by Box and Draper (1965) is a power of the determinant $|\mathbf{v}(\theta)|$, whose elements appear in Eq. 3. They used contour plots of this function to determine approximate 95% probability regions for a two-parameter model. This technique is very useful for studying models with few parameters and is convenient with modern graphic software.

For larger problems, early workers minimized $|\mathbf{v}(\theta)|$ by search techniques, which proved to be slow and gave only point estimates of θ. Newton-like algorithms for minimization of $|\mathbf{v}(\theta)|$ and for interval estimation of θ were given by Stewart and Sørensen (1976, 1981) and by Bates and Watts (1985, 1987). Related algorithms for likelihood-based estimation were developed by Bard (1968, 1974) and were extended by Klaus and Rippin (1979) and Steiner et al. (1986).

Several generalizations of the problem considered by Box and Draper (1965) have been treated in the literature, and some interesting applications have appeared. These results and some extensions are described below, along with available software for further applications.

Problem Types

The general problem considered here is the estimation of

the parameter vector θ and covariance matrix Σ in a model:

$$y_{ui} = f_i(x_u,\theta) + \epsilon_{ui} \quad (u = 1,\ldots,n;\ i = 1,\ldots,m) \tag{1}$$

of a multiresponse data array $y = \{y_{ui}\}$. Each integer u from 1 to n will denote an independent event, in which $m_u \le m$ responses are observed. The function $f_i(x_u,\theta)$ is an expectation model for response i at the experimental design point x_u. By "estimation" we mean not just the computation of most probable values of θ and Σ, but the determination of their joint posterior probability distribution and corresponding interval estimates.

The errors ϵ_{ui} are modeled by an m-dimensional Normal distribution, with expectations $E(\epsilon_{ui}) = 0$ and covariances $\Sigma_{ij} := E(\epsilon_{ui}\epsilon_{uj})$. Use of this distribution with Eq. 1 yields a density function $p(y\,|\,\theta,\Sigma)$ for future observations at the design points $\{x_1,\ldots,x_n\}$. Once data y are available, the likelihood function $l(\theta,\Sigma\,|\,y)$ is constructed in the manner of Fisher (1922) as the function $p(y\,|\,\theta,\Sigma)$ with y now given, whereas θ and Σ are free to vary. Multiplication of the likelihood by a suitable prior density $p(\theta,\Sigma)$ in accordance with Bayes' theorem gives the posterior density function $p(\theta,\Sigma\,|\,y)$, which contains all current information on θ and Σ. These constructions are summarized below for several problem types.

Type 1: full y and full unknown Σ

The estimation of θ and a full covariance matrix Σ from a full data array y was treated by Box and Draper (1965). Here every experiment has a full set of m responses, as shown in Table 1. The likelihood function

$$l(\theta,\Sigma\,|\,y) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}\mathbf{v}(\theta)]\right\} \tag{2}$$

is meaningful over the positive definite range of Σ. The symmetric matrix $\mathbf{v}(\theta)$ with elements

$$v_{ij}(\theta) = \sum_{u=1}^{n} [y_{ui} - f_i(x_u,\theta)][y_{uj} - f_j(x_u,\theta)] \tag{3}$$

was assumed to be nonsingular, and can be made so by restriction of the set of responses. The Jeffreys noninformative prior for Σ was found to be

Table 1. Full y and Σ Structures

	y_{u1}	y_{u2}	y_{u3}	y_{u4}
$u=1$	+	+	+	+
$u=2$	+	+	+	+
$u=3$	+	+	+	+
$u=4$	+	+	+	+
$u=5$	+	+	+	+
$u=6$	+	+	+	+
$u=7$	+	+	+	+
$u=8$	+	+	+	+
	Σ_{j1}	Σ_{j2}	Σ_{j3}	Σ_{j4}
Σ_{1i}	+	+	+	+
Σ_{2i}	+	+	+	+
Σ_{3i}	+	+	+	+
Σ_{4i}	+	+	+	+

$$p(\Sigma) \propto |\Sigma|^{-(m+1)/2} \tag{4}$$

by an analysis summarized by Box and Tiao (1973, p. 475). A different prior was proposed by Stewart (1987), but is no longer recommended.

Using Eqs. 2 and 4, and a locally uniform prior for θ, Box and Draper obtained the posterior density function

$$p(\theta,\Sigma\,|\,y) \propto |\Sigma|^{-(n+m+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}\mathbf{v}(\theta)]\right\} \tag{5}$$

for the parameters of the expectation and error models. Seeking a simpler formula, they integrated Eq. 5 over the positive definite range of Σ^{-1} and found

$$p(\theta\,|\,y) \propto |\mathbf{v}(\theta)|^{-n/2} \tag{6}$$

as the *marginal* posterior density function for the parameters of the expectation model. The mode $\hat{\theta}$ of this function occurs at the minimum of $|\mathbf{v}(\theta)|$, thus satisfying a matrix generalization of least squares.

Another way of reducing Eq. 5 is to replace Σ by its conditional maximum-density estimate, $\tilde{\Sigma}(\theta)\,|\,y$. As shown in the Appendix, this gives the modified posterior density function:

$$\tilde{p}(\theta\,|\,y) := p(\theta,\tilde{\Sigma}(\theta)\,|\,y) \propto |\mathbf{v}(\theta)|^{-(n+m+1)/2} \tag{7}$$

which has the same modal θ value as Eqs. 5 and 6. Equation 7 gives a sharper posterior distribution of θ than does Eq. 6; thus, it leads to narrower interval estimates of θ.

Equation 5 gives the fullest information, and was used by Stewart and Sørensen (1981) to obtain modal and interval estimates of θ and Σ for a sample problem. The modal θ occurs at the minimum of $|\mathbf{v}(\theta)|$, and the modal covariance estimate is

$$\hat{\Sigma} = \mathbf{v}(\hat{\theta})/(n+m+1) \tag{8}$$

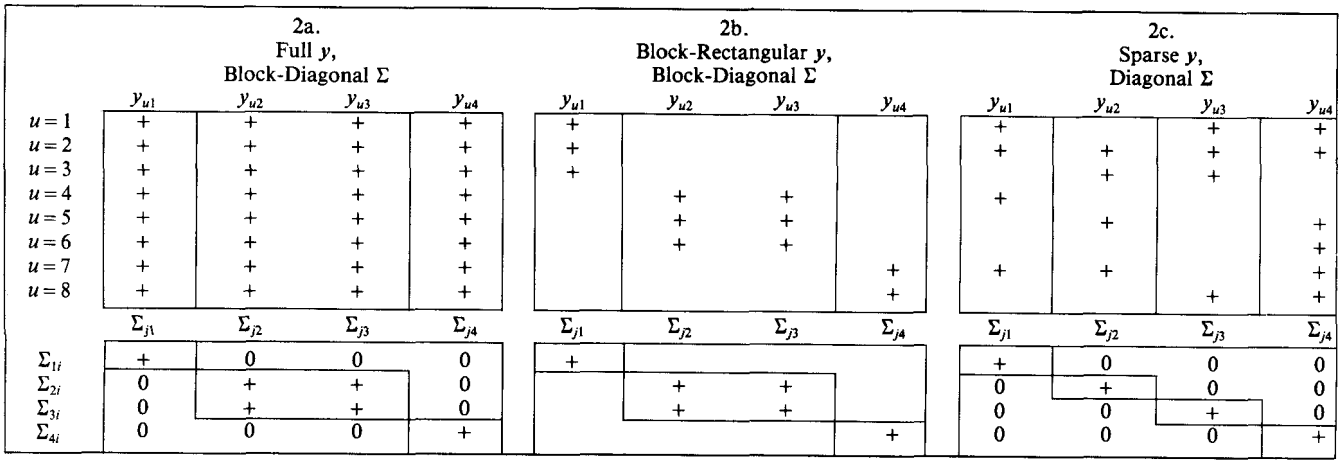
as shown in their article and in the Appendix. Users of Eq. 6 or 7 can compute this point estimate from their final matrix $\mathbf{v}(\hat{\theta})$, though the other information about Σ has been suppressed in deriving those equations from Eq. 5.

