Biochem. J. (1987) 245, 67–74 (Printed in Great Britain)

Analytical methods for fitting integrated rate equations

A discontinuous assay

Elizabeth A. BOEKER
Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-0300, U.S.A.

The integrated rate equation for reactions with stoichiometry A→P+Q is:

$$e_0 t = -C_f \cdot \ln(1-\Delta P/A_0) + C_1 \Delta P + \tfrac{1}{2} C_2 (\Delta P)^2$$

where the coefficients C are linear or quadratic functions of the kinetic constants and the initial substrate and product concentrations. I have used the 21 progress curves described in the accompanying paper [Cox & Boeker (1987) Biochem. J. **245**, 59–65] to develop computer-based analytical and statistical techniques for extracting kinetic constants by fitting this equation. The coefficients C were calculated by an unweighted non-linear regression: first approximations were obtained from a multiple regression of t on ΔP and were refined by the Gauss–Newton method. The procedure converged in six iterations or less. The bias in the coefficients C was estimated by four methods and did not appear to be significant. The residuals in the progress curves appear to be normally distributed and do not correlate with the amount of product produced. Variances for C_f, C_1 and C_2 were estimated by four resampling procedures, which gave essentially identical results, and by matrix inversion, which came close to the others. The reliability of C_2 can also be estimated by using an analysis-of-variance method that does not require resampling. The final kinetic constants were calculated by standard multiple regression, weighting each coefficient according to its variance. The weighted residuals from this procedure were normally distributed.

INTRODUCTION

In this paper I develop computational procedures that extract kinetic constants from integrated rate equations (Boeker, 1984*a*,*b*, 1985) for a relatively simple enzyme system. For the stoichiometry A→P+Q, the equations to be fit are:

$$e_0 t = -C_f \cdot \ln\left(1-\frac{\Delta P}{A_0}\right) + C_1 \Delta P + \tfrac{1}{2} C_2 (\Delta P)^2 \quad (1)$$

$$C_f = \frac{J_0}{J_A k_{cat.}} + \frac{J_P}{J_A k_{cat.}}(A_0+P_0) + \frac{J_Q}{J_A k_{cat.}}(A_0+Q_0) + \frac{J_{PQ}}{J_A k_{cat.}}(A_0+P_0)(A_0+Q_0) \quad (2)$$

$$C_1 = \frac{1}{k_{cat.}} - \frac{J_P+J_Q}{J_A k_{cat.}} + \frac{J_{AP}}{J_A k_{cat.}} P_0 + \frac{J_{AQ}}{J_A k_{cat.}} Q_0 - \frac{J_{PQ}}{J_A k_{cat.}}(A_0+P_0+Q_0) \quad (3)$$

$$C_2 = \frac{J_{AP}+J_{AQ}-J_{PQ}}{J_A k_{cat.}} \quad (4)$$

The terminology of these equations is described in detail in the accompanying paper (Cox & Boeker, 1987). Eqn. (1) describes any progress curve in terms of three coefficients, C_f, C_1 and C_2. These coefficients are related to the Dalziel (1957) constants for the enzyme and the initial conditions of the progress curve by eqns. (2)–(4).

The first step of the calculation is to obtain values of C_f, C_1 and C_2 such that the calculated progress curve predicts the actual data. The difficulty here is that it is the measurement of ΔP, rather than t, which is ordinarily subject to experimental error. Eqn. (1) is 'backwards' with respect to the dependent and independent variables. An ordinary multiple regression of t on ΔP gives a curve in which $\Sigma(t_i-\hat{t}_i)^2$ is minimized, rather than $\Sigma(\Delta P_i-\Delta\hat{P}_i)^2$.

The direct approach to this problem, which would be to solve for ΔP, is not possible. Estimates of C_f, C_1 and C_2 that minimize the product residuals must instead be obtained by a non-linear regression, a technique that has been applied to enzyme kinetic data in other contexts (Wilkinson, 1961; Fernley, 1974; Darvey *et al.*, 1975;

Notation used: A, P and Q and A_0, P_0 and Q_0 are respectively the instantaneous and initial concentrations of substrates and products; ΔP is $P-P_0$, the net change in product concentration at time t. The coefficients J are collections of microscopic rate constants that, for a particular mechanism, result directly from a King & Altman (1956) derivation. The coefficients C are defined by eqns. (2)–(4). $k_{cat.}$ is the catalytic constant or turnover number and e_0 is the enzyme concentration. The subscript i indicates a particular measurement, a circumflex (^) indicates a best-fit (rather than a measured) value, and an overbar (¯) indicates an average value.

Duggleby & Morrison, 1977, 1978; Duggleby, 1981; Matyska & Kovář, 1985). For non-linear regression, a first estimate of C_f, C_1 and C_2 is made and the sum of $(\Delta P_i - \Delta \hat{P}_i)^2$ is formed, by repeatedly solving eqn. (1) (numerically) for $\Delta \hat{P}_i$. Better approximations are obtained, and the process is repeated until a minimum sum is reached. I have used a 'backwards' regression of t on ΔP to obtain the initial estimates, and the Gauss–Newton method for refining them. I have also examined the structure of the experimental error, the effect of weighting the data, and the possible bias introduced by the non-linear procedure.

