

# Fitting ecological and physiological data to rectangular hyperbolae: a comparison of methods using Monte Carlo simulations

John A. Berges\*, David J. S. Montagnes\*\*, C. L. Hurd, P. J. Harrison

Department of Oceanography, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

**ABSTRACT:** Recommendations were developed to help aquatic scientists select which curve fitting method should be used to fit data that are expected to follow rectangular hyperbolic relationships. Rectangular hyperbolae of the form  $V = (V_{\max}S)/(K_m + S)$ , where  $V$  is a biological rate and  $S$  is the concentration of some substrate, are widely used by researchers to model the kinetics of processes such as enzyme activity versus substrate concentration, nutrient uptake versus nutrient concentration, and grazing and growth rate versus prey concentration. A variety of procedures exist to estimate the parameters  $V_{\max}$  (the rate of the process at saturating substrate concentration) and  $K_m$  (the concentration of  $S$  at which half the saturated rate is achieved). There has been extensive discussion in the biochemical and ecological literature as to which fitting method is most appropriate, based largely on theoretical and statistical considerations. However, the assumptions inherent in these fitting procedures are typically violated by the data obtained in many field and laboratory studies, e.g. the measurement of  $S$  has an associated error, or error levels in the measurement of  $V$  may not be constant across  $S$ . Thus, there is a problem predicting *a priori* which fitting method should be used. In this study, this problem was approached using Monte Carlo simulations. Data sets with known  $V_{\max}$  and  $K_m$  were constructed for 5 different data cases, ranging from data sets where saturation was not achieved to data sets where very few sub-saturated measurements were available. Random, normally distributed errors were assigned to each point based on a 10%, 20% or 50% constant or variable error in the estimate of  $V$ , or 20% error in both  $S$  and  $V$ . Six fitting procedures were applied including linear methods (Lineweaver-Burk; Eadie-Hofstee; Hanes-Woolf), the median method (Eisenthal and Cornish-Bowden), and non-linear least-squares methods (Cleland-Wilkinson; Tseng-Hsu). Non-linear methods were generally superior, but for data sets with low error (10%) all methods gave almost equally accurate results. Data with constant error were more difficult to fit than those where error varied with  $V$ . Criteria for selecting a fitting method based on data characteristics are discussed and applied to actual data sets.

**KEY WORDS:** Michaelis-Menten model · Curve fitting · Enzyme activity · Nutrient uptake kinetics · Plankton feeding · Bias · Experimental design

## INTRODUCTION

Rectangular hyperbolic models, often referred to as Michaelis-Menten (Michaelis & Menten 1913), Monod (1942), Droop (1983) and Holling (1966) Type II models, are widely used by biologists working in aquatic

systems to model processes such as enzyme kinetics (e.g. Packard 1979), heterotrophic microbial activity (e.g. Li 1983), nutrient uptake or growth rate of autotrophs (e.g. MacIsaac & Dugdale 1969), and ingestion, grazing or growth rates of unicellular and multicellular heterotrophs (e.g. Mullin et al. 1975, Verity 1991). Such hyperbolae are typically of the form:

$$V = (V_{\max} S)/(K_m + S) \quad (1)$$

where  $V$  is the biological rate,  $S$  is the concentration of some substrate required in the process, and the fitted constants  $V_{\max}$  and  $K_m$  represent the maximal rate

Present addresses:

\*Oceanic and Atmospheric Sciences Division, Brookhaven National Laboratory, Upton, New York 11973, USA

\*\*School of Oceanography WB-10, University of Washington, Seattle, Washington 98195, USA

at substrate concentration and the substrate concentration at half the maximal rate, respectively.

A variety of methods exist to fit this equation to a data set and include using transformations to linearize the data and iterative non-linear procedures. It has been noted that fitting data using different methods can give different, and sometimes contradictory, results (e.g. Packard 1979, Hurd & Dring 1990). Thus, it is important to determine which method is best for a given data set. Moreover, such information is essential to the correct interpretation of the parameters in published data sets that have been fitted using different methods. There are extensive discussions to guide the investigator in the choice of a fitting procedure in the biochemical and physiological literature, suggesting for example that many fitting methods using linear transformations are biased (see Wilkinson 1961, Cleland 1967, Eisenthal & Cornish-Bowden 1974, Lam 1981, Nelder 1991, Johnson 1992b). However, the data typically obtained from aquatic systems violate several key assumptions of commonly used fitting procedures: (1) errors in the measurement of  $V$  can be high (>10% coefficient of variation) and can be either constant with  $S$  or can vary in proportion to  $S$ ; (2)  $S$  itself is not measured without error; and (3) data are available for limited ranges of  $S$ . Depending on the procedure used, these violations of assumptions may introduce significant error into determination of  $V_{\max}$  and  $K_m$ . There have been considerations of these problems in which field data have been used to assess the accuracy of different procedures (e.g. Mullin et al. 1975, Li 1983). Ultimately, however, such an approach relies on a subjective judgment of which fit is most 'correct', since the actual parameters are unknown.

Alternatively, a number of biochemists have attempted to use Monte Carlo approaches to the problem, by creating data sets with known  $V_{\max}$  and  $K_m$  and adding random errors (e.g. Cornish-Bowden et al. 1978, Currie 1982, Tseng & Hsu 1990). These studies are of limited value to aquatic researchers because they have focused on data sets with unrealistically small errors, they have ignored error in  $S$ , they have used regular spacing of data points across  $S$ , and/or they have compared only a limited number of linear and non-linear fitting methods. Moreover, many of these studies are largely statistical in nature, and are published in areas of the literature not usually read by aquatic researchers.