Bard (1974) gave formulas for parameter estimation based on the likelihood function in Eq. 2. The antilogarithm of his "concentrated likelihood" function $\tilde{\mathfrak{L}}(\theta)$ corresponds to $l(\theta,\tilde{\Sigma}(\theta)\,|\,y)$, which is proportional to the density function given in Eq. 6. His result for $\hat{\Sigma}(\theta)$ is like the form in Eq. 8, but has a divisor n rather than $(n+m+1)$. Clearly, sharper estimates are obtained by use of the posterior density, Eq. 5, rather than just the likelihood and by optimizing rather than integrating over the range of Σ^{-1}.

Type 2: block-rectangular y and Σ

When the errors form independent rectangular blocks as in Table 2, we call the data structure block-rectangular. The lines of data in a block need not be consecutive, though for clarity they are so arranged in Table 2. The covariances Σ_{ij} between blocks are either irrelevant to the data (as in Table 2b) or are set to zero on physical grounds (as in Tables 2a and 2c), so

Table 2. Block-Rectangular Data and Covariance Structures



that only the within-block elements of Σ need be estimated. Equations 4-7 then take the forms

$$p(\Sigma_b) \propto |\Sigma_b|^{-(m_b+1)/2} \qquad b=1,\ldots,B \quad (9)$$

$$p(\theta,\Sigma_b|y_b) \propto |\Sigma_b|^{-(n_b+m_b+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{tr}[\Sigma_b^{-1}\mathbf{v}_b(\theta)]\right\} \qquad b=1,\ldots,B \quad (10)$$

$$p(\theta|y_b) \propto (\mathbf{v}_b(\theta)|^{-n_b/2} \qquad b=1,\ldots,B \quad (11)$$

$$\tilde{p}(\theta|y_b) := p(\theta,\tilde{\Sigma}_b(\theta)|y_b) \propto |\mathbf{v}_b(\theta)|^{-(n_b+m_b+1)/2} \qquad b=1,\ldots,B \quad (12)$$

for the individual blocks $b=1,\ldots,B$ of responses. Here the matrices $\mathbf{v}_b(\theta)$ must be nonsingular, and can be made so by restricting the subsets (blocks) of responses used.

Combining Eqs. 10 for all the blocks, one obtains the posterior density function

$$p(\theta,\Sigma|y) \propto \prod_{b=1}^{B} |\Sigma_b|^{-(n_b+m_b+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{tr}[\Sigma_b^{-1}\mathbf{v}_b(\theta)]\right\}. \quad (13)$$

This equation could be used directly to estimate θ and Σ, but historically this has not been done; rather, a combination of Eqs. 11 has been used.

Combining Eqs. 11 or integrating Eq. 13 over the permitted ranges of $\Sigma_1,\ldots,\Sigma_B$, one obtains the marginal posterior density

$$p(\theta|y) \propto \prod_{b=1}^{B} |\mathbf{v}_b(\theta)|^{-n_b/2} \quad (14)$$

first given by Box and Draper (1972). Here, as with Eq. 6, information about Σ has been suppressed to get a simple alternative to the full posterior density function. This time the simplification is less successful, because Eq. 14 does not generally give a mode $\hat{\theta}$ consistent with Eq. 13.

Multiplication of Eqs. 12 suggests the new formula:

$$\tilde{p}(\theta|y) := p(\theta,\tilde{\Sigma}(\theta)|y) \propto \prod_{b=1}^{B} |\mathbf{v}_b(\theta)|^{-(n_b+m_b+1)/2}. \quad (15)$$

This result is derived in the Appendix by maximizing $p(\Sigma,\theta)$ with respect to Σ at each value of θ. Equation 15 and the covariance matrix estimates

$$\hat{\Sigma}_b = \mathbf{v}_b(\hat{\theta})/(n_b+m_b+1) \qquad b=1,\ldots,B \quad (16)$$

give the same mode $(\hat{\theta},\hat{\Sigma})$ as Eq. 13 and give nearly the same interval estimates for the θ's. Equations 15 and 16 are recommended for parameter estimation for block-rectangular data structures, pending the availability of software for Eq. 13.

Type 3: irregular y and full or partial Σ

If the error matrix $\epsilon(\theta) := y - f(x,\theta)$ does not consist of independent rectangular blocks, we call the data structure irregular. The matrix y in Table 3 corresponds to such a structure unless Σ is considered to be diagonal as in Table 2c.

One way of dealing with irregular data is to treat the missing observations y_a as additional parameters and estimate them as part of θ by use of Eq. 6 or 7. This approach is tedious and is also incorrect, since its use for a data structure like that in

Table 3. Irregular Data Structure and Relevant Covariances

	y_{u1}	y_{u2}	y_{u3}	y_{u4}
$u=1$	+			
$u=2$		+		
$u=3$			+	
$u=4$				+
$u=5$	+	+		
$u=6$	+		+	
$u=7$		+	+	
$u=8$		+		+
	Σ_{j1}	Σ_{j2}	Σ_{y3}	Σ_{j4}
Σ_{1i}	+	+	+	
Σ_{2i}	+	+	+	+
Σ_{3i}	+	+	+	
Σ_{4i}		+		+

Table 2b or 2c would replace the exponents in Eq. 14 or 15 by artificial values.

It is better to apply Bayes' theorem directly. The likelihood function can be written as

$$l(\theta,\Sigma|y)\propto\left[\prod_{u=1}^{n}|\Sigma_u|^{-1/2}\right]$$
$$\times\exp\left\{-\frac{1}{2}\sum_{u=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}\Sigma_u^{ij}[y_{ui}-f_i(x_u,\theta)][y_{uj}-f_j(x_u,\theta)]\right\}. \quad (17)$$

Here Σ_u is constructed from Σ by replacing row i and column i with those parts of a unit matrix, for each response i that is missing from experiment u. Klaus and Rippin (1979) used such a function in their algorithm for likelihood-based estimation from irregular data.