Eqns. (2)–(4), for C_f, C_1 and C_2 in terms of the initial concentration of substrates and products, can be fitted with ordinary multiple-regression techniques. The coefficients that result are the Dalziel constants for the enzyme. For these regressions to be meaningful, the values of C_f, C_1 and C_2 must be weighted according to their uncertainties. This weighting is critically important. Because the product terms in eqn. (1), $-\ln(1-\Delta P/A_0)$, ΔP and $(\Delta P)^2$, are highly correlated, it may be possible to fit a progress curve by using a range of values in the coefficients C. For example, if C_f is decreased, a corresponding increase in C_1 and/or C_2 may produce a curve with very nearly as good a fit. Since the seriousness of this problem will depend greatly on the shape of the progress curve (i.e. on the initial conditions), the uncertainties in C_f, C_1 and C_2 must be calculated and used as weighting factors.

There are two experimental situations. In the first, exemplified by a spectrophotometric progress curve, all the points result from a single addition of enzyme to substrate. The points are not statistically independent, but progress curves are easily come by. It is sensible simply to do a number of progress curves for each set of initial conditions and to calculate the mean and standard deviation for each coefficient C.

In the second situation the assay is discontinuous, as is the radioactivity assay in the accompanying paper (Cox & Boeker, 1987). Each point in each progress curve results from a separate addition of enzyme to substrate; the points on any one curve can reasonably be considered to be statistically independent. Although each such curve is a great deal of work, it should be possible to obtain coefficient variances without repeating the curve.

There is no standard statistical procedure for calculating variance with non-linear regression. In the present paper I examine two data-simulation (Monte Carlo) methods and two non-parametric methods, the jackknife and the bootstrap, as well as a matrix-inversion method that is not theoretically sound but is computationally cheap. Monte Carlo methods and the jackknife have been used before in order to estimate variance in enzyme kinetic data (Cornish-Bowden & Wong, 1978; Duggleby, 1979, 1980; Matyska & Kovář, 1985).

COMPUTATIONAL METHODS

The experimental data used in this paper are the 21 time courses reported in the accompanying paper (Cox & Boeker, 1987). The comparison were carried out in FORTRAN on a VAX 11/780 computer; the programs are available from the author on request.

Numerical solution to the non-linear equation

Eqn. (1) can be rewritten as:

$$0 = -C_f \cdot \ln(1-\Delta P/A_0) + C_1 A_0 (\Delta P/A_0) + \tfrac{1}{2} C_2 A_0^2 (\Delta P/A_0)^2 - e_0 t$$

The root of the rewritten equation, $\Delta P/A_0$, represents the fractional reaction, and must lie between 0 and 1. This root is required in order to calculate both the sum of squares and the partial derivatives needed for the non-linear regression. The root was first obtained by a simple bisection method [see King (1984) or any standard numerical-analysis text]. However, since the calculation is required seven times for each data point for each iteration of the non-linear regression, bisection slowed the overall computation unnecessarily. The final procedure adopted was a modification of Newton's method (King, 1984). An upper bound was first established and the root was approached from above in order to avoid the discontinuity at $\Delta P/A_0 = 1$. The routine was terminated when the root changed < 0.00001 between iterations. If the value of the function was > 0 at $\Delta P/A_0 < 0.0001$ or < 0 at $\Delta P/A_0 > 0.9999$, the value of the root was taken to be 0 or 1 respectively.

Multiple regressions

For the non-linear regression, multiple regressions are required to calculate both the initial estimates and the corrections at each iteration. They are also needed to find the best-fit Dalziel constants once C_f, C_1 and C_2 are known. Considering the system $\mathbf{XB} = \mathbf{Y}$ where $\mathbf{X}$ is an $m \times n$ matrix of the independent variables, $\mathbf{B}$ is the $n \times 1$ vector of the coefficients to be calculated, and $\mathbf{Y}$ is the $m \times 1$ vector of the dependent variables, multiple regressions were carried out by forming $\mathbf{X}^t\mathbf{XB} = \mathbf{X}^t\mathbf{Y}$ (Neter & Wasserman, 1974) and using Gaussian elimination with scaled partial pivoting (King, 1984) to solve for $\mathbf{B}$. For calculations of the Dalziel constants, the procedure was weighted according to the variance of each y_i. Where required, variances were obtained for each b_i by inverting $\mathbf{X}^t\mathbf{X}$ and multiplying the diagonal elements by the appropriately weighted $\Sigma(y_i - \hat{y}_i)^2/(m-n)$.

Non-linear regression

For each progress curve, initial estimates of C_f, C_1 and C_2 were obtained by running a 'backwards' regression of t on ΔP for those data where $\Delta P/A_0 < 0.85$. The initial estimates were then refined by calculating three sets of partial derivatives, $(\partial P/\partial C_f)_i$, $(\partial P/\partial C_1)_i$ and $(\partial P/\partial C_2)_i$, and running a multiple regression of $\Delta P_i - \Delta \hat{P}_i$ against them. The coefficients of the multiple regression are the corrections to C_f, C_1 and C_2. The derivatives were calculated numerically; each coefficient C was changed by $+1\%$ and -1% of its most recent value, and the corresponding value of $\Delta \hat{P}_i$ was found. The size of ΔC was not especially important; changes of up to $+5\%$ and -5% appeared to work equally well. The process was terminated when $\Sigma(\Delta P_i - \Delta \hat{P}_i)^2$ declined by 0.01% or less. A more stringent convergence criterion did not change the final result significantly, and some-

times led to oscillation around the minimum sum of squares.

