ISSN 0140 3818
SSSU 36

THE REGRESSION ANALYSIS OF SOME SHIP MODEL RESISTANCE DATA

by A.F. Molland and M.J. Watson

Ship Science Report No. 36

June 1988

THE REGRESSION ANALYSIS OF SOME SHIP MODEL RESISTANCE DATA

By

A.F. Molland and M.J. Watson

SUMMARY

The background to the investigation and development of new regression equations suitable for the prediction of total hull resistance is described.

The investigation indicated that, given an adequate number of data, equations derived at discrete increments of speed-length ratio and with a limited number of hull definition variables can provide very satisfactory preliminary resistance predictions. Standard errors are of the order of 3% to 6%.

1. INTRODUCTION

The application of multiple linear regression techniques to the analysis of ship model resistance data has taken place over several years and has, for example, been reported in Refs. 1 to 7. Refs. 4 and 6, in particular, appraise the applications of the techniques, their strengths and shortcomings.

This paper reports, and comments briefly, on an analysis of some of the model resistance data at British Maritime Technology Ltd (BMT). It was considered that the production of prediction formulae, resulting from a statistical analysis of these model data, would enhance the methods of predicting ship power at the early design stage. The results of the analysis indicate that this basic objective has been achieved.

A more detailed account of the work is reported in Ref. 8.

2. APPROACH TO THE ANALYSIS

A number of statistical software packages were considered, and that chosen as being the most suitable for this investigation was the B.M.D.P. package (Ref. 9) which offers stepwise multiple linear regression and all the necessary pre and post analysis statistical tests.

In view of the large amount of data to be handled and the likely uses of the predictor equations it was desirable that the analysis should be based on the information available from the original standard B.M.T. resistance data sheets. This amounted to © $_{400}$ (based on total resistance with Froude extrapolation) for a range of speeds for given hull particulars.

Only single screw vessels in the loaded condition were considered in the current analysis. Bulbous bow models were not included. (It was decided that the investigation of bulbous bows and ballast conditions could be treated with complementary analyses at some later stage). Specialist vessels such as tugs, trawlers and patrol boats were also excluded from the analysis.

All the resistance data had been recorded at $v/\sqrt{L}$ increments of 0.02. Thus large numbers of data were available at most of these speed increments, without the need for any cross fairing. This property of the data base made it highly suitable for regression at discrete speeds rather than attempting to formulate or include terms in the regression equation to take account of speed dependent wave interference effects. $v/\sqrt{L}$ values having large numbers of data cases associated with them had the cases divided into two groups according to the magnitude of C_B.

Even after grouping for C_B a large amount of data existed at most speeds; thus a representative and manageable sample (rather than all the data) was used to derive each regression equation. Examples of the total number of cases available over the speed range (after sorting/editing etc) were as follows:

$v/\sqrt{L}$	0.40	0.50	0.60	0.70	0.80	0.90	0.94	0.96
No.	119	526	869	863	520	219	110	86

3. DEVELOPMENT OF THE EQUATIONS

The variables in the data base defining the hull form were L, B, T, C_B, C_M, LCB, HA and ⓢ, and these were grouped into a functional equation of the form:

$$\textcircled{c} = f[L/B,\ B/T,\ C_B,\ C_M,\ LCB,\ HA,\ \textcircled{s}]$$

For the regression analysis a curvilinear model was used in which the independent variables entered into the analysis were made up of the seven first order variables, their squares, cross products, cubes and third order cross-terms.

The development of each equation (at each $v/\sqrt{L}$ increment) entailed the investigation of correlations and multicollinearities between the independent variables, relative significance of the variable groups, residual analysis and estimates of its predictive qualities in terms of R^2 and Standard Error.

The use of the stepwise regression package resulted in a number of equations at each speed, all showing similar predictive qualities. Of the possible choices it was desirable to choose an equation that was simple in structure and consistent with equations for neighbouring $v/\sqrt{L}$ increments.

4. THE EQUATIONS

4.1 Variables

In order to achieve a reasonably consistent set of variable groupings across the speed range, the `best` equation (in the statistical sense) for a particular speed was not necessarily chosen. The resulting equations feature only eight different variable groupings, with between three and five of these appearing in any one equation. The eight groups are as follows: (C_B); (C_B x L/T); (C_B^2 x L/B); (L/B); (B/T); (LCB); (LCB^2 x HA); (HA). (L/T in the second group results from the product of L/B and B/T).