Use of Eq. 17 with Eq. 4 and a uniform $p(\theta)$ leads to the posterior density function

$$p(\theta,\Sigma|y)\propto|\Sigma|^{-(m+1)/2}\left[\prod_{u=1}^{n}|\Sigma_u|^{-1/2}\right]$$
$$\times\exp\left\{-\frac{1}{2}\sum_{u=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}\Sigma_u^{ij}[y_{ui}-f_i(x_u,\theta)][y_{uj}-f_j(x_u,\theta)]\right\} \quad (18)$$

which Stewart and Sørensen (1981) used to analyze irregular data.

Type 4: irregular y and heteroscedastic Σ

When one or more responses vary widely, it may be useful to treat their variances and covariances as functions of the predicted values of the responses. One possible model is the power function

$$\Sigma_{uij}=\omega_{ij}|f_i(x_u,\theta)f_j(x_u,\theta)|^{(\gamma_i+\gamma_j)/2} \quad i,j=1,\ldots,m \quad (19)$$

although other forms are safer for functions $f_i(x_u,\theta)$ that range through zero. Insertion of this expression into Eqs. 17 and 18, and inclusion of all independent values ω_{ij} and γ_i as parameters, give the corresponding likelihood and posterior density functions. Some results of this approach are indicated later.

Software and Computational Issues

Several portable FORTRAN programs are available for parameter estimation in nonlinear models of multiresponse data structures. The ones known to us are listed in Table 4 with current information from their authors. Each of these programs also can do nonlinear least squares.

For multiresponse problems, each of the programs minimizes an objective function tailored to the data structure and covariance assumptions. Bates and Watts use the determinant $|\mathbf{v}(\theta)|$ as the objective; the other programs use logarithmic functions of a posterior density or likelihood. Interval estimates of the parameters are computed after the minimization.

The programs differ in their repertoires of covariance models. DDTXPX and DMRCVG are written for problems of type 1. SimuSolve includes type 1 and diagonal-Σ versions of types 2 and 4. RKPES handles types 1–3. GREG has done all problem types defined above, but the code for types 3 and 4 is not yet portable.

Each of the programs uses iterative descent, based on local quadratic expansions of its objective function in terms of the parameters. The Bates-Watts and RKPES programs compute descent steps via a Levenberg-Marquardt modification of the quadratic expansion, whereas GREG and SimuSolve predict the descent steps by minimizing the current quadratic expansion over a feasible region of the parameter space. GREG uses a pivoting tolerance to select the estimable parameters and thus deals efficiently with overparametrized models.

Interval estimates of the parameters are computed in each of the programs by use of an approximate posterior density or likelihood function constructed from the final quadratic expansion. The construction of such functions is described by Stewart and Sørensen (1976, 1981) and by Stewart (1987).

Each of the programs accepts the user's mathematical model, $f(x,\theta)$, as a set of subroutines in FORTRAN (or in ACSL for SimuSolve). User-coded derivatives $\partial f_i(x_u,\theta)/\partial\theta_r$ are required in DDTXPX, optional in GREG and SimuSolve, and not used in RKPES. A general procedure for computing these derivatives for nonlinear models was given and used by Stewart and

Table 4. Programs for Multiresponse Nonlinear Parameter Estimation with Unknown Covariance Matrix Σ

Program	GREG*	SimuSolve**	DDTXPX, DMRCVG†	RKPES‡
Authors	Stewart, Caracotsios, Sørensen	Steiner, Blau, Agin	Bates, Watts	Klaus, Rippin
Problems Treated	Types 1, 2; 3, 4 pending	Types 1, 2c; 4 for diagonal Σ	Type 1	Types 1, 2, 3
Step Computation	Quadratic minimization on trust region	Quadratic minimization on trust region	Gauss-Newton or Levenberg-Marquardt	Levenberg-Marquardt
Rank Tests of v or Σ	Cholesky§ divisors vs. data resolution	QR§ divisors	Eigenvalues	Cholesky divisors
Singularities Removed by	GREG	User	User	User
User Language	FORTRAN	FORTRAN or ACSL	FORTRAN	FORTRAN

*From Sørensen (1982), Caracotsios (1986), and Stewart et al. (1992).
**From Steiner et al. (1986).
†From Bates and Watts (1984, 1987).
‡From Klaus and Rippin (1979).
§From Dongarra et al. (1979).

Table 5. Data of Fuguitt and Hawkins*

Expt. u	Feed	T, °C	t_u, min	w_u†	y_{u1} (A)	y_{u2} $(A+B)$	y_{u3} $(A+B+C)$	y_{u4} (E)
1	A	189.5	1,230	1	88.3	‡	96.2	2.2
2**	A	189.5	1,230	1	88.2	‡	95.7	1.3
3	A	189.5	3,060	2	76.4	‡	92.7	2.8
4	A	189.5	4,920	2	64.8	‡	88.9	5.8
5	A	189.5	7,800	2	50.3	‡	84.7	9.3
6	A	189.5	10,680	2	37.5	‡	82.0	12.0
7	A	189.5	15,030	2	25.9	‡	77.1	17.0
8	A	189.5	22,620	2	14.0	‡	73.9	21.0
9	A	204.5	440	2	86.6	‡	95.3	0.6
10	A	204.5	825	2	75.0	‡	91.5	1.6
11	A	204.5	1,200	2	66.0	‡	88.8	3.4
12	A	204.5	1,500	2	59.4	‡	86.4	5.1
13	A	204.5	2,040	2	48.9	‡	83.0	8.3
14	A	204.5	3,060	2	32.8	‡	77.8	13.8
15	A	204.5	6,060	2	11.5	‡	70.4	22.5
16	A	189.5	36,420	2	4.5	7.4	70.5	25.7
17	A	204.5	16,020	2	—	3.1	66.2	28.6
18	A	225.0	3,000	1	—	3.0	66.0	28.0
19**	A	225.0	3,000	1	—	4.0	66.0	28.0
20	A	245.0	630	1	—	4.0	65.0	27.0
21**	A	245.0	630	1	—	5.0	65.0	27.0
22	A	265.0	120	1	—	7.0	65.0	23.0
23**	A	265.0	120	1	—	7.0	65.0	24.0
24	A	285.0	30	1	—	11.0	66.0	19.0
25**	A	285.0	30	1	—	9.0	66.0	19.0
26	D	189.5	1,020	1	—	—	—	80.0
27	D	189.5	3,990	1	—	—	—	87.3
28**	D	189.5	3,990	1	—	—	—	87.3
29	D	189.5	6,780	1	—	—	—	87.5
30	D	189.5	8,220	1	—	—	—	86.5
31	D	189.5	13,260	1	—	—	—	88.5
32	D	189.5	14,760	1	—	—	—	89.8
33	D	204.5	3,480	1	—	—	—	87.5
34	D	204.5	5,700	1	—	—	—	86.8
35	E	189.5	8,880	1	—	—	—	91.9
36**	E	189.5	8,880	1	—	—	—	92.0
37	E	189.5	14,340	1	—	—	—	89.8
38	E	189.5	23,400	1	—	—	—	89.7
39**	E	189.5	23,400	1	—	—	—	88.5
40	E	204.5	5,700	1	—	—	—	88.4
41	E	204.5	8,100	1	—	—	—	87.9

*From Fuguitt and Hawkins (1945, 1947) and Stewart and Sørensen (1981).
**Replicate of the preceding test.
†w_u is the number of tests combined in experiment u.
‡Originally reported but not observed; see text.