Variance of C_f, C_1 and C_2

The first two methods were Monte Carlo methods in which the time courses were simulated repeatedly. In each case 'perfect' values of ΔP_i were calculated from the actual values of t and the best-fit values of C_f, C_1 and C_2. An error was then introduced according to either an absolute ($\Delta P_i = \Delta P_i + z_i\sqrt{MSE}$) or a proportional [$\Delta P_i = \Delta P_i(1 + z_i\sqrt{MSE})$] error model. z_i was obtained from a randon number generator and had a standard normal distribution; MSE was the (actual, not simulated) $\Sigma(\Delta P_i - \hat{P}_i)^2$/(number of points -3) for the time course. C_f, C_1 and C_2 were then calculated for each simulation just as they had been for the original time course. The procedure was repeated 100 times for each error model for each time course, and the mean and variance of the simulated values of C_f, C_1 and C_2 were calculated. The bias was estimated from the difference between the original coefficients C and the mean of the simulated coefficients: bias $= \hat{C}(1 - \hat{C}/\overline{C})$.

The jackknife and the bootstrap (Efron, 1982) are non-parametric methods that depend on resampling the available data. For the jackknife, resampled coefficients C are calculated at each point on the progress curve. For example, for the fifth point in a set of 23, C_f would first be calculated from the full set and then from a reduced set of 22 with the fifth point omitted. The resampled C_f is then $23 \times C_f$(full set) minus $22 \times C_f$ (reduced set). The procedure is repeated at each point, and the mean and variance of the resampled values of C_f are calculated. The bias is the difference between the original value of C_f and the resampled mean.

The bootstrap treats the original data set as the maximum likelihood estimator of the set of all possible data for that progress curve. For each repetition, the points on a particular progress curve were randomly sampled, with replacement, until there were as many as in the original progress curve. The coefficients C were then calculated and, after 100 repetitions, their means and variances. The bias is again the difference between the original coefficient and its resampled mean.

Analysis of variance

For C_2, t was calculated from:

$$t = \sqrt{\frac{SSRc' - SSRr'}{MSEc}}$$

SSRc′ is the regression sum of squares for the complete fit (including C_f, C_1 and C_2), adjusted for the non-additivity of the sums of squares. That is:

$$SSRc' = \frac{SSRc}{SSRc + SSEc} \times TSS$$

where $SSRc = \Sigma(\Delta\hat{P}_i - \Delta\overline{P})^2$, $SSEc = \Sigma(\Delta P_i - \Delta\hat{P}_i)^2$ and $TSS = \Sigma(\Delta P_i - \Delta\overline{P})^2$. SSRr′ is the corresponding adjusted sum of squares for the reduced fit, including only C_f and C_1. MSEc is SSEc/(number of points -3).

The values of t used for comparison were C_2 divided by the standard deviation of C_2 as calculated from the jackknife.

Dalziel constants

Of the two products of arginine decarboxylase, only agmatine was used for product-inhibition studies in the accompanying paper (Cox & Boeker, 1987). For notation purposes, agmatine is arbitrarily designated Q without implying a particular reaction order. In the absence of the second product, eqns. (2) and (3) now simplify somewhat. From eqn. (2), a regression of C_f against A_i, $A_i + Q_i$ and $A_i(A_i + Q_i)$ gives four of the seven Dalziel constants. From eqn. (3), regression of C_1 against Q_i and $A_i + Q_i$ gives two more. C_2 should be constant and should give the final Dalziel term; all that is required is the average of the observations. Regressions and averages were in all cases weighted by the appropriate variances. $J_{PQ}/J_A k_{cat.}$ is calculated from both C_f and C_1; the final value was taken to be the weighted average.

RESULTS

The Gauss–Newton method for non-linear regression converges only when the initial estimates of the unknown parameters are reasonably good. I have used a 'backwards' multiple regression of t on ΔP, making use only of values of $\Delta P/A_0$ between 0 and 0.85. Because the progress curves flatten out between 0.85 and 1.0, and because this regression minimizes the error along the $e_0 t$ (horizontal) axis, including data late in the progress curve introduces large errors.

A comparison between the initial estimates and the final values of C_f, C_1 and C_2 is shown in Table 1. For C_f and C_1, the estimates are good to within an order of magnitude of the final result, and are generally within a factor of 3-fold. The estimates for C_2 appear to be worse. However, C_2 has a very small absolute value and is very

Table 1. Data weighting and approximations

Initial concentration (mM)		Correlation between e_i and P_i*	Estimated value ÷ final value		
Arginine	Agmatine		C_f	C_1	C_2
0.05	–	0.14	0.78	4.67	2.37
0.2	–	0.16	0.43	3.93	1.29
0.5	–	0.06	1.47	0.61	−5.38
1.0	–	0.30	1.44	0.72	8.64
2.0	–	0.05	1.58	0.83	−7.16
5.0	–	0	0.64	1.08	0.79
10	–	0	1.30	0.94	5.42
20	–	0.09	2.31	0.79	1.74
30	–	0	2.09	0.59	5.41
0.5	10	0.10	0.48	1.55	4.02
0.5	20	0.23	0.89	1.05	2.62
0.5	40	0.18	0.67	5.03	−1.09
0.5	80	0.15	0.33	6.48	−9.46
5.0	40	0.08	0.67	1.69	0.05
10	40	0.04	0.91	1.03	−0.04
20	40	0.14	0.71	1.30	0.52
30	40	0.27	†	3.75	−2.47
30	10	0.11	0.41	1.33	−1.77
30	20	0.07	1.00	1.02	1.04
30	60	0	0.69	1.56	0.06
30	80	0.13	0.70	1.53	−0.04

* Pearson correlation coefficient.
† The estimate was negative; see the text.