In this investigation, randomly generated data sets with known  $V_{\max}$  and  $K_m$  and with different error levels and distributions of points were used in order to compare 6 fitting procedures. In each case, results were evaluated in terms of both accuracy (i.e. the distance of the parameter estimates from the true values), and precision (i.e. the percentage of estimates falling within a

certain range of the true value). Generalizations were then developed for different types of data sets, and validated using data from a range of aquatic experiments.

## MATERIALS AND METHODS

**Data generation.** Five data cases were devised (Fig. 1): Case 1, where data were geometrically distributed across  $S$ ; Case 2, where there were no data for  $S$  less than  $K_m$ ; Case 3, where there were no data for  $S$  greater than  $K_m$ ; Case 4, where there were data only for  $S$  greater than  $2 \times K_m$  or less than  $0.5 \times K_m$ ; and Case 5, where all data fell between  $2 \times K_m$  and  $0.5 \times K_m$ . For each case, data sets were generated with  $V_{\max} = 10$  and  $K_m = 2$ . Random errors (generated using a SYSTAT procedure; Wilkinson 1990) were assigned in an additive manner:

$$V = (V_{\max} S)/(K_m + S) + \varepsilon \quad (2)$$

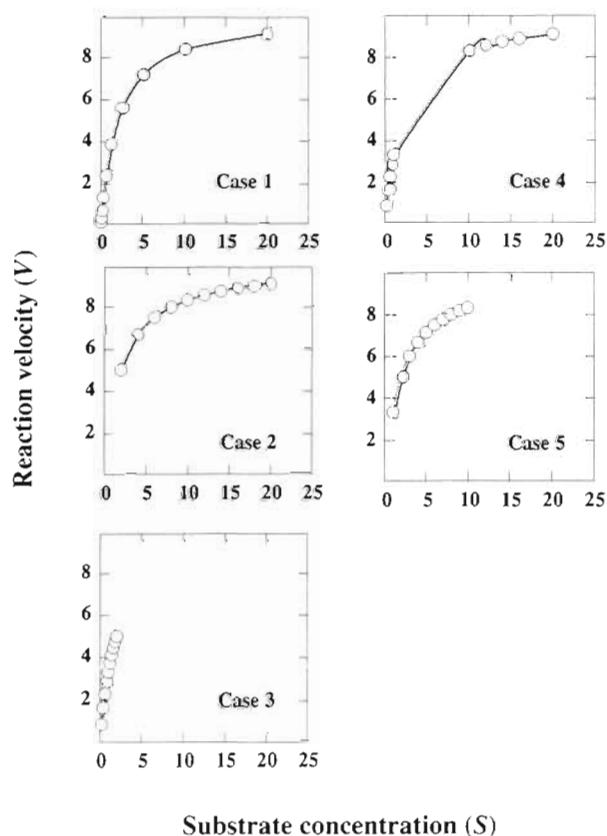


Fig. 1. Data cases considered. Cases represent geometrically distributed data (Case 1), data where no points are lower than  $K_m$  (Case 2), data where no points are higher than  $K_m$  (Case 3), data where only points higher than  $2 \times K_m$  or lower than  $0.5 \times K_m$  are available (Case 4), and data where all points fall between  $2 \times K_m$  and  $0.5 \times K_m$  (Case 5). In each case, data sets of 10 points were generated with  $V_{\max} = 10$  and  $K_m = 2$

where the error,  $\varepsilon$ , was normally distributed with mean of zero. In one series of data sets, errors were added such that errors were a constant percentage of  $V_{\max}/2$ , either 10, 20, or 50% in  $V$  alone, or 20% (independent) error in both  $S$  and  $V$  (see Fig. 2). This was termed a constant error; the standard deviation in  $V$  was constant regardless of  $S$  (analogous to the constant errors used by Dowd & Riggs 1965, or the 'simple' errors used by Cornish-Bowden et al. 1978). In the second series of data sets, errors were a constant fraction of the value of  $V$ , again either 10, 20, or 50% in  $V$  alone, or 20% (independent) error in both  $S$  and  $V$  (Fig. 2). This was termed a variable error; in this situation the coefficient of variation of  $V$  was constant regardless of  $S$  (analogous to the 'increasing' errors used by Dowd & Riggs 1965, or the 'variable' errors used by Cornish-Bowden et al. 1978). In each case, 100 data sets of 10 points each were generated, i.e. representing 100 individual experiments where  $V$  was measured at 10 different  $S$ -values.

**Fitting procedures.** Fitting procedures were implemented on a 486 AT-type personal computer using software written by the authors or modified from published software, as noted below. Six basic procedures were compared. The names describing each method were chosen on the basis of common use in the literature; Cornish-Bowden (1979) provides a discussion of priorities.

*Lineweaver-Burk (LB; Lineweaver & Burk 1934).* This double-reciprocal transformation of Eq. (1) (represented graphically by a plot of  $1/V$  versus  $1/S$ ) gives:  $1/V = (K_m/V_{\max})(1/S) + (1/V_{\max})$ , which represents a straight line with slope  $(K_m/V_{\max})$  and intercept  $(1/V_{\max})$ . Slope and intercept were estimated by least-squares linear regression (Steel & Torrie 1980) and parameters calculated. The regressions were not weighted.

*Eadie-Hofstee (EH; Hofstee 1959).* This rearrangement of Eq. (1) gives:  $V = -K_m(V/S) + V_{\max}$  (represented graphically as a straight line on a plot of  $V$  versus  $V/S$ , and also known as an Augustinsson plot) from which parameters are estimated as for the LB method.

*Hanes-Woolf (HW; Hanes 1932).* This rearrangement of Eq. (1) gives:  $S/V = K_m/V_{\max} + S(1/V_{\max})$ , (represented graphically as a straight line on a plot of  $S/V$  versus  $S$ , and referred to also as a Hanes plot or a Woolf plot) from which parameters may be estimated as above.