It is noted that C_M and Ⓢ do not feature in the equations. Their significance was generally small and at several speeds their effect was insignificant. For consistency throughout the equations, therefore, these variables were omitted.

The variable groupings finally adopted in each of the equations are given in Table 1.

In general, the move to an equation with a more consistent grouping led in most cases to only small increases in standard error. It is seen from Table 1 that the standard errors for the equations for the lower C_B ranges are all less than 4% whilst those for the higher C_B ranges tend to lie between 4% and 6%. The higher C_B range equations broadly represent the `overdriven` case (when the resistance curve tends to rise more steeply) and a larger scatter in the basic data might well exist for these cases.

An investigation of confidence limits indicated that provided proposed designs have parameters within the limits defined by data availability then the 95% confidence interval is likely to be within ±2.2 times the standard error (given in Table 1).

4.2 Relative Importance of the Terms in the Regression Equations

The approximate proportion of each term in the equation is given in Table 2. For each speed, the value of each term (based on the mean value of the variables) together with the maximum range (based on the limits of the variables) is expressed as a percentage of the mean resistance coefficient.

As might be expected from the physics of the problem, the (C_B x L/T), (${C_B}^2$ x L/B) and (C_B) terms predominate. (L/B) is relatively important, particularly in the equations for the larger C_B range. The contributions of (LCB) and (LCB^2 x HA) are generally not large, but change significantly with speed. The contribution of HA is relatively small.

5. EXAMPLE TESTS OF EQUATIONS

In order to test the equations, a program was written which outputs ⓒ$_{400}$ at each $v/\sqrt{L}$ increment, and is also linked to a plotter to provide visual inspection of the results. The lines in the plots (Fig. 2) are not smoothed and are produced merely by drawing from point to point.

5.1 Comparison with Models from the Data Base

Comparisons of predictions by the regression equations with representative experimental results from the data base are given in Fig. 1. The squares represent the predicted values and the line represents the observed values for each model. A filled square on the plot indicates that the model was included in the regression data at that $v/\sqrt{L}$ increment. It is clear from these plots that the discrepancy between observed and predicted results is independent of whether or not the observed model was included in the regression data.

5.2 Tests of Changes in Basic Variables

As examples, the results of systematic changes in B/T and LCB using the equations are given in Fig. 2. The use of such a technique provides the user with an immediate broad appraisal of the likely sensitivity of ⓒ to changes in a particular variable over the speed range. Such a technique is, of course, strictly constrained within the limits of the data used in developing the equations.

5.3 Comparison with Other Regression Formulae

A limited comparison was made with the published regression formulae of Holtrop, Ref. 7. The wave resistance formula in this reference includes speed as a variable and the regression coefficients appear to have been derived using several ship types. Fig. 3 shows the results of the comparison. Holtrop comments that his wave resistance formula is only partially successful at low and moderate speeds, and this is apparent in Fig. 3. However, although it is not made clear in Ref. 7 or earlier references, it appears that the formulae are intended to be used primarily for the design speed; this would also account for some of the discrepancies seen in Fig. 3.

Fig. 3 also shows predictions using the equations proposed by Sabit, Ref. 2. These are seen to be quite good approximations to the observed curves, but the speed range of Sabit`s equations is very limited.

Inspection of Figs. 1 and 3 indicates the justification in the present approach of considering a limited ship type (or limited range of ship types) and, if adequate quantity of data allows, the derivation of equations at discrete values of $v/\sqrt{L}$, thus eliminating speed as a variable in the equation.

6. CONCLUDING COMMENTS

a) Most of the regression equations produced have a standard error of between about 3% and 4%. In some cases, involving higher block coefficients and/or higher speeds, the standard error lies between 4% and 6%. It was found that the 95% confidence interval for the equations is likely to be between ± 2.2 times the standard error.

It should be noted that, given the limited number of parameters available in the data base for defining the hull form, the feasible limits of accuracy may have been reached. For example, changes in fore and aft end section shape will not be reflected in the standard hull parameters used in the regression analysis.

b) Tests of the equations showed a fair degree of stability both over the speed range and for changes in variables. As might be expected, there is some instability and lack of accuracy near the limits of the data.

Relatively small local undulations in the predictions over a speed range do occur. However, considering the standard errors of the individual equations and the fact that there is a separate equation for each speed, this would be expected.

c) The proposed regression equations indicate that using discrete values of $v/\sqrt{L}$, suitable groupings of C_B and independent variables made up of first order terms, squares and cross products of the starting variables, equations can be derived which can predict total resistance with a good order of accuracy. The application of the large number of regression equations resulting from this approach is simple and practical if the equations are included in a computer based resistance prediction analysis.