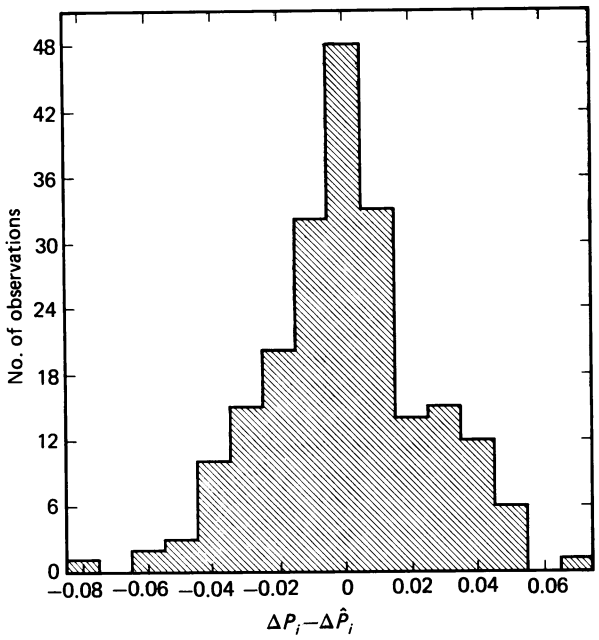

Fig. 1. Distribution of residuals

Data are for the first nine rows of Tables 3–5.

uncertain; the estimates are better than they appear. For the last entry in Table 1, for example, the estimated and final values were 0.004 and −0.092 respectively.

Occasionally, the estimated value of C_f was less than zero; the non-linear regression then invariably diverged. This occurred for the 17th progress curve in Table 1, as well as in several of the simulated curves. All of the terms in eqn. (2) have positive signs and, since the negative logarithm of a fraction is also positive, C_f itself should under all conditions be positive. (This is not true for either C_1 or C_2.) The simple expedient of replacing the negative estimate with a very small positive one (0.001) led to convergence in all of these cases.

With the use of these 'backwards-regression' estimates and the computational procedure described in the Computational methods section, the non-linear regression converges very quickly; three or four iterations were generally all that were required, and never more than six.

The regression was initially carried out without weighting the data. It is frequently true that the error in kinetic data is proportional to ΔP, rather than independent of it (Storer *et al.*, 1975). The data should then be weighted as $1/(\Delta P)^2$. It is not possible, from the experimental procedures in the accompanying paper (Cox & Boeker, 1987), to make a statement *a priori* about the error structure of these data. One approach to this problem is to calculate the residuals, $e_i = \Delta P_i - \Delta \hat{P}_i$, and ask whether they are correlated with ΔP. The results are shown in Table 1. The average value of the correlation coefficient is only 0.11 ± 0.09. And, as shown in Fig. 1, these residuals appear to be normally distributed. When a second, weighted, fit was carried out for all of the time courses, the values of C_f, C_1 and C_2 were essentially unchanged from the first fit. Evidently the independent error model is at least as good as the proportional one.

A potential problem with this non-linear regression is that, although the results are minimum variance estimators of C_f, C_1 and C_2, they are not necessarily unbiased. Four of the methods used in the next section to estimate variance also estimate the bias in the calculated values of C_f, C_1 and C_2. These results are summarized in Table 2. In general, the methods give bias estimates that do not all have the same sign, suggesting that the true bias may be close to 0. Further, the bias values for C_f and C_1 are quite small. The values for C_2 are both large and scattered; this is presumably caused by the fact that C_2 is both small and uncertain. It does not seem likely that bias of the magnitude suggested by these results will affect the subsequent calculations. The standard deviations of C_f, C_1 and C_2 (next section) are generally larger than the bias values in Table 2.

The results of five possible techniques for obtaining the standard deviations of C_f, C_1 and C_2 are shown in Tables 3–5. The values shown in columns 1 and 2 are based on repeated simulations, introducing a normally distributed random error. The results will be valid only if the measurement errors in the progress curves are normally distributed. The distribution of residuals for the first nine time courses in Tables 3–5 is shown in Fig. 1; the distributions for the remaining progress curves are similar.

Results from two non-parametric techniques are shown in columns 4 and 5. These values are similar to those in columns 1 and 2. The results in column 3 are from the 'backwards' regression of t on ΔP used to make the initial estimates of C_f, C_1 and C_2. This procedure is not theoretically sound; the variance–co-variance matrix depends on t rather than ΔP. I have nevertheless tried it,

Table 2. Estimates of bias in the coefficients

The methods are described in detail in the text; method 1, data simulation with the use of an independent error model; method 2, simulation with the use of a proportional error model; method 3, the jackknife; method 4, the bootstrap. Mean bias is averaged over the time courses and expressed as a percentage of the mean of the appropriate coefficient. Standard deviation is of the sample. The distributions did not appear to be normal.