*Eisenthal and Cornish-Bowden (EC; Eisenthal & Cornish-Bowden 1974).* In this median method, the points of intersection of the lines defined for each ( $V$ ,  $S$ ) pair by  $y = (V/S)x + V$  are calculated. The median value of  $x$  corresponds to  $K_m$ , and the median value of  $y$  corresponds to  $V_{\max}$ . The program was adapted from Myers et al. (1990), with improvements to the sorting routine.

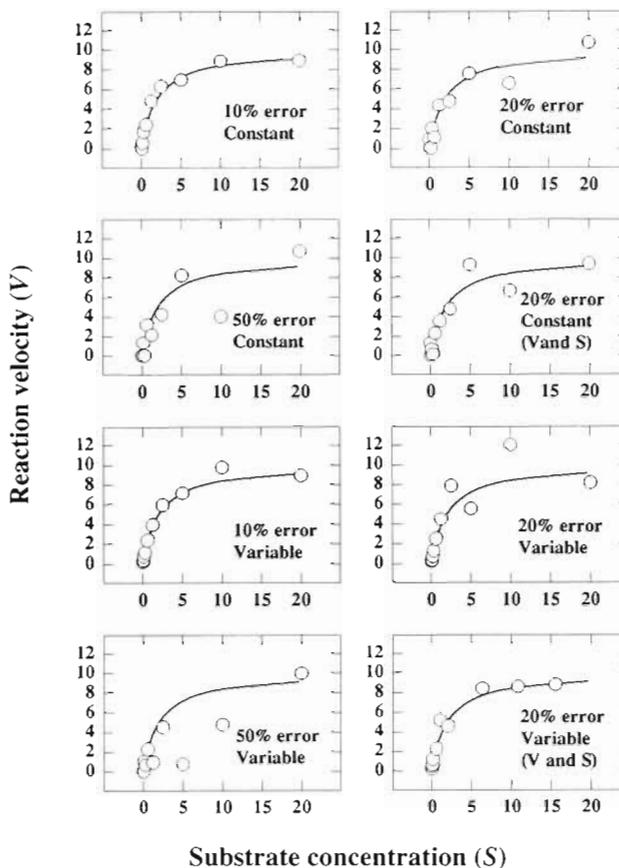


Fig. 2. Examples of error levels (as percentages of  $V$  and  $S$ ) assigned for Case 1. Constant error levels are set as percentages of  $0.5 \times V_{\max}$ . Errors were assigned in a normal distribution

*Cleland-Wilkinson (W; Wilkinson 1961).* This method first estimates values of  $V_{\max}$  and  $K_m$  from a linear plot (identical to HW), then uses an iterative non-linear method to directly fit the equation to the data using least squares as a criterion (see Johnson 1992b).

*Tseng-Hsu (TH; Hsu & Tseng 1989).* This method uses a least-squares criterion as for the W method, but uses a random search technique. First, estimates were made of the ranges in which  $V_{\max}$  and  $K_m$  occurred (0 to 10 for  $K_m$  and 0 to 50 for  $V_{\max}$ ). Within this range, 500 random pairs of points were picked. Least squares were calculated for each pair of estimates and the 10  $V_{\max}$ - $K_m$  pairs giving the smallest residuals were identified. The search range was redefined based on the variance of the best 10 parameter pairs and the process was repeated up to 100 times, or until no further changes in parameter estimates were found.

In addition to the latter 2 non-linear methods, selected data were fitted using different algorithms and commercially available packages including Simplex and Semi-Newton (SYSTAT NLIN procedure;

Wilkinson 1990), and Marquardt-Levenberg (Sigma-plot 5.0, Jandel Scientific) methods.

**Real data sets.** The above fitting procedures were also applied to 3 real data sets, for comparison: (1) nitrate reductase activity versus nitrate concentration in a marine microalga (Berges 1993), representing data with low, constant error; (2) phosphate uptake versus phosphate concentration in a marine macroalga (Hurd & Dring 1990), representing data with high, variable error; and (3) growth rate versus prey concentration in a marine ciliate (Montagnes 1993, note these data were translated to give x- and y-intercepts of zero), representing data with high, constant error.

**RESULTS**

Table 1 shows typical results of the 6 fitting methods for data generated for Cases 1 and 3. These cases illustrate the quality of the data, and were selected because they show the best (Case 1) and the poorest (Case 3) fits. It was also found that parameters estimated using three non-linear fitting routines in 2 commercial packages corresponded almost exactly to those found for the TH method; differences were found only in third or fourth decimal places of parameter estimates. Thus, only the data from the TH method are presented in the following results. Median values were chosen for presentation because fre-

Table 1. Results of model-fitting procedures (Lineweaver-Burk, LB; Eadie-Hofstee, EH; Hanes-Woolf, HW; Eisenthal and Cornish-Bowden, ECB; Cleland-Wilkinson, W; or Tseng-Hsu, TH) for Case 1 (CS1) and Case 3 (CS3) data. Notation describes error as constant (C) or variable (V) and error levels as a percentage (10, 20 or 50). Error in both S and V is denoted XY Values represent medians of estimated parameters of the data. The true values are  $V_{max} = 10$ , and  $K_m = 2$ . Numbers in parentheses represent the percentages of estimates which fell outside the ranges of 1.2 to 2.8 for  $K_m$  and 6 to 14 for  $V_{max}$