Sørensen (1976), and has been implemented by Caracotsios (1985, 1986, 1992) in extensions of the integrator DASSL (Petzold, 1982). An analytic derivative algorithm is available for linear differential equation systems with distinct eigenvalues (Jennrich and Bright, 1976; Bates and Watts, 1985).

Singularities of $\mathbf{v}$ or Σ are handled in various ways. GREG uses a modified version of Subroutine DCHDC (Dongarra et al., 1979) to test which columns in the current $\mathbf{v}$ or Σ are linearly independent according to the rounding level of the data, and removes a response from the working set if it fails this test (that is, if the response is already fitted within its rounding error). Bates and Watts (1984) use eigenvalue-eigenvector analysis of the initial $\mathbf{v}$ to select a working set of responses. Any subsequent singularity of $\mathbf{v}$ in their algorithm will cause an error stop, and the user must then resubmit the problem with a smaller set of responses; the same will occur with SimuSolve or RKPES.

Some special features are of interest. GREG permits optimal selection of the next experiment for precise parameter estimation or for estimation of additional parameters of the model. SimuSolve and RKPES give extensive graphical output.

Example from Chemical Kinetics

Fuguitt and Hawkins (1945, 1947) published extensive experiments on the liquid-phase thermal reactions of α-pinene and its decomposition products. The components considered are listed here in the order of boiling point, which is also the order of their recovery in the distillation analyses of the reaction mixtures.

A. α-Pinene	$C_{10}H_{16}$
B. α- and β-Pyronene	$C_{10}H_{16}$
C. Dipentene	$C_{10}H_{16}$
D. *allo*-Ocimene	$C_{10}H_{16}$
E. Dimer	$C_{20}H_{32}$

The reaction conditions and yields are shown in Table 5. Clearly, the data structure is highly irregular.

Table 5 differs from the original tabulations of Fuguitt and Hawkins (1945, 1947) in three respects:

1. The yields are mass-balanced. One component (here D) can thus be omitted without any loss of information.
2. The remaining yields are grouped as distillation fractions: A, (A + B), (A + B + C), and E. These responses express the measurements more directly than the component yields originally reported.
3. The yields of B originally reported for experiments 1–15 are ignored, since these were constructed by linear combination of the other yields (Fuguitt and Hawkins, 1947) and thus contain no independent information.

The first 15 experiments have a rectangular data structure with $m = 3$, so they can be analyzed via the Box-Draper density of Eq. 6. Such analyses were done for experiments 1–8 by Box et al. (1973) and by Stewart and Sørensen (1976).

The full data of Table 5 were analyzed by Stewart and Sørensen (1981) via Eq. 18 with four responses and the reaction model of Figure 1. This model gives the following differential equations for the concentrations:

$$\frac{d\phi_A}{dt} = -(k_1 + k_2)\phi_A - 2k_5\phi_A^2$$

$$\frac{d\phi_B}{dt} = -k_{-3}\phi_B + k_3\phi_D$$

$$\frac{d\phi_C}{dt} = k_1\phi_A$$

$$\frac{d\phi_D}{dt} = k_2\phi_A + k_{-3}\phi_B - k_3\phi_D - 2k_4\phi_D^2 + 2k_{-4}\phi_E$$

$$\frac{d\phi_E}{dt} = k_5\phi_A^2 + k_4\phi_D^2 - k_{-4}\phi_E$$

Here the densities of the mixture and all species are treated as equal, and the molar concentrations ϕ_i are expressed relative to the molar density of pure liquid α-pinene at the reaction temperature. The resulting initial concentrations for the pure reactants are 1.0 for α-pinene, 1.0 for *allo*-ocimene, and 0.5 for the dimer. The rate coefficients are represented as Arrhenius functions,

$$\ln k_j = \ln k_{jB} - (1/T - 1/T_B)(E_j/R)$$
$$= \theta_j - (1/T - 1/T_B)\theta_{j+5} \qquad j = 1,\ldots,5$$

$$\ln(k_3/k_{-3}) = \ln K_{3B} - (1/T - 1/T_B)(\Delta E_3/R)$$
$$= -\theta_{11}/T_B - (1/T - 1/T_B)\theta_{13}$$

$$\ln(k_4/k_{-4}) = \ln K_{4B} - (1/T - 1/T_B)(\Delta E_4/R)$$
$$= -\theta_{12}/T_B - (1/T - 1/T_B)\theta_{14}$$

with k_j in min^{-1}, absolute temperature T in Kelvins, and a base temperature T_B of 478.5 K. The prior densities used were Eq. 4 for $p(\Sigma)$ in the usable (positive definite) range of Σ, and a uniform $p(\theta)$ in the permitted range of θ. The symmetric matrix Σ was parameterized with its elements on and below the diagonal.

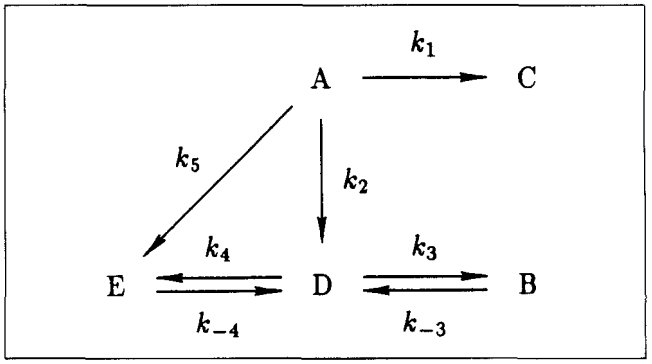


Figure 1. Reaction scheme for Example 1

Reactions −3 and 5 were not included by Fuguitt and Hawkins (1945, 1947) nor by Box et al. (1973). Reaction −3 was included to obtain dimer as an equilibrium product, as indicated by experiments 26 through 41; without this reaction, α- and β-pyronene would be the only long-time products. Reaction 5 was postulated because the short-time data on α-pinene conversion indicated the dimer to be an initial reaction product.

A structured count of the data indicated that all parameters except the covariance element Σ_{21} might be estimable if the test conditions x_u were well chosen. However, the replicate pairs $u = 18$–19, 20–21, 22–23, and 24–25 give duplicate values of y_{u3} and of y_{u2} or y_{u4}, and there are few other observations of the response pairs (3,2) and (4,2). Consequently, Σ_{21}, Σ_{32}, and Σ_{42} proved inestimable and were set to zero. The remaining parameters were estimated via Eq. 18, using the objective function $S(\theta,\Sigma) = -2\ln p(\theta,\Sigma \mid y) + c$. The estimates for this example were computed with GREG.