	C_f		C_1		C_2	
Method	Mean bias	Standard deviation	Mean bias	Standard deviation	Mean bias	Standard deviation
1	1.7	2.5	−6.1	11.4	95	198
2	2.2	3.5	−1.4	14.1	−121	659
3	0.4	4.3	−7.0	22.0	67	167
4	−0.8	3.5	6.3	18.8	−116	300

Table 3. Values and standard deviations for C_f

The methods, which are described in detail in the text, are: method 1, data simulation with the use of an independent error model; method 2, simulation with the use of a proportional error model; method 3, 'backwards' regression; method 4, the jackknife; method 5, the bootstrap. The units of C_f and the standard deviations are $\mu g \cdot ml^{-1} \cdot min$.

Initial concentration (mM)		Value of	Standard deviation from				
Arginine	Agmatine	C_f	Method 1	Method 2	Method 3	Method 4	Method 5
0.05	–	1.99	0.46	0.71	0.36	0.62	0.78
0.2	–	0.42	0.29	0.35	0.57	0.36	0.29
0.5	–	1.24	0.37	0.50	0.60	0.55	0.50
1.0	–	1.57	0.29	0.41	0.36	0.47	0.38
2.0	–	1.78	0.28	0.39	0.75	0.33	0.35
5.0	–	2.41	1.42	1.67	4.10	1.14	1.22
10	–	5.74	1.16	1.62	2.98	0.90	1.72
20	–	10.4	5.3	7.8	9.6	7.0	5.6
30	–	25.4	8.2	10.6	23.6	8.6	8.2
0.5	10	2.02	0.89	1.03	1.85	0.93	0.86
0.5	20	2.77	0.37	0.45	0.57	0.51	0.49
0.5	40	7.19	2.23	2.71	3.47	1.96	2.40
0.5	80	12.5	4.0	4.6	7.9	3.4	3.5
5.0	40	14.4	0.9	1.6	1.8	1.2	1.1
10	40	13.4	1.0	1.9	2.0	1.9	1.6
20	40	37.5	3.3	4.0	6.4	3.8	4.1
30	40	62.2	12.7	14.3	22.6	13.8	13.2
30	10	30.8	4.1	5.5	6.9	4.9	4.9
30	20	37.9	5.6	7.6	8.2	7.0	7.0
30	60	59.3	14.0	15.9	25.8	11.7	11.7
30	80	98.3	13.2	25.3	11.4	15.4	16.3
Correlation with column 1*				0.95	0.91	0.98	0.98
Correlation with column 2*					0.77	0.97	0.98

* Pearson correlation coefficient.

simply because it is computationally economical. Although the values in column 3 are noticeably different from those in columns 1, 2, 4 and 5, they are remarkably well correlated. If regressions are run of the columns against each other, it becomes clear that the values in column 3 are, relative to those in the other columns, consistently lower for C_f (factor of 0.54–0.69-fold) and higher for C_2 (factor of 1.8–2.9-fold). The remaining columns do not show any strong patterns when compared with each other.

Shown in Table 6 are the Dalziel constants and their standard deviations, calculated from the values and variances of C_f, C_1 and C_2 in Tables 2–5 by using the relationships shown in eqns. (2)–(4). Again, methods 1, 2, 4 and 5 produce the same result; method 3 is close. The distributions of the residuals for C_f and C_1 are shown in Fig. 2.

In the accompanying paper (Cox & Boeker, 1987) we conclude, on the basis of a t test, that C_2 is not significantly different from 0 for 19 of the 21 progress curves. Depending on the final calculation desired, this result can mean that the minimum procedure required is to estimate C_f, C_1 and C_2 by non-linear regression, obtain the jackknife variance estimate by resampling the data according to the number of points, calculate t for C_2, and then repeat the fitting procedure and variance calculation with C_2 held at 0.

A possible way to reduce this cumbersome calculation is to calculate t for C_2 using analysis of variance, thereby avoiding the need for the first resampling procedure. The problem with analysis of variance in this system is that the sums of squares are not additive, i.e.:

$$\Sigma(\Delta P_i - \Delta \hat{P}_i)^2 + \Sigma\,(\Delta \hat{P}_i - \Delta \overline{P})^2 \neq \Sigma\,(\Delta P_i - \Delta \overline{P})^2$$

However, the sums of squares are close to being additive, suggesting that a properly adjusted analysis of variance might still be of value. For example, for the first time course in Tables 3–5, $\Sigma(\Delta P_i - \Delta \hat{P}_i)^2 = 0.0043$, $\Sigma(\Delta \hat{P}_i - \Delta \overline{P})^2 = 1.4253$, sum $= 1.4296$, and $\Sigma(\Delta P_i - \Delta \overline{P})^2 = 1.4336$.

An analysis-of-variance calculation that adjusts for non-additivity in the sums of squares is described in the Computational methods section. Values of t calculated in this way are compared with values calculated by using the jackknife estimate of variance in Fig. 3. The two estimates fail to correspond in only three cases, all with t values less than 1.1. For 23 points, the minimum number in these progress curves, the critical value of t for a 95% confidence limit is 2.07: for 90% it is 1.71. The analysis-of-variance test makes no wrong decisions, and it in general corresponds very well with t values calculated from the standard deviation of C_2.