	Estimated $V_{max}$						Estimated $K_m$					
	LB	EH	HW	ECB	W	TH	LB	EH	HW	ECB	W	TH
CS1C10	0.03 (96)	5.37 (72)	-1.35 (65)	10.15 (6)	10.03 (0)	10.03 (0)	0.09 (98)	0.48 (97)	-15.60 (87)	2.06 (35)	1.97 (1)	1.97 (1)
CS1C20	0.02 (98)	4.72 (92)	-0.43 (72)	9.90 (19)	10.09 (1)	10.09 (1)	0.11 (93)	0.15 (99)	-14.50 (77)	1.65 (66)	1.98 (22)	1.98 (22)
CS1C50	0.01 (100)	4.32 (95)	-0.36 (95)	10.43 (39)	10.00 (35)	10.00 (13)	0.12 (96)	0.02 (100)	-14.73 (100)	1.26 (76)	1.64 (42)	1.68 (56)
CS1CXY20	0.02 (98)	4.72 (92)	-0.43 (72)	9.90 (19)	10.09 (17)	10.09 (1)	0.11 (93)	0.15 (99)	-14.50 (77)	1.65 (66)	1.98 (27)	1.98 (22)
CS1V10	9.97 (35)	9.76 (0)	10.07 (0)	10.14 (0)	10.12 (0)	10.12 (0)	2.04 (46)	1.92 (0)	2.07 (0)	2.02 (0)	2.05 (17)	2.05 (5)
CS1V20	6.99 (56)	8.81 (1)	9.94 (2)	10.10 (6)	10.27 (25)	10.27 (5)	1.32 (65)	1.65 (5)	2.19 (10)	2.03 (11)	2.10 (29)	2.10 (34)
CS1V50	1.08 (91)	5.76 (54)	8.10 (59)	11.35 (37)	11.48 (52)	10.67 (31)	0.21 (90)	0.74 (79)	2.42 (86)	2.09 (51)	2.71 (44)	2.31 (75)
CS1VXY20	6.59 (69)	8.03 (2)	9.87 (4)	9.74 (5)	10.10 (7)	10.10 (4)	1.08 (74)	1.33 (34)	2.05 (68)	2.05 (14)	1.92 (32)	1.92 (32)
CS3C10	5.11 (73)	4.96 (74)	9.17 (48)	10.43 (29)	10.40 (21)	10.60 (21)	0.57 (80)	0.47 (95)	1.69 (61)	2.06 (51)	2.05 (47)	2.08 (47)
CS3C20	3.64 (86)	3.15 (97)	5.57 (71)	8.20 (56)	9.37 (83)	10.15 (52)	0.29 (90)	-0.03 (100)	0.59 (73)	1.16 (77)	1.86 (72)	1.93 (72)
CS3C50	0.00 (97)	1.82 (100)	-0.01 (96)	4.56 (71)	9.63 (72)	11.02 (68)	-0.19 (96)	-0.43 (100)	-1.58 (96)	0.09 (88)	1.50 (83)	1.72 (82)
CS3CXY20	3.64 (86)	3.15 (97)	5.57 (71)	8.20 (56)	9.37 (53)	10.15 (52)	0.29 (90)	-0.03 (100)	0.59 (73)	1.16 (77)	1.86 (72)	1.93 (72)
CS3V10	10.07 (21)	8.40 (1)	10.35 (5)	10.40 (7)	10.27 (15)	10.29 (9)	2.04 (33)	1.53 (23)	2.14 (19)	2.08 (16)	2.06 (30)	2.11 (28)
CS3V20	8.57 (39)	5.08 (76)	9.47 (32)	10.06 (33)	10.67 (56)	10.11 (37)	1.64 (53)	0.55 (94)	1.90 (39)	2.03 (48)	2.22 (60)	2.07 (57)
CS3V50	1.87 (84)	1.54 (99)	1.39 (86)	8.08 (72)	10.33 (84)	11.60 (69)	0.27 (86)	-0.50 (100)	0.13 (86)	1.33 (76)	2.31 (81)	2.37 (81)
CS3VXY20	7.65 (43)	4.66 (89)	8.00 (28)	8.56 (29)	8.01 (31)	8.12 (28)	1.32 (64)	0.36 (98)	1.42 (30)	1.58 (49)	1.30 (64)	1.34 (62)