Table6. Parameter Estimates for α-Pinene Reaction Models*

Parameter		Estimates for 4-Reaction Model**	Estimates for 5-Reaction Model**
θ_1 =	$\ln k_1(T_B)$	−8.331 ± 0.024	−8.333 ± 0.025
θ_2 =	$\ln k_2(T_B)$	−8.898 ± 0.029	−8.961 ± 0.054
θ_3 =	$\ln k_3(T_B)$	−8.242 ± 0.341	−8.196 ± 0.325
θ_4 =	$\ln k_4(T_B)$	−5.389 ± 0.081	−5.438 ± 0.087
θ_5 =	$\ln k_5(T_B)$		−11.945 ± 0.698
θ_6 =	E_1/R	19,814 ± 428	19,785 ± 457
θ_7 =	E_2/R	20,828 ± 474	20,890 ± 536
θ_8 =	E_3/R	17,336 ± 4079	17,212 + 4,203
θ_9 =	E_4/R	10,321 ± 915	10,322 ± 918
θ_{10} =	E_5/R		19,957 ±†
θ_{11} =	$-T_B \ln K_3(T_B)$	269 ± 83	279 ± 83
θ_{12} =	$-T_B \ln K_4(T_B)$	−1,976 ± 64	−1,985 ± 63
θ_{13} =	$\Delta E_3/R$	−336 ± 950	−259 ± 958
θ_{14} =	$\Delta E_4/R$	−3,873 ± 1,624	−3,781 ± 1,555
Σ_{11}		0.696 ± 0.419	0.784 ± 0.492
Σ_{21}		0.000 ±†	0.000 ±†
Σ_{22}		0.391 ± 0.359	0.376 ± 0.348
Σ_{31}		0.358 ± 0.412	0.426 ± 0.456
Σ_{32}		0.000 ±†	0.000 ±†
Σ_{33}		0.706 ± 0.426	0.732 ± 0.444
Σ_{41}		−0.248 ± 0.344	−0.294 ± 0.354
Σ_{42}		0.000 ±†	0.000 ±†
Σ_{43}		−0.504 ± 0.317	−0.493 ± 0.314
Σ_{44}		0.744 ± 0.304	0.654 ± 0.282

*From Stewart and Sørensen (1981).
**95% probability intervals.
†Parameters not estimable from the data.

A first minimization of $S(\theta,\Sigma)$, with reaction 5 omitted, converged within 20 iterations to the parameter estimates in Table 6. The 95% probability intervals show the θ's to be estimated quite precisely. The error covariances Σ_{ij} are estimated less precisely, as expected from the limited amount of data on several pairs of responses. The interval estimates were calculated from the final quadratic approximation of $S(\theta,\Sigma)$, which yields a Normal density function $p(\theta_e,\Sigma|y)$ for the vector θ_e of estimable parameters as described by Stewart and Sørensen (1981).

A second minimization of S was done with the full five-reaction model and gave the parameter estimates shown in the last column of Table 6. A comparison of residuals showed that this model fits the data better.

The significance of the fifth reaction can be tested more objectively by calculating the posterior probability $Pr(k_{5B}>0|y)$. This can be done from the interval estimate of θ_5; we simply replace $\theta_5 := \ln k_{5B}$ by $\theta_5' := k_{5B}/\hat{k}_{5B}$ and use a uniform prior density for the new parameter θ_5'. The first and second derivatives of S at the mode are the same for both parameterizations, so the interval widths for θ_5 and θ_5' agree. The resulting 95% probability interval for $k_{5B}/\hat{k}_{5B}$ is 1 ± 0.698 and corresponds to ±1.96 standard deviations from the mode, since we are using a Normal distribution approximation for the interval estimation. Consequently, a value of zero for $k_{5B}/\hat{k}_{5B}$ deviates from the mode by $1.96(0-1)/0.698=-2.88$ standard deviations, from which we find $Pr(k_{5b}>0|y)=0.998$ by use of a table of the Normal distribution. This result strongly favors the inclusion of the fifth reaction.

The significance of the fifth reaction can be tested in another way, noting that the modal S value decreases from 41.06 to 34.09 when the additional parameter k_{5B} is estimated. Comparison of the difference $\Delta S=6.97$ against the χ^2 distribution with one degree of freedom gives $Pr(k_{5B}>0|y)=0.992$, which is close to the value 0.998 found above. The χ^2 comparison given by Stewart and Sørensen (1981) used 20 degrees of freedom and was an unduly stringent test for significance of one parameter.

Summary of Applications

Multiresponse modeling investigations are summarized in Table 7 for various chemical engineering problems. All data types except that of Table 2a are represented. Six of the investigations used data of the type in Table 1, though investigations 2 and 3 actually used lines 1-8 of the irregular data set in Table 5. Three investigations (9, 10, 11) involved independent blocks of data from different laboratories. Investigations 1, 2 and 6 gave point estimates $\hat{\theta}$ only; the others included interval estimates of the θ's.

The first real-life application of Eq. 6 was made by Mezaki and Butt (1968) to a four-component chemical reaction system. They estimated the six parameters at each temperature by minimizing $|\mathbf{v}(\theta)|$, and also by minimizing the total sum of squares, $\mathrm{tr}[\mathbf{v}(\theta)]$. The parameter estimates by the two methods differed significantly, and the minimum-$|\mathbf{v}|$ estimates were more consistent with previous literature. The greater steepness of the $|\mathbf{v}(\theta)|$ function gave faster convergence of the numerical searches. A pattern search around the final θ was used to ensure that a local minimum of $|\mathbf{v}(\theta)|$ had been found.

Investigation 2, by Box et al. (1973), is a classic study of causes and cures for singularities of $\mathbf{v}(\theta)$. For the data of Fuguitt and Hawkins (1945, 1947), use of the five reported responses gave a singular $\mathbf{v}(\theta)$ within rounding error for all parameter values, making the minimum-$|\mathbf{v}|$ criterion useless. To correct this situation the matrix y^Ty was analyzed, along with the original article, to select good subsets of responses; the minimization then went smoothly with a determinant $|\mathbf{v}(\theta)|$ of