DISCUSSION

The multiple-regression method used throughout the calculations is straightforward and computationally efficient, but can be numerically unstable if the columns of the input matrix are close to being linearly dependent.

Table 4. Values and standard deviations for C_1

The methods, which are described in detail in the text, are: method 1, data simulation with the use of an independent error model; method 2, simulation with the use of a proportional error model; method 3, 'backwards' regression; method 4, the jackknife; method 5, the bootstrap. The units of C_1 and the standard deviations are $\mu g \cdot ml^{-1} \cdot min \cdot mM^{-1}$.

Initial concentration (mM)		Value of	Standard deviations from				
Arginine	Agmatine	C_1	Method 1	Method 2	Method 3	Method 4	Method 5
0.05	–	1.82	7.58	12.4	4.80	10.9	13.6
0.2	–	6.99	1.02	1.30	1.75	1.28	1.05
0.5	–	2.22	0.54	0.76	0.78	0.80	0.72
1.0	–	2.13	0.21	0.29	0.22	0.34	0.25
2.0	–	2.10	0.10	0.14	0.25	0.10	0.12
5.0	–	2.60	0.18	0.24	0.56	0.19	0.17
10	–	1.92	0.09	0.11	0.20	0.09	0.13
20	–	2.91	0.24	0.26	0.32	0.33	0.29
30	–	1.39	0.20	0.27	0.52	0.21	0.20
0.5	10	3.54	1.24	1.44	2.42	1.41	1.35
0.5	20	4.20	0.48	0.66	0.72	0.77	0.68
0.5	40	1.18	3.18	4.44	4.00	3.06	3.76
0.5	80	2.19	5.68	7.28	9.30	5.74	5.53
5.0	40	1.05	0.15	0.24	0.23	0.18	0.18
10	40	2.24	0.08	0.14	0.13	0.13	0.10
20	40	1.30	0.13	0.16	0.19	0.15	0.15
30	40	0.56	0.31	0.38	0.64	0.35	0.31
30	10	1.24	0.10	0.14	0.14	0.13	0.12
30	20	1.29	0.14	0.20	0.17	0.17	0.17
30	60	1.01	0.38	0.39	0.53	0.31	0.28
30	80	1.33	0.34	0.71	0.25	0.40	0.41
Correlation with column 1*				0.91	0.90	0.90	0.87
Correlation with column 2*					0.76	1.00	0.99

* Pearson correlation coefficient.

This is not likely to be a problem in calculating the Dalziel constants for this stoichiometry, but could be a problem for the non-linear regressions, where the product terms corresponding to C_f, C_1 and C_2 are in fact highly correlated. Instability of this sort shows up as bad corrections in the non-linear regression, leading to divergence rather than convergence. For the 2100 calculations performed for the bootstrap estimation of variance, for example, only six failed to converge. The regression technique used is evidently adequate for these data. More stable, but also more complex, regression techniques are available if they are required in the future (Lawson & Hanson, 1974; Rice, 1981).

In Table 6, $J_{AP}/J_A k_{cat.}$ is a negative number, regardless of the computational method. This is of course physically impossible; $J_{AP}/J_A k_{cat.}$ is the uncompetitive inhibition constant for CO_2. For the experiments analysed here, where CO_2 was never addded initially, $J_{AP}/J_A k_{cat.}$ can only be calculated from C_2; see eqn. (4). In the accompanying paper (Cox & Boeker, 1987) we conclude that C_2 has a significant value in at most two of the progress curves; use of C_2 values in further calculations is not justified. I have considered them in the present paper only in order to make the fullest possible analysis of the various computational methods.

The calculations required to extract kinetic constants from integrated rate equations are complex. This is due to the need for non-linear regression in the first part of the program, and especially to the demand for variances along with values of C_f, C_1 and C_2. The multiple regression in the second part of the calculation can be performed with any standard statistical package.

The complex nature of the calculations need not pose a barrier to the use of complete progress curves. The most suitable of the techniques examined in the present paper have been incorporated into two computer programs, one for extracting C_f, C_1 and C_2 and their variances from a progress curve, and a second for calculating Dalziel constants and variances from a set of curves. The programs were originally written in FORTRAN and implemented on a VAX 11/780 computer, but have been rewritten in BASIC and can be run on an IBM-PC or similar small computer. They apply specifically to the stoichiometry $A \rightarrow P+Q$ with a discontinuous assay, but could presumably be adapted for use with a continuous assay. They are available from the author on request.

The first of these programs carries out a non-linear regression to obtain C_f, C_1 and C_2, tests C_2 for significance by using analysis of variance, repeats the fit with $C_2 = 0$ if necessary, and performs a jackknife calculation of variance. The jackknife was chosen for two reasons. It is non-parametric, and therefore avoids any assumptions due to normality or problems due to outlying data points. And it is more conservative of computer time than any of the remaining three successful methods. Each of these requires that the data be simulated or resampled some large number of times, whereas the jackknife requires only as many resamplings as there are points, and even this could possibly be reduced by

Table 5. Values and standard deviations for C_2

The methods, which are described in detail in the text, are: method 1, data simulation with the use of an independent error model; method 2, simulation with the use of a proportional error model; method 3, 'backwards' regression; method 4, the jackknife; method 5, the bootstrap. The units of C_2 and the standard deviations are $\mu g \cdot ml^{-1} \cdot min \cdot mM^{-2}$.