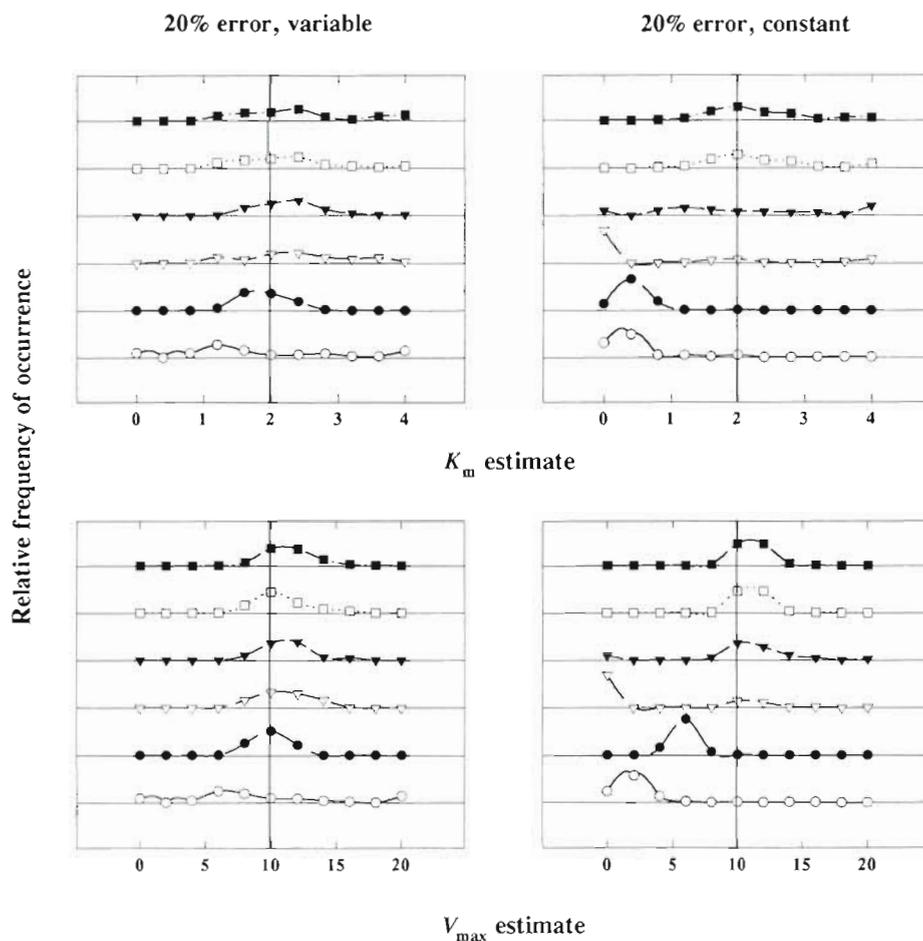


Fig. 3. Examples of frequency distributions of  $V_{\max}$  and  $K_m$  estimates for various fitting procedures for Case 1 data with 20% variable or constant error. Procedures are: (○) Line-weaver-Burk, (●) Eadie-Hofstee, (▽) Hanes-Woolf, (▼) Eisenthal and Cornish-Bowden, (□) Cleland-Wilkinson, and (■) Tseng-Hsu. Y-axis scale is relative percentage. True values of  $V_{\max}$  and  $K_m$  are 10 and 2, respectively

quency distributions of estimates were often asymmetrical (Fig. 3).

Table 2 summarizes the results of fitting procedures by case and by error levels. In order to rank the success of fitting procedures, 2 criteria were selected: the accuracy of the estimate (i.e. how close the median estimate was to the true value) and the precision of the estimate (i.e. what percentage of estimates fell within a specified range). Enzyme activity and growth rate data often give coefficients of variation of 20% or less. Thus a 20% or smaller departure of the median estimate from the true parameter estimate was considered acceptable. High variability is also a common feature of biological data, so precision was evaluated based on the percentage of estimates that fell outside a range of  $\pm 20\%$  of the true parameter value, i.e. 8 to 12 for  $V_{\max}$  and 1.6 to 2.4 for  $K_m$ . In Table 2, median estimates that fell within 20% of the true value and in which fewer than 50% of the 100 estimates were outside the respective range were marked with a single asterisk. If less than 20% of the 100 estimates fell outside the range a double asterisk was used.

In general, linear models performed well only at low error (10%) and when error varied with  $V$ . The EH

technique seldom predicted  $V_{\max}$  within the stated criteria of accuracy and precision, but was more successful for predicting  $K_m$ . LB and HW methods estimated both parameters equally well.

The ECB method was more successful than linear methods as error increased, but W and TH methods showed the greatest level of accuracy and precision in predicting  $K_m$  and  $V_{\max}$ . Even for these non-linear least-squares methods, however, parameters were poorly estimated at high error levels. The estimates were worst when few points at high  $S$  were available (e.g. Cases 3 and 5). In contrast to linear methods, the constant error cases we investigated in the present study appeared to be fitted better than variable cases for non-linear methods, particularly at high error levels. The TH method was marginally better in terms of accuracy of estimates than W, but this was always in cases where initial estimates for W (derived from a linear method estimate in the procedure used) were substantially in error. The methods were equivalent when suitable initial estimates were provided.

Despite cautions about the effects of error in  $S$  (see Leatherbarrow 1990), the results of analyses with errors in both  $S$  and  $V$  (i.e. the XY error levels) were, in

Table 2. Summary of the results of fitting procedures for  $V_{max}$  and  $K_m$ . Conditions marked with •• indicate cases where the median of the estimates was within 20% of the true parameter values (10 or 2) and less than 20% of the estimates fell outside the ranges of 6 to 14 for  $V_{max}$  or 1.2 to 2.8 for  $K_m$ . Conditions marked with • indicate that the median was within 20% of the true values, and less than 50% of the estimates fell outside the specified ranges

	$V_{max}$ estimates					$K_m$ estimates										
	C10	C20	C50	CXY20	V10	V20	V50	VXY20	C10	C20	C50	CXY20	V10	V20	V50	VXY20
<b>Lineweaver-Burk</b>																
Case 1					•				••	•		•	•			
Case 2	••				••				••	•		•	•			
Case 3					••								•			
Case 4					••				•			••	••	••		
Case 5	••				••				•			••	••	•		
<b>Eadie-Hofstee</b>																
Case 1									••	•		•	••	••		
Case 2	••								••	•		•	••	••		
Case 3													••	••		
Case 4					•				••			••	••	••		
Case 5	••								••			••	••	••		
<b>Hanes-Woolf</b>																
Case 1					••				••			•	••	••		
Case 2	••	••			••				••	•		•	•	•		
Case 3					••				••			•	••	•		
Case 4	••				••				••			••	••	••		••
Case 5	••	••			••				••	•		•	••	•		
<b>Eisenthal-Cornish-Bowden</b>																
Case 1	••	••			••				•			•	••	••		••
Case 2	••	••			••				••	•		•	•	••		•
Case 3					•				•			••	•	•		•
Case 4	••				••				••	•		•	••	••		••
Case 5	••	••			••				••	•		•	••	•		•
<b>Cleland-Wilkinson</b>																
Case 1	••	••	•		••				••	•		•	••	•		•
Case 2	••	••			••				••	•		•	••	••		•
Case 3					••				•			•	•	•		•
Case 4	••	••	••		••				••	•		•	••	••		••
Case 5	••	••			••				••	•		•	••	•		•
<b>Tseng-Hsu</b>																
Case 1	••	••	••		••				••	•		•	••	•		•
Case 2	••	••	••		••				••	•		•	••	••		•
Case 3			•		•				•			•	•	•		•
Case 4	••	••	••		••				••	•		•	••	•		••
Case 5	••	••			••				••	•		•	••	•		•