ACKNOWLEDGEMENTS

The work described in this paper was carried out at the University of Southampton and was funded jointly by the S.E.R.C. and B.M.T. Ltd. The data for the analysis were kindly supplied by B.M.T.Ltd and the paper is published by permission of the Company.

NOMENCLATURE

B : Breadth

C_B : Block coefficient

C_M : Midship area coefficient

HA : Half angle of entrance

L : Length between perpendiculars

LCB : Longitudinal centre of buoyancy

T : Draught

ⓢ : Wetted surface area coefficient

ⓒ$_{400}$: Resistance coefficient (for 400 ft. ship)

$v/\sqrt{L}$: Speed-length ratio (speed knots, length ft.)

SE : Standard Error

R^2 : 1 - ESS/TSS
(ESS = error sum of squares; TSS = Total sum of squares)

REFERENCES

1. Doust, D.J. : `Optimised Trawler Forms`. Trans N.E.C.I. Vol. 79, 1962-63.

2. Sabit, A.S. : `Regression Analysis of the Resistance Results of the B.S.R.A. Series`. International Shipbuilding Progress, Volume 18, 1971.

3. Swift, P.M., Nowacki, H. and Fischer, J.P. : `Estimation of Great Lakes Bulk Carrier Resistance Based on Model Test Data Regression`. Marine Technology, Vol. 10, 1973.

4. Fairlie-Clarke, A.C. : `Regression Analysis of Ship Data`. International Shipbuilding Progress, Vol. 22, 1975.

5. Holtrop, J. and Mennen, G.G.J. : `An Approximate Power Prediction Method`. International Shipbuilding Progress, Vol. 29, 1982.

6. Coles, P.D. : `Prediction of Hull Resistance for Merchant Ship Forms`. M.Sc. Dissertation, Faculty of Mathematical Studies, University of Southampton, 1983.

7. Holtrop, J. : `A Statistical Re-analysis of Resistance and Propulsion Data`. International Shipbuilding Progress, Vol. 31, 1984.

8. Molland, A.F. and Watson, M.J.: 'The Regression Analysis of Ship Model Resistance Data'. SERC/BMT Cooperative Research Project GR/C/9746.0, Report to SERC, November 1986.

9. Biomedical Computer programs, P-Series (BMDP-83 Manual). University of California Press.

$v/\sqrt{L}$	C_B Range	$C_B \times L/T$	LCB	B/T	L/B	$LCB^2 \times HA$	HA	$C_B^2 \times L/B$	C_B	SE%
0.40	All C_B	●	●	●		●				3.2
0.42	All C_B	●	●	●		●				3.4
0.44	All C_B	●	●	●		●				3.1
0.46	All C_B	●	●		●	●				2.9
0.48	All C_B	●	●		●	●				2.7
0.50	All C_B	●	●		●	●				2.8
0.52	All C_B	●	●		●					2.6
0.54	All C_B	●	●		●					2.6
0.56	All C_B	●	●		●					2.6
0.58	All C_B	●	●		●					3.2
0.60	<0.81	●	●		●					2.7
0.60	>0.77	●	●		●	●				4.8
0.62	<0.80	●	●		●				●	2.8
0.62	>0.76	●	●		●	●		●		4.5
0.64	<0.79	●			●	●	●		●	2.6
0.64	>0.75	●	●		●	●		●		4.7
0.66	<0.78	●			●	●	●		●	2.7
0.66	>0.74	●	●		●	●		●		5.3
0.68	<0.77	●			●	●	●		●	3.0
0.68	>0.73	●	●		●	●		●		5.1
0.70	<0.75	●			●	●	●		●	3.6
0.70	>0.71	●	●		●	●		●		4.1
0.72	<0.74	●			●	●	●		●	3.8
0.72	>0.70	●			●	●		●		4.1
0.74	<0.72	●			●		●		●	3.3
0.74	>0.68	●	●		●	●		●		4.4
0.76	<0.71	●	●		●		●		●	3.3
0.76	>0.67	●			●	●		●		4.8
0.78	<0.69	●	●		●		●		●	3.2
0.78	>0.65	●			●	●		●		5.5
0.80	<0.68	●	●		●		●		●	3.6
0.80	>0.64	●	●		●	●		●		6.7
0.82	<0.66	●	●		●		●		●	3.6
0.82	>0.62	●	●		●	●		●		5.5
0.84	<0.65	●	●		●		●		●	3.8
0.84	>0.61	●	●		●	●		●		6.7
0.86	<0.65	●	●		●	●			●	3.1
0.86	>0.61	●	●		●	●		●		4.4
0.88	All C_B	●	●			●		●		4.1
0.90	All C_B	●	●			●		●		4.8
0.92	All C_B	●	●			●		●		5.4
0.94	All C_B	●	●			●		●		6.0