Initial concentration (mM)		Value of	Standard deviation from				
Arginine	Agmatine	C_2	Method 1	Method 2	Method 3	Method 4	Method 5
0.05	–	339	484	651	260	583	696
0.2	–	35.7	32.5	32.5	26.5	35.1	29.7
0.5	–	0.99	5.32	6.55	4.24	7.27	6.63
1.0	–	−0.21	1.13	1.36	0.69	1.49	1.27
2.0	–	0.095	0.29	0.35	0.32	0.34	0.33
5.0	–	0.029	0.27	0.26	0.27	0.23	0.24
10	–	−0.012	0.051	0.061	0.052	0.038	0.063
20	–	−0.098	0.066	0.080	0.047	0.080	0.076
30	–	−0.022	0.038	0.040	0.045	0.043	0.040
0.5	10	3.01	14.2	14.1	12.7	13.1	11.9
0.5	20	3.50	6.16	6.09	4.60	6.92	6.74
0.5	40	−8.81	32.8	30.7	27.5	27.6	33.7
0.5	80	−9.50	58.0	54.2	62.0	44.6	47.6
5.0	40	−0.53	0.14	0.18	0.14	0.16	0.14
10	40	−0.035	0.041	0.058	0.034	0.066	0.054
20	40	−0.16	0.033	0.030	0.030	0.035	0.036
30	40	−0.062	0.054	0.046	0.051	0.050	0.051
30	10	−0.022	0.020	0.020	0.014	0.019	0.020
30	20	−0.024	0.026	0.027	0.017	0.028	0.028
30	60	−0.049	0.053	0.055	0.054	0.053	0.053
30	80	−0.092	0.045	0.071	0.023	0.050	0.052
Correlation with column 1*				1.00	0.99	1.00	1.00
Correlation with column 2*					0.98	1.00	1.00

* Pearson correlation coefficient.

using the grouped jackknife (Cornish-Bowden & Wong, 1978).

Computationally, the most conservative method for obtaining variances is the 'backwards' regression of t on ΔP. Although not theoretically sound, this method produced surprisingly good results. If computer speed is a serious problem, variances obtained by this method would be preferable to no variances at all.

The second program calculates Dalziel constants by weighted multiple regressions (eqns. 2–4), taking into account the possibility that the initial conditions may be such that only A_0 is varied, or only A_0 and P_0, or etc.;

Table 6. Dalziel values

The methods, which are described in detail in the text, are: method 1, data simulation with the use of an independent error model; method 2, simulation with the use of a proportional model; method 3, 'backwards' regression; method 4, the jackknife; method 5, the bootstrap.

Dalziel parameter	Units	Mean value ± standard deviation of mean, obtained by the use of variances from				
		Method 1	Method 2	Method 3	Method 4	Method 5
$1/k_{cat.}$	$(\mu mol/min \text{ per } \mu g)^{-1}$	2.59±0.23	2.67±0.25	2.19±0.23	2.56±0.20	2.70±0.26
$J_0/J_A k_{cat.}$	$mM \cdot (\mu mol/min \text{ per } \mu g)^{-1}$	1.01±0.31	0.91±0.26	1.49±0.25	0.93±0.27	0.81±0.28
$J_P/J_A k_{cat.}$	$(\mu mol/min \text{ per } \mu g)^{-1}$	0.103±0.026	0.099±0.022	0.077±0.025	0.116±0.023	0.118±0.024
$J_Q/J_A k_{cat.}$	$(\mu mol/min \text{ per } \mu g)^{-1}$	0.16±0.18	0.20±0.16	−0.09±0.20	0.17±0.12	0.20±0.17
$J_{PQ}/J_A k_{cat.}$	$mM^{-1} \cdot (\mu mol/min \text{ per } \mu g)^{-1}$	0.021±0.004	0.020±0.004	0.026±0.004	0.020±0.003	0.020±0.004
$J_{AP}/J_A k_{cat.}$	$mM^{-1} \cdot (\mu mol/min \text{ per } \mu g)^{-1}$	0.028±0.011	0.025±0.013	0.026±0.012	0.023±0.012	0.028±0.012
$J_{AQ}/J_A k_{cat.}$	$mM^{-1} \cdot (\mu mol/min \text{ per } \mu g)^{-1}$	−0.055±0.016	−0.055±0.018	−0.052±0.015	−0.047±0.017	−0.054±0.017
Correlation with column 1*			1.00	0.96	1.00	1.00
Correlation with column 2*				0.94	1.00	1.00

* Pearson correlation coefficient.