Table 3. Comparison of estimates of the parameters of rectangular hyperbolae ( $V_{max}$  and  $K_m$ ) for 3 real data sets, using different fitting methods. Fitting procedures are: Lineweaver-Burk (LB), Eadie-Hofstee (EH), Hanes-Woolf (HW), Eisenthal Cornish-Bowden (ECB), Cleland-Wilkinson (W), and Tseng-Hsu (TH). Data sets are pictured in Fig. 4C, and described in the text

Fitting method	NR vs $[\text{NO}_3^-]$		$\text{PO}_4^{3-}$ vs $[\text{PO}_4^{3-}]$		Growth rate vs [prey]	
	$V_{max}$	$K_m$	$V_{max}$	$K_m$	$V_{max}$	$K_m (\times 10^3)$
LB	-0.015	-0.026	18.60	10.02	0.375	0.103
EH	4.153	-0.006	9.82	1.57	1.01	0.295
HW	-0.310	-0.161	11.13	3.52	-9.58	-202
ECB	8.876	0.022	17.07	7.52	1.59	4.89
W	8.119	0.018	15.42	5.39	1.63	6.93
TH	8.119	0.018	15.42	5.39	1.63	6.93

fact, very similar to cases where the same error was introduced only in  $V$  at the same level.

The results of the Monte Carlo simulation were reflected in real data set analyses. Fig. 4A represents enzyme kinetic data with a low (10%), constant error and a distribution similar to Case 1. From Table 2, it would be predicted that such a data set would not be fit well by linear methods. As shown in both Fig. 4A and Table 3, linear estimates are clearly inadequate. If the 2 data points with the lowest  $S$ -values were deleted from this data set, however, all of the linear methods gave results much closer to the non-linear techniques. In the nutrient uptake data set represented in Fig. 4B, data are similar to Case 2, with high (20 to 50%) and variable error. In general, linear techniques predicted lower  $K_m$ -values and lower  $V_{max}$ -values. Table 2 suggests that, while none of the methods is entirely satisfactory, non-linear methods should provide better results in this case. For the growth rate versus prey concentration data, a case of high (20 to 50%) and constant error, linear methods again gave the poorest results (Fig. 4C; Table 3), as predicted from the simulation (Table 2).

## DISCUSSION

The finding that non-linear methods were superior to linear transformation methods is not surprising; this has been pointed out previously (e. g. Wilkinson 1961, Cornish-Bowden 1979, Currie 1982). It should, however, be noted that linear transformation methods do perform well for data with low, variable error. In general, non-linear least-squares methods appear to be the most useful approach to fitting data to rectangular hyperbolae, provided adequate initial parameter estimates are given. In addition to superior accuracy and precision, non-linear models provide greater flexibility

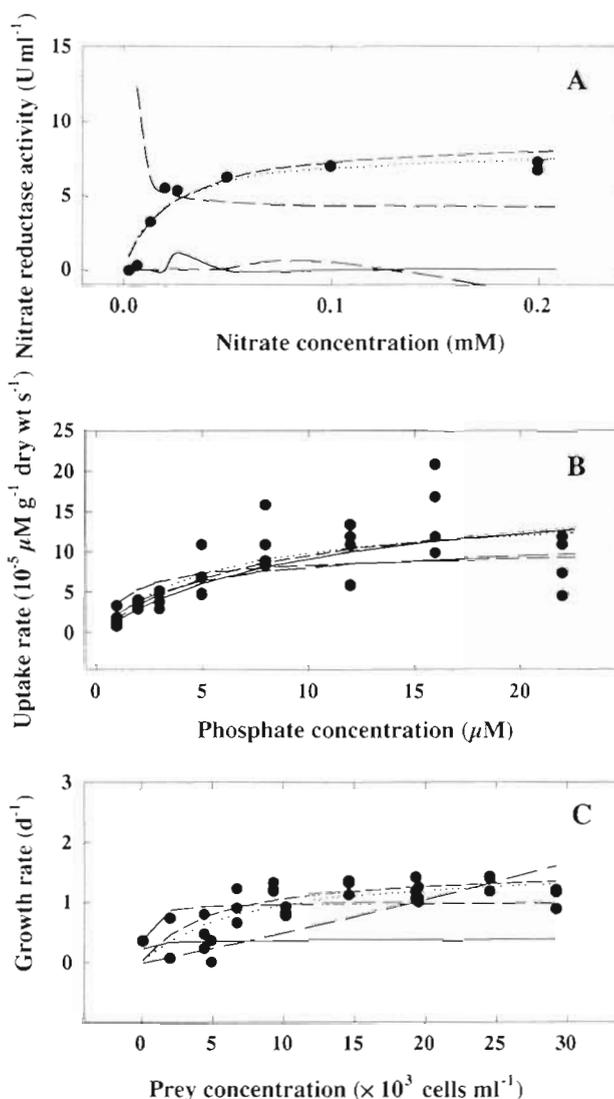


Fig. 4. Examples of real data sets fit to rectangular hyperbolae using different fitting methods: (—) Lineweaver-Burk, (—) Eadie-Hofstee, (—) Hanes-Woolf, (---) Eisenthal and Cornish-Bowden, (····) Cleland-Wilkinson or Tseng-Hsu. (A) Nitrate reductase activity versus nitrate concentration in extracts of the diatom *Thalassiosira pseudonana* (from Berges 1993), (B) phosphate uptake versus concentration in the marine macroalgae *Fucus spiralis* (from Hurd & Dring 1990) and (C) growth rate versus prey concentration for the marine ciliate *Strombidium* sp. feeding on the marine alga *Rhodomonas* sp. (modified from Montagnes 1993). Parameters for each fit are given in Table 3