● indicates groups used

TABLE 1 GROUPINGS OF VARIABLES AT EACH SPEED

v/√L	C_B Range	Const	C_BxL/T	LCB	B/T	L/B	LCB^{2}xHA	HA	$C_B{}^2$xL/B	C_B
0.40	All C_B	78	58±22	-10±7	-30±5	-	4±4	-	-	-
0.42	All C_B	69	58±22	-9±7	-22±4	-	4±4	-	-	-
0.44	All C_B	64	57±21	-8±8	-18±3	-	5±5	-	-	-
0.46	All C_B	37	44±17	-5±5	-	20±3	4±4	-	-	-
0.48	All C_B	42	45±18	-4±5	-	14±2	3±3	-	-	-
0.50	All C_B	36	48±21	-4±5	-	17±3	3±3	-	-	-
0.52	All C_B	48	44±18	-2±3	-	9±1	-	-	-	-
0.54	All C_B	47	44±18	-1±2	-	10±1	-	-	-	-
0.56	All C_B	40	45±19	-2±3	-	17±2	-	-	-	-
0.58	All C_B	40	49±21	-1±3	-	12±2	-	-	-	-
0.60	<0.81	43	47±19	-1±4	-	11±2	-	-	-	-
0.60	>0.77	21	55±19	-21±11	-	26±3	20±18	-	-	-
0.62	<0.80	66	54±23	-1±2	-	6±1	-	-	-	-25±3
0.62	>0.76	15	42±15	-16±10	-	-20±3	18±16	-	62±14	-
0.64	<0.79	79	56±24	-	-	13±2	-2±2	4±5	-	-49±6
0.64	>0.75	36	30±11	-9±6	-	-50±7	13±13	-	80±19	-
0.66	<0.78	79	54±22	-	-	12±2	2±2	5±3	-	-52±6
0.66	>0.74	29	34±12	-11±9	-	-40±6	19±19	-	70±17	-
0.68	<0.77	82	55±23	-	-	15±2	3±2	9±6	-	-63±7
0.68	>0.73	26	37±14	-10±9	-	-45±6	19±19	-	73±18	-
0.70	<0.75	85	56±23	-	-	12±2	2±2	11±7	-	-66±7
0.70	>0.71	28	45±17	-3±3	-	-31±4	13±13	-	47±13	-
0.72	<0.74	67	52±21	-	-	11±2	2±1	7±4	-	-39±4
0.72	>0.70	34	39±15	-	-	-35±5	12±12	-	50±14	-
0.74	<0.72	80	53±21	-	-	10±1	-	13±7	-	-56±5
0.74	>0.68	25	43±17	-2±3	-	-38±5	16±15	-	55±17	-
0.76	<0.71	80	46±18	0±2	-	18±3	-	12±7	-	-56±5
0.76	>0.67	29	43±16	-	-	-63±9	10±9	-	80±23	-
0.78	<0.69	92	50±19	0±2	-	14±2	-	13±7	-	-69±5
0.78	>0.65	29	47±18	-	-	-62±9	10±9	-	76±23	-
0.80	<0.68	94	48±18	-1±3	-	19±3	-	14±7	-	-73±5
0.80	>0.64	28	44±17	+1±5	-	-47±7	10±8	-	63±19	-
0.82	<0.66	109	45±16	-2±5	-	19±3	-	8±3	-	-78±4
0.82	>0.62	29	50±20	+1±8	-	-37±5	9±7	-	48±15	-
0.84	<0.65	96	46±16	-3±6	-	24±3	-	9±3	-	-72±3
0.84	>0.61	30	52±21	0±7	-	-35±5	9±6	-	43±14	-
0.86	<0.65	88	46±16	-6±10	-	13±2	3±2	-	-	-44±2
0.86	>0.61	43	44±18	0±10	-	-20±3	4±2	-	28±6	-
0.88	All C_B	36	44±18	-1±11	-	-	7±4	-	14±5	-
0.90	All C_B	28	42±16	-1±10	-	-	5±2	-	25±8	-
0.92	All C_B	18	48±18	-3±9	-	-	3±1	-	34±9	-
0.94	All C_B	20	37±13	-5±9	-	-	2±2	-	46±10	-