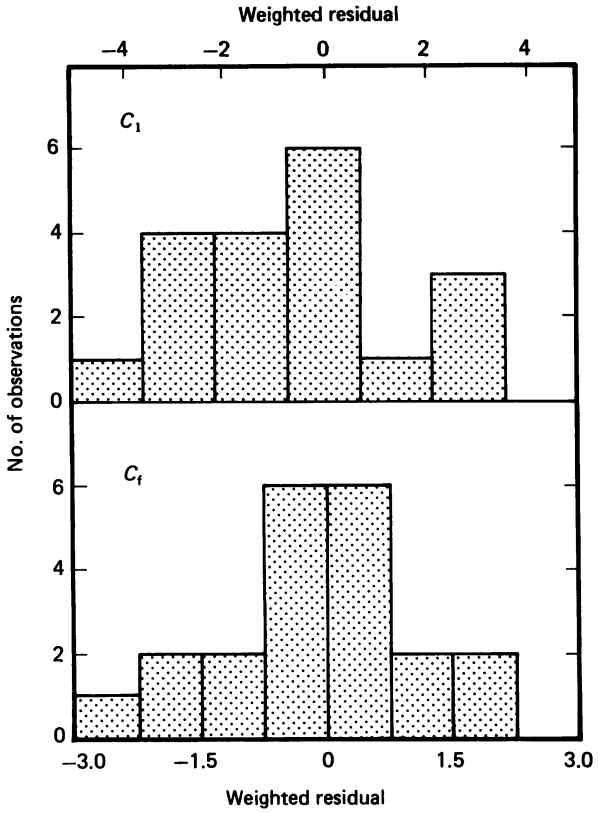

Fig. 2. Distribution of weighted residuals when the Dalziel constants are obtained by fitting eqns. (2) and (3)

The quantities plotted on the ordinate are $(C_i - \hat{C}_i)/s_{C_i}$.

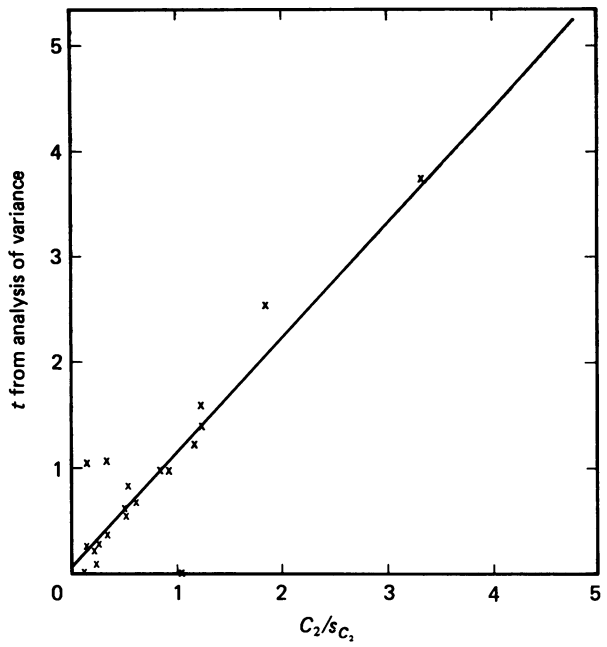

Fig. 3. Relationship between t values calculated for C_2 from jackknife variance values (C_2/s_{C_2}) and analysis of variance

Details of the calculations are given in the Computational methods section. The regression line is $y = 0.07 + 1.09\,x$.

it extracts whatever information is possible for the conditions used. The advantage of these programs is that standard deviations are automatically calculated for the final Dalziel constants; an effort has been made to reduce the subjectivity of decisions made about terms in the empirical rate equation.

This work was supported by Grant GM34065 from the National Institutes of Health. I thank Dr. Ronald Canfield for statistical help.

REFERENCES

Boeker, E. A. (1984*a*) Experientia **40**, 453–456
Boeker, E. A. (1984*b*) Biochem. J. **223**, 15–22
Boeker, E. A. (1985) Biochem. J. **226**, 29–35
Cornish-Bowden, A. J. & Wong, J. T.-F. (1978) Biochem. J. **175**, 969–976
Cox, T. T. & Boeker, E. A. (1987) Biochem. J. **245**, 59–65
Dalziel, K. (1957) Acta Chem. Scand. **11**, 1706–1723
Darvey, I. G., Shrager, R. & Kohn, L. D. (1975) J. Biol. Chem. **250**, 4696–4701
Duggleby, R. G. (1979) Biochem. J. **181**, 255–256
Duggleby, R. G. (1980) Eur. J. Biochem. **109**, 93–96
Duggleby, R. G. (1981) Anal. Biochem. **110**, 9–18
Duggleby, R. G. & Morrison, J. F. (1977) Biochim. Biophys. Acta **481**, 297–312
Duggleby, R. G. & Morrison, J. F. (1978) Biochim. Biophys. Acta **526**, 398–409
Efron, B. (1982) The Jackknife, the Bootstrap, and other Resampling Plans, Society for Industrial and Applied Mathematics, Philadelphia
Fernley, H. H. (1974) Eur. J. Biochem. **43**, 377–378
King, E. L. & Altman, C. (1956) J. Phys. Chem. **60**, 1375–1378
King, J. T. (1984) Introduction to Numerical Computation, McGraw-Hill, New York
Lawson, C. L. & Hanson, R. J. (1974) Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs
Matyska, L. & Kovář, J. (1985) Biochem. J. **231**, 171–177
Neter, J. & Wasserman, W. (1974) Applied Linear Statistical Models, R. D. Irwin, Homewood
Rice, J. R. (1981) Matrix Computations and Mathematical Software, McGraw-Hill, New York
Storer, A. C., Darlison, M. G. & Cornish-Bowden, A. J. (1975) Biochem. J. **151**, 361–367
Wilkinson, G. N. (1961) Biochem. J. **80**, 324–332

Received 4 August 1986/8 December 1986; accepted 2 March 1987