in that they readily permit consideration of cases where there are non-zero intercepts (e.g. Leatherbarrow 1990), or inhibition constants (e.g. Li 1983), which are problematic to accommodate in linear fitting methods. Of the non-linear fitting methods available, commercially available non-linear fitting packages using methods such as Gauss-Newton, Simplex, or Mar-

quardt-Levenberg may be most suitable for routine use; the agreement of these methods is not unusual because they use the same criterion (least squares) for determining the best fit. They differ in the algorithm used to find the best fit, and thus in relative efficiency and complexity of manipulations. Users are likely to select a method based on computing power available. For example, while the TH method may provide good parameter estimates, it is inefficient and difficult to apply to data sets larger than 25 points (see Cornish-Bowden 1991). The principle disadvantages of non-linear versus linear methods are that the non-linear methods are computationally complex and that they are sensitive to initial parameter estimates (see Jones & Taransky 1991). The former problem is no longer difficult to overcome, given the ready availability of powerful personal computers (see Duggleby 1991), but the latter problem remains very real. In the present study, for example, the W techniques sometimes suffered from poor initial estimates.

Generalizing the results of the present study may be limited due to the number of cases considered. Clearly, given the wide variety of applications of rectangular hyperbolic models, the number of possible patterns of error distribution are enormous. We did not, for example, examine how readily the results of this simulation can be applied to a case where error decreases with  $S$  or was Poisson distributed. Thus, an investigator with data displaying a radically different error distribution should consider comparing different non-linear fitting methods.

#### **Error of the parameter estimates and statistical considerations**

Another potential advantage of non-linear methods is that they can provide error estimates for kinetic parameters that are necessary for statistical evaluation of data. However, for non-linear methods there is no exact theory to calculate a confidence interval. Many fitting packages provide 'asymptotic standard errors', but these are probably inaccurate in many cases. Variance/covariance matrices and Monte Carlo simulations to find joint confidence intervals are preferred, but more complex methods (Jones & Taransky 1991, Johnson 1992b). Cornish-Bowden et al. (1978) have shown that distribution-free confidence limits can also be derived for data fitted using the ECB method. However, for the data considered in the present study, the problems with the ECB method itself make this approach less desirable for certain data sets. A better strategy might be to replicate the entire experiment and derive parameter estimates (but not confidence limits) from each replicate. The mean and standard

error of these estimates could then be used to estimate the true parameters and their error

In some cases, it might be useful to consider for what purpose the confidence limits are needed. If, for example, the purpose is to demonstrate a difference between 2 curves in which the values of the parameters are not important in themselves (or the precise form of the relationship is not known), this may be more efficiently done using an analysis of variance (ANOVA; Johnson 1992a) or a multivariate analysis (MANOVA) (Potvin et al. 1990). Considerations of these procedures will also affect the optimal experimental design.

#### **Design recommendations**

The level of error and whether this error is constant or variable will dictate which methods can be used with satisfactory results, but the design of the experiment (i.e. the specific concentrations of  $S$  selected, the number of data points and the spacing of these points) will also play an important role (see Duggleby & Clark 1991). Based on the results of the present study, there are clearly data distributions that are not amenable to analysis even by non-linear methods. Data corresponding to Case 3, for example, where no  $V_{\text{max}}$  plateau is obtained, are particularly problematic. More accurate and precise values result when all data points are above  $K_m$ , suggesting that values on the plateau are critical. Currie (1982) found using simulations that non-linear methods performed best with data distributed geometrically (i.e. as in Case 1), but the results of the present study do not support this contention; fits to data distributed as Cases 2 and 4 were as good or better (see Table 2).

Johnson (1992b) has pointed out that increasing the number of independent data points gives better results than increasing the number of replicates at single  $S$  value, for non-linear fitting procedures (i.e. 30 different concentrations of  $S$  is superior to 3 replicates each at 10 different concentrations). Furthermore, data must not be smoothed since this will violate assumptions of random sampling. The 3 key assumptions remain however: (1) that the error in  $S$  is much smaller than the error in  $V$ ; (2) that  $V$  is normally distributed at a single value of  $S$ ; and (3) that the model to which the data are fit is correct (see Johnson 1992b). This study has not considered whether the rectangular hyperbola is an appropriate model for a given data set. Indeed, there is evidence that nutrient uptake kinetics and growth versus food concentration may not fit hyperbolic kinetics perfectly (see Mullin et al. 1975, Condrey 1982, Hurd & Dring 1990). Unfortunately, in real data sets, it may often be impossible to distinguish a lack of fit due to error from a lack of fit due to an unsuitable model.

In summary, this study has provided objective criteria for selecting a fitting method for a given data set, based on the data distribution and error level and structure. It has also provided a means to assess whether previously published parameter estimates using different fitting methods are satisfactory or whether they should be reanalyzed.