Table 7. Multiresponse Investigations with Unknown Covariance Matrix

Investigation	Problem Type	m or $\{m_b\}$	Model Eqs.	Estimation Algorithm	No. of θ's Estimated
1. Ethanol dehydration (Mezaki and Butt, 1968)	1	4	ODE**	Search (Sobol, 1963)	6
2. α-Pinene decomposition (Box et al., 1973)	1	3	ODE**	Search (Hooke and Jeeves, 1962)	5
3. α-Pinene decomposition (Stewart and Sørensen, 1976)	1	3	ODE**	GREG	6
4. α-Pinene decomposition (Stewart and Sørensen, 1981)	3	4	ODE**	GREG	12-13
5. *o*-Xylene oxidation (Boag et al., 1978)	1	5	AE*	Search (Powell, 1964)	15
6. Oil shale pyrolysis (Ziegel and Gorman, 1980)	1	1-3	ODE**	Search (Rosenbrock, 1960)	3-6
7. Petrochemical process (Ziegel and Gorman, 1980; Stewart and Sørensen, 1976, 1981)	1	4	AE*	Search (Rosenbrock, 1960)	28-36
	1	4	AE*	GREG	28-35
8. Dow Test Problem (Biegler et al., 1986; Caracotsios, 1986)	3	2	ODAE†	GREG	7
(Biegler et al., 1986)	2c	2-4	ODE**, ODAE†	Various	7-9
9. Hydrogenation (Stewart et al., 1988)	2b	{1,1}	AE*	GREG	58-70
10. Catalytic reforming (Weidman and Stewart, 1990)	2b	{5,1}	PDAE‡	GREG	11-12
11. Distillation tray models (Young and Stewart, 1992)	2b	{1,1,1}	PDAE‡	GREG	152-157

*Algebraic equations.
**Ordinary differential equations.
†Ordinary differential and algebraic equations.
‡Partial differential and algebraic equations.

order three. McLean et al. (1979) recommended doing the rank tests on $\mathbf{v}(\theta)$ rather than $\mathbf{y}^T\mathbf{y}$; all the programs in Table 4 do this.

Investigation 3 gave a more detailed treatment of the data considered by Box et al. (1973), using an early version of Subroutine GREG. Local Taylor expansions of $|\mathbf{v}(\theta)|$ were used to compute constrained Newton steps toward the minimum $|\mathbf{v}|$ and to obtain interval estimates of θ. Local rank testing of $\mathbf{v}$ was included in the algorithm, and an extension of Newton's method was used to compute the parametric sensitivities $\partial f_i(\mathbf{x}_u)/\partial\theta_r$ directly.

Investigation 4 treated the full data of Table 5 as described in the previous section, with modal and interval estimation of θ and Σ via Eq. 18 for this irregular data structure.

Investigation 5 is a pioneering example of sequential design and analysis of multiresponse experiments. The vanadia-catalyzed oxidation of *o*-xylene was investigated in a recycle reactor at a temperature of 320°C. A comprehensive reaction model based on earlier studies was postulated, with 15 kinetic parameters. The parameters were estimated after 15 preliminary runs, and every fourth run thereafter, by minimizing the determinant $|\mathbf{v}(\theta)|$ with the search algorithm of Powell (1964). Each designed set of four runs was selected from a grid of candidates to minimize the predicted volume of the 15-parameter joint confidence region by the method of Draper and Hunter (1966). The joint confidence region shrank considerably over the early iterations of this process, and the improvement continued at a decreasing rate through the next 16 runs. Five final runs were then performed and a reduced model was fitted, retaining eight significant parameters.

Investigations 6 and 7 were reported by Ziegel and Gorman (1980). The oil shale pyrolysis study was based on data from the literature and gave a considerably improved representation by means of a multiresponse model which was fitted by minimizing $|\mathbf{v}(\theta)|$. The petrochemical process was not named, for proprietary reasons, but the study is notable as an early application of multiresponse, multiparameter modeling to a complex practical process. Models of the latter process were also investigated successfully with the subroutine GREG, as indicated in Table 7.

Investigation 8 is a set of contributed solutions to a test problem posed by the Dow Chemical Company (Biegler et al., 1986). Data on four responses were provided, along with a nine-parameter kinetic model and a four-parameter model for the variances as functions of the responses. Five contributed solutions are reported by Biegler et al. (1986), including ours which is described briefly here. The four responses satisfy two linear equality constraints (material balances); thus, the covariance matrix has rank 2 at most, and this value was used in our solution. There were occasional missing observations as in Table 3 or Table 2c. We treated Σ as a nondiagonal matrix; consequently, the variance model provided by Dow needed to be generalized to the form in Eq. 19, which is an extension of Table 3 rather than Table 2c. The resulting final parameter estimates showed Σ to be nondiagonal with high probability and showed moderate heteroscedasticity, with γ_i values of 0.0 (lower bound) and 0.190 ± 0.118 for the responses [HABM] and [AB], respectively. The five contributed solutions all represent the data well, but the parameter estimates differ appreciably because of differences in parametrization and error models.

Investigation 9 dealt with reaction rate models for the Pt-catalyzed hydrogenation of propylene. Computations via Eq. 14 were given for 20 reaction models, the best of which were constructed from evidence on surface species along with the kinetic data. The two responses were observed in separate experiments in two different laboratories. Corrections for catalyst deactivation during each of the 28 days of experiments added 55 ($=2\times28-1$) parameters to each model; earlier workers with these data implicitly used a comparable number of parameters in their graphical adjustments for changes in catalyst activity. The models were easily discriminated by ranking them according to the function

$$S'(\hat{\theta}) := \text{const.} + \sum_{b=1}^{2} n_b \ln|\mathbf{v}_b(\hat{\theta})/(n_b - n_{eb})| \qquad (20)$$

based on Eq. 14. Here n_{eb} is the number of estimated parameters used in simulating block b of the data. Each determinant $|\mathbf{v}(\theta)/(n_b - n_{eb})|$ corresponds here to a sample estimate of the variance of a single response, conditional on the current value of θ.

Investigation 10 was a study of fixed-bed reactor models and their application to the data of Hettinger et al. (1955) on catalytic reforming of C_7 hydrocarbons. The posterior density function $p(\theta|y)$ proposed by Stewart (1987) was used to estimate the rate and equilibrium parameters of various reaction schemes, two of which were reported in the article. The data were analyzed with and without models for the intraparticle and boundary-layer transport. The more detailed transport model led to a two-dimensional differential-algebraic equation system, which was solved via finite-element discretization in one dimension and adaptive stepwise integration in the other (Weidman, 1990). Parametric sensitivities were calculated during the integrations via the algorithm of Caracotsios and Stewart (1985). The detailed transport model was able to describe the data with a simpler reaction scheme.

Investigation 11 dealt with transport models for predicting the fractionation performance of sieve trays. Data from three laboratories were analyzed. There was one response (output composition) per experiment, after mass-balance adjustments; thus, the covariance matrix was diagonal as in Table 2c. For each model, three to nine predictive parameters were fitted, plus 149 parameters representing adjusted $\mathbf{x}_u$ values for the experiments. The criterion $S'(\hat{\theta})$ of Eq. 20 proved useful here again for model selection, along with interval tests for the individual parameters, and led to a promising four-parameter model for sieve-tray fractionation performance. Since the data structure consisted of single-response blocks, the minimization of $-2\ln p(\theta|y)$ gave optimally weighted least-squares solutions; the same was true in investigation 9.

Conclusion

Bayesian and likelihood approaches have been summarized here for parameter estimation from various types of multiresponse data with unknown covariance matrix Σ. The Bayesian approaches considered here use the noninformative prior of Jeffreys (1961) for Σ and give the most objective estimates. These methods are preferable to weighted least squares and to use of a specified Σ, since they estimate Σ and θ optimally from the data provided.