TABLE 2 PROPORTIONS OF THE TERMS IN EACH REGRESSION EQUATION
(as a percentage of mean resistance coefficient)

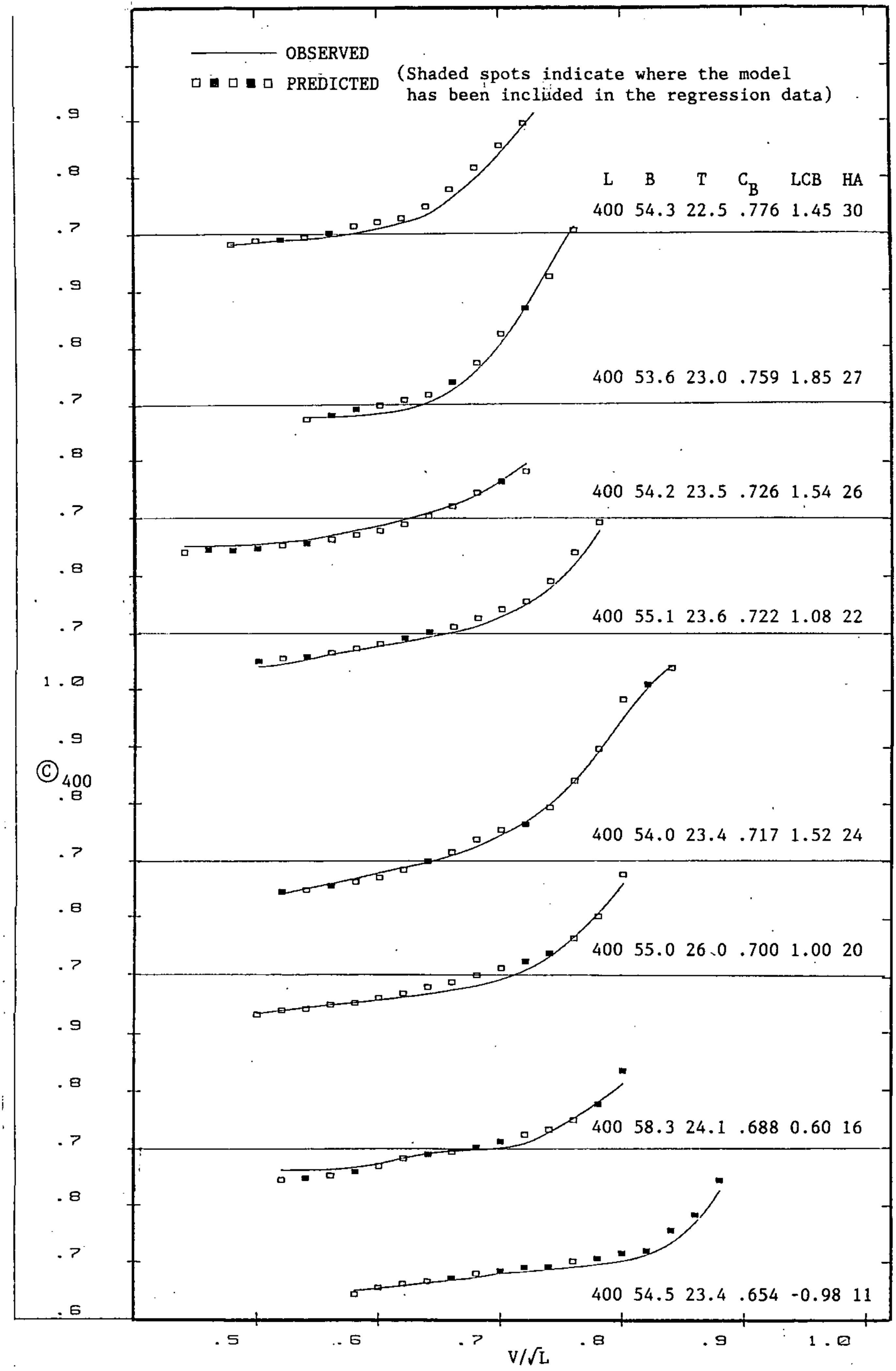

Fig. 1 OBSERVED AND PREDICTED RESULTS