## LITERATURE CITED

- Berges, J. A. (1993). Enzymes as indices of growth rate and nitrate metabolism in marine phytoplankton. Ph.D. thesis, University of British Columbia, Vancouver
- Cleland, W. W. (1967). The statistical analysis of enzyme kinetic data. *Adv. Enzymol.* 29: 1–32
- Condrey, R. E. (1982). Ingestion-limited growth of aquatic animals: the case for Blackman kinetics. *Can. J. Fish. Aquat. Sci.* 39: 1585–1595
- Cornish-Bowden, A. (1979). *Fundamentals of enzyme kinetics*. Butterworths, London
- Cornish-Bowden, A. (1991). Parameter estimating procedure for the Michaelis-Menten model: reply to Tseng and Hsu. *J. theor. Biol.* 153: 437–440
- Cornish-Bowden, A., Eisenthal, R. (1974). Statistical considerations in the estimation of enzyme kinetic parameters by the direct linear plot and other methods. *Biochem. J.* 139: 721–730
- Cornish-Bowden, A., Porter, W. R., Trager, W. F. (1978). Evaluation of distribution-free confidence limits for enzyme kinetic parameters. *J. theor. Biol.* 74: 163–175
- Currie, D. J. (1982). Estimating Michaelis-Menten parameters: bias, variance and experimental design. *Biometrics* 38: 907–919
- Dowd, J. E., Riggs, D. S. (1965). A comparison of estimates of Michaelis-Menten kinetic constants from various linear transformations. *J. biol. Chem.* 240: 863–869
- Droop, M. R. (1983). Twenty-five years of algal growth kinetics. *Botanica mar.* 26: 99–112
- Duggleby, R. G. (1991). Analysis of biochemical data by non-linear regression: is it a waste of time? *Trends Biochem. Sci.* 16: 51–52
- Duggleby, R. G., Clarke, R. B. (1991). Experimental design for estimating the parameters of the Michaelis-Menten equation from progress curves of enzyme-catalyzed reactions. *Biochim. Biophys. Acta* 1080: 231–236
- Eisenthal, R., Cornish-Bowden, A. (1974). The direct linear plot: a new graphical procedure for estimating enzyme kinetic parameters. *Biochem. J.* 139: 715–720
- Hanes, L. (1932). Studies of plant amylases. I. The effect of starch concentration upon the velocity of hydrolysis by the amylase of germinated barley. *Biochem. J.* 26: 1406–1421
- Hofstee, B. H. N. (1959). Non-inverted versus inverted plots in enzyme kinetics. *Nature* 184: 1296–1298
- Holling, C. S. (1966). The functional response of invertebrate predators to prey density. *Mem. Entomol. Soc. Can.* 48: 1–86
- Hsu, J. P., Tseng, S. (1989). Estimating the adjustable parameters of the double Michaelis-Menten model. *J. theor. Biol.* 137: 355–363
- Hurd, C. L., Dring, M. J. (1990). Phosphate uptake by intertidal algae in relation to zonation and season. *Mar. Biol.* 107: 281–289
- Johnson, L. A. (1992a). Analysis of variance of parameter estimates: F tests and t tests. *Analyt. Biochem.* 206: 195–201
- Johnson, M. L. (1992b). Why, when and how biochemists should use least squares. *Analyt. Biochem.* 206: 215–225
- Jones, M. E., Taransky, K. (1991). Least-squares estimation of enzyme parameters. *Comp. Biol. Med.* 21: 459–464
- Lam, C. F. (1981). *Techniques for the modelling of enzyme kinetic mechanisms*. John Wiley and Sons, Chichester
- Leatherbarrow, R. J. (1990). Use of non-linear regression to analyze enzyme kinetic data: application to situations of substrate contamination and background subtraction. *Analyt. Biochem.* 184: 274–278
- Li, W. K. W. (1983). Consideration of errors in estimating kinetic parameters based on Michaelis-Menten formalism in microbial ecology. *Limnol. Oceanogr.* 28: 185–190
- Lineweaver, H., Burk, D. (1934). Determination of enzyme dissociation constants. *J. Am. Chem. Soc.* 56: 658–666
- MacIsaac, J. J., Dugdale, R. C. (1969). The kinetics of nitrate and ammonia uptake by natural populations of marine phytoplankton. *Deep Sea Res.* 16: 45–57
- Michaelis, M., Menten, M. L. (1913). Kinetics of invertase action. *Z. Biochem.* 49: 333–369
- Monod, J. (1942). *Recherches sur la croissance des cultures bactériennes*, 2nd edn. Herman et Cie, Paris
- Montagnes, D. J. S. (1993). An investigation of the growth and feeding responses of oligotrich ciliates to food types and concentrations: an approach assessing the potential of marine planktonic ciliate blooms. Ph.D. thesis, University of British Columbia, Vancouver
- Mullin, M. M., Stewart, E. F., Fuglister, F. J. (1975). Ingestion by planktonic grazers as a function of concentration of food. *Limnol. Oceanogr.* 20: 259–262
- Myers, R., Olson, S., Maloy, S. (1990). Computer programs for the rapid determination of enzyme kinetics on MS-DOS compatible microcomputers. *Comp. appl. Biosci.* 6: 51–53
- Nelder, J. A. (1991). Generalized linear models for enzyme-kinetic data. *Biometrics* 47: 1605–1615
- Packard, T. T. (1979). Half-saturation constants for nitrate reductase and nitrate translocation in marine phytoplankton. *Deep Sea Res.* 26A: 321–326
- Potvin, C., Lechowicz, M. J., Tardif, S. (1990). The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures. *Ecology* 71: 1389–1400
- Steel, R. G. D., Torrie, J. H. (1980). *Principles and procedures of statistics*, 2nd edn. McGraw-Hill, New York
- Tseng, S., Hsu, J. P. (1990). A comparison of the parameter estimating procedures for the Michaelis-Menten model. *J. theor. Biol.* 145: 457–464
- Verity, P. G. (1991). Measurement and simulation of prey uptake by marine planktonic ciliates fed plastidic and aplastidic nanoplankton. *Limnol. Oceanogr.* 36: 729–750
- Wilkinson, G. N. (1961). Statistical estimations in enzyme kinetics. *Biochem. J.* 80: 324–332
- Wilkinson, L. (1990). *SYSTAT: the system for statistics*. SYSTAT, Inc., Evanston, IL

*This article was submitted to the editor*

*Manuscript first received: April 5, 1994*

*Revised version accepted: August 4, 1994*