The posterior density function is the key to Bayesian parameter estimation, both in least squares and in multiresponse estimation. Its mode gives point estimates of the parameters, and its spread can be used to calculate their intervals of given probability content. These intervals show which parameters are really needed to represent the data; they should always be determined and reported.

Most investigations with unspecified Σ have used procedures based on Eq. 6 or 14, which exclude Σ from the parameter set (but implicitly estimate its mode nonetheless; see Eqs. 8 and 16). We have given new formulas of this type, Eqs. 7 and 15, which give sharper estimates and are more consistent with the full posterior density functions of Eqs. 5 and 13.

Several software programs have been described which do both least squares and multiresponse parameter estimation. Some multiresponse applications have been reviewed. Persons interested in any of the programs listed in Table 4 should contact the respective authors.

Acknowledgment

The authors are pleased to acknowledge the encouragement received from George Box and Norman Draper, as well as the fiscal support of the National Science Foundation (Grant CBT-8618571) and the U.S. Department of Energy, Office of Basic Energy Sciences (Grant DE-FG02-84ER13291).

Notation

b = block number, with range from 1 to B
$f_i(x_u,\theta)$ = expectation model for response i in event u
$f(x,\theta)$ = array of expectation models $f_i(x_u,\theta)$ for observations
E_j/R = activation energy of reaction j, K
$\Delta E_j/R$ = standard energy of reaction j, K
k_j, k_{-j} = forward and reverse rate coefficients for reaction j
K_j = equilibrium constant for reaction j
$l(\theta,\Sigma|y)$ = likelihood function in Eq. 2 or 17
m = number of working responses in data array
m_b = number of working responses in block b
m_u = number of working responses in event u
n = number of events in data set
n_b = number of events in block b
n_e = number of parameters estimated
$p(x)$ = prior probability density in space of x
$p(\theta,\Sigma|y)$ = posterior density function in Eq. 5 or 18
$p(\theta|y)$ = marginal posterior density function in Eq. 6
$\tilde{p}(\theta|y)$ = modified posterior density function in Eq. 7
R = gas constant, consistent units
$S(\theta,\Sigma)$ = $-2\ln p(\theta,\Sigma|y)$
t = time, consistent units
T = absolute temperature, K
T_B = base temperature for parametrization of reaction model
u = event number in Eq. 1
x_u = vector of experimental settings for event u

Greek letters

γ_i = exponent for response i in Eq. 19
$\epsilon_{ui}(\theta)$ = error in observation y_{ui} according to Eq. 1
θ = vector of parameters in Eq. 1
$\hat{\theta}$ = modal value of θ
Σ = error covariance matrix
$\hat{\Sigma}$ = modal value of Σ
$\tilde{\Sigma}(\theta)|y$ = maximum-density estimate of Σ, conditional on θ and y
ω_{ij} = reference value of Σ_{ij} in Eq. 19

Subscripts

b = of response block b
B = at reference temperature T_B
e = of estimated parameters
u = of event u

Functions and operations

$E(z)$ = expectation of random variable z
$(l:=r)$ = notation l is defined as expression r
Pr = cumulative probability
$\mathrm{tr}(A)$ = trace of matrix A
$|A|$ = determinant of matrix A
$\mathbf{A}^T$ = transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$ = inverse of matrix $\mathbf{A}$
$\mathbf{A}^{ij}$ = element i, j of $\mathbf{A}^{-1}$
$\propto$ = proportionality sign
Π = product
Σ = summation

Literature Cited

Aitken, A. C., "On Least Squares and Linear Combination of Observations," *Proc. of the Royal Soc. of Edinburgh*, **55**, 42 (1935).
Bard, Y., "Nonlinear Estimation and Programming," Program 360D 13.6.003, IBM, New York (1968).
Bard, Y., *Nonlinear Parameter Estimation*, Academic Press, New York (1974).
Bates, D. M., and D. G. Watts, "A Multiresponse Gauss-Newton Algorithm," *Commun. in Statistics—Simul. and Comput.*, **13**, 705 (1984).
Bates, D. M., and D. G. Watts, "Multiresponse Estimation with Special Application to Systems of Linear Differential Equations," *Technometrics*, **27**, 329 (1985).
Bates, D. M., and D. G. Watts, "A Generalized Gauss-Newton Procedure for Multiresponse Parameter Estimation," *SIAM J. of Scientific and Statistical Comput.*, **7**, 49 (1987).
Bayes, T. R., "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Trans. of the Royal Soc. of London*, **53**, 370 (1763); reprinted in *Biometrika*, **45**, 293 (1958).
Biegler, L. T., J. J. Damiano, and G. E. Blau, "Nonlinear Parameter Estimation: a Case Study Comparison," *AIChE J.*, **32**, 29 (1986).
Boag, I. F., D. W. Bacon, and J. Downie, "Using a Statistical Multiresponse Method of Experimental Design in a Reaction Network Study," *Can. J. of Chem. Eng.*, **56**, 389 (1978).
Box, G. E. P., and N. R. Draper, "The Bayesian Estimation of Common Parameters from Several Responses," *Biometrika*, **52**, 355 (1965).
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 (1973).
Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA (1973).
Box, M. J., and N. R. Draper, "Estimation and Design Criteria for Multiresponse Nonlinear Models with Nonhomogeneous Variance," *Appl. Statistics*, **21**, 13 (1972).
Box, M. J., N. R. Draper, and W. G. Hunter, "Missing Values in Multiresponse Nonlinear Data Fitting," *Technometrics*, **12**, 613 (1970).
Caracotsios, M., "Model Parametric Sensitivity Analysis and Nonlinear Parameter Estimation: Theory and Applications," Ph.D. Thesis, Dept. of Chemical Engineering, Univ. of Wisconsin, Madison (1986).
Caracotsios, M., and W. E. Stewart, "Sensitivity Analysis of Initial Value Problems with Mixed ODEs and Algebraic Equations," *Comp. & Chem. Eng.*, **9**, 359 (1985).
Caracotsios, M., and W. E. Stewart, "Sensitivity Analysis of Initial-Boundary-Value Problems with Mixed Algebraic and Partial Differential Equations," *Comp. & Chem. Eng.*, submitted (1992).
Deming, W. E., *Statistical Adjustment of Data*, Wiley, New York (1943).
Dongarra, J. J., J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users' Guide*, SIAM, Philadelphia (1979).
Draper, N., and W. G. Hunter, "Design of Experiments for Parameter Estimation in Multiresponse Situations," *Biometrika*, **53**, 525 (1966).
Fisher, R. A., "On the Mathematical Foundations of Theoretical

Statistics," *Philosophical Trans. of the Royal Soc. of London*, Ser. A, **222**, 309 (1922).
Fuguitt, R. E., and J. E. Hawkins, "The Liquid-Phase Thermal Isomerization of α-Pinene," *J. of the Amer. Chem. Soc.*, **67**, 242 (1945).
Fuguitt, R. E., and J. E. Hawkins, "Rate of the Thermal Isomerization of α-Pinene in the Liquid Phase," *J. of the Amer. Chem. Soc.*, **69**, 319 (1947).
Gauss, C. F., *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*, Perthes et Besser, Hamburg (1809); reprinted as *Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, translated by C. H. Davis, Dover, New York (1963).
Hettinger, W. P., Jr., C. D. Keith, J. L. Gring, and J. W. Teter, "Hydroforming Reactions: Effect of Certain Catalyst Properties and Poisons," *Ind. Eng. Chem.*, **47**, 719 (1955).
Hooke, R., and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems," *J. of the Assoc. of Comput. Machinery*, **8**, 212 (1962).
Jeffreys, H., *Theory of Probability*, 3rd ed., Clarendon Press, Oxford (1961).
Jennrich, R. I., and P. B. Bright, "Fitting Systems of Linear Differential Equations Using Computer Generated Exact Derivatives," *Technometrics*, **18**, 385 (1976).
Klaus, R., and D. W. T. Rippin, "A New Flexible and Easy-to-Use General-Purpose Regression Program," *Comput. & Chem. Eng.*, **3**, 105 (1979).
McLean, D. D., D. J. Pritchard, D. W. Bacon, and J. Downie, "Singularities in Multiresponse Modeling," *Technometrics*, **21**, 291 (1979).
Mezaki, R., and J. B. Butt, "Estimation of Rate Constants from Multiresponse Kinetic Data," *Ind. and Eng. Chemistry Fund.*, **7**, 120 (1968).
Petzold, L. R., "A Description of DASSL: a Differential/Algebraic System Solver," *Sandia Tech. Rep.*, 82-8637 (1982).
Powell, M. J. D., "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives," *Comput. J.*, **7**, 155 (1964).
Rosenbrock, H. H., "An Automatic Method for Finding the Greatest or Least Value of a Function," *Comput. J.*, **3**, 175 (1960).
Sobol, M., "NU Scoop Linear Surface Minimization Routine," IBM Share Program 1551 (1963).
Steiner, E. C., G. E. Blau, and G. L. Agin, *Introductory Guide to SIMUSOLVE*, Mitchell & Gauthier, Concord, MA (1986).
Stewart, W. E., and J. P. Sørensen, "Sensitivity and Regression of Multicomponent Reactor Models," *Int. Symp. on Chem. Reaction Eng.*, DECHEMA, Frankfurt, **I**, 12 (1976).
Stewart, W. E., and J. P. Sørensen, "Bayesian Estimation of Common Parameters from Multiresponse Data with Missing Observations," *Technometrics*, **23**, 131 (1981); **24**, 91 (1982).
Stewart, W. E., "Multiresponse Parameter Estimation with a New and Noninformative Prior," *Biometrika*, **74**, 557 (1987).
Stewart, W. E., R. H. Shabaker, and Y. T. Lu, "Kinetics of Propylene Hydrogenation on Platinum-Alumina," *Chem. Eng. Sci.*, **43**, 2257 (1988); **44**, 205 (1989).
Stewart, W. E., M. Caracotsios, and J. P. Sørensen, "GREG Software Package Documentation," Dept. of Chemical Engineering, Univ. of Wisconsin—Madison (1992).
Sørensen, J. P., "Simulation, Regression and Control of Chemical Reactors by Collocation Techniques," Doctoral Thesis, Danmarks tekniske Højskole, Lyngby (1982).
Weidman, D. L., and W. E. Stewart, "Catalyst Particle Modeling in Fixed-Bed Reactors," *Chem. Eng. Sci.*, **45**, 2155 (1990).
Young, T. C., and W. E. Stewart, "Correlation of Fractionation Tray Performance with a Distributed Boundary-Layer Model," *AIChE J.*, **38**(4), 592 (1992).
Ziegel, E. R., and J. W. Gorman, "Kinetic Modeling with Multiresponse Data," *Technometrics*, **22**, 139 (1980).

Appendix: Proof of Eqs. 15 and 16

The following identities for real nonsingular square matrices A and B are derived by Bard (1974):

$$\frac{\partial \ln |A|}{\partial A} = (A^{-1})^T \tag{A1}$$

$$\frac{\partial A^{kl}}{\partial A_{ij}} = -A^{ki}A^{jl} \tag{A2}$$

$$\frac{\partial}{\partial A}\,\mathrm{tr}(A^T B) = B \tag{A3}$$

Here the elements of dA have been treated as independent, so that $\partial/\partial A^T$ is a full square matrix. The transposition symbol T is omitted hereafter, since the matrices considered are symmetric. Elements of A^{-1} are indicated by superscript indices.

For analysis of Eq. 13, define the function

$$F(\theta,\Sigma) = -2\ln p(\theta,\Sigma) + c$$
$$= \sum_{b=1}^{B} (n_b + m_b + 1)\ln|\Sigma_b| + \sum_{b=1}^{B} \mathrm{tr}[\Sigma_b^{-1}\mathbf{v}_b(\theta)] \tag{A4}$$

for nonsingular Σ_b and $\mathbf{v}_b(\theta)$ in each block b. Differentiation with respect to Σ_b^{-1} for a particular block b gives

$$\frac{\partial F}{\partial \Sigma_b^{-1}} = -(n_b + m_b + 1)\Sigma_b + \mathbf{v}_b(\theta) \tag{A5}$$

after use of Eqs. A1 and A3. When this derivative matrix is zero, $\partial F/\partial \Sigma_b$ is also zero in view of Eq. A2. The resulting stationary value of F is a local minimum with respect to Σ_b, and the corresponding conditional estimate of Σ_b is:

$$[\tilde{\Sigma}_b(\theta)\,|\,y] = \mathbf{v}_b(\theta)/(n_b + m_b + 1). \tag{A6}$$

Hence,

$$[\tilde{\Sigma}_b^{-1}(\theta)\,|\,y] = (n_b + m_b + 1)\mathbf{v}_b^{-1}(\theta) \tag{A7}$$

and substitution into Eq. A4 gives:

$$F(\theta,\tilde{\Sigma}(\theta)\,|\,y) = -2\ln p(\theta,\tilde{\Sigma}(\theta)\,|\,y) + c$$
$$= \sum_{b=1}^{B} (n_b + m_b + 1)\ln\left|\frac{\mathbf{v}_b(\theta)}{(n_b + m_b + 1)}\right|$$
$$+ \sum_{b=1}^{B} (n_b + m_b + 1)\mathrm{tr}[\mathbf{v}_b^{-1}(\theta)\mathbf{v}_b(\theta)]$$
$$= \sum_{b=1}^{B} (n_b + m_b + 1)\ln|\mathbf{v}_b(\theta)| + \text{const.} \tag{A8}$$

Solving for $p(\theta,\tilde{\Sigma}(\theta)\,|\,y)$ then gives Eq. 15. It follows that the modes of Eqs. 13 and 15 correspond, since they occur at the minimum (or minima) of F. Application of Eq. A6 at the modal value $\hat{\theta}$ verifies Eq. 16.

Manuscript received Oct. 21, 1991, and revision received Mar. 20, 1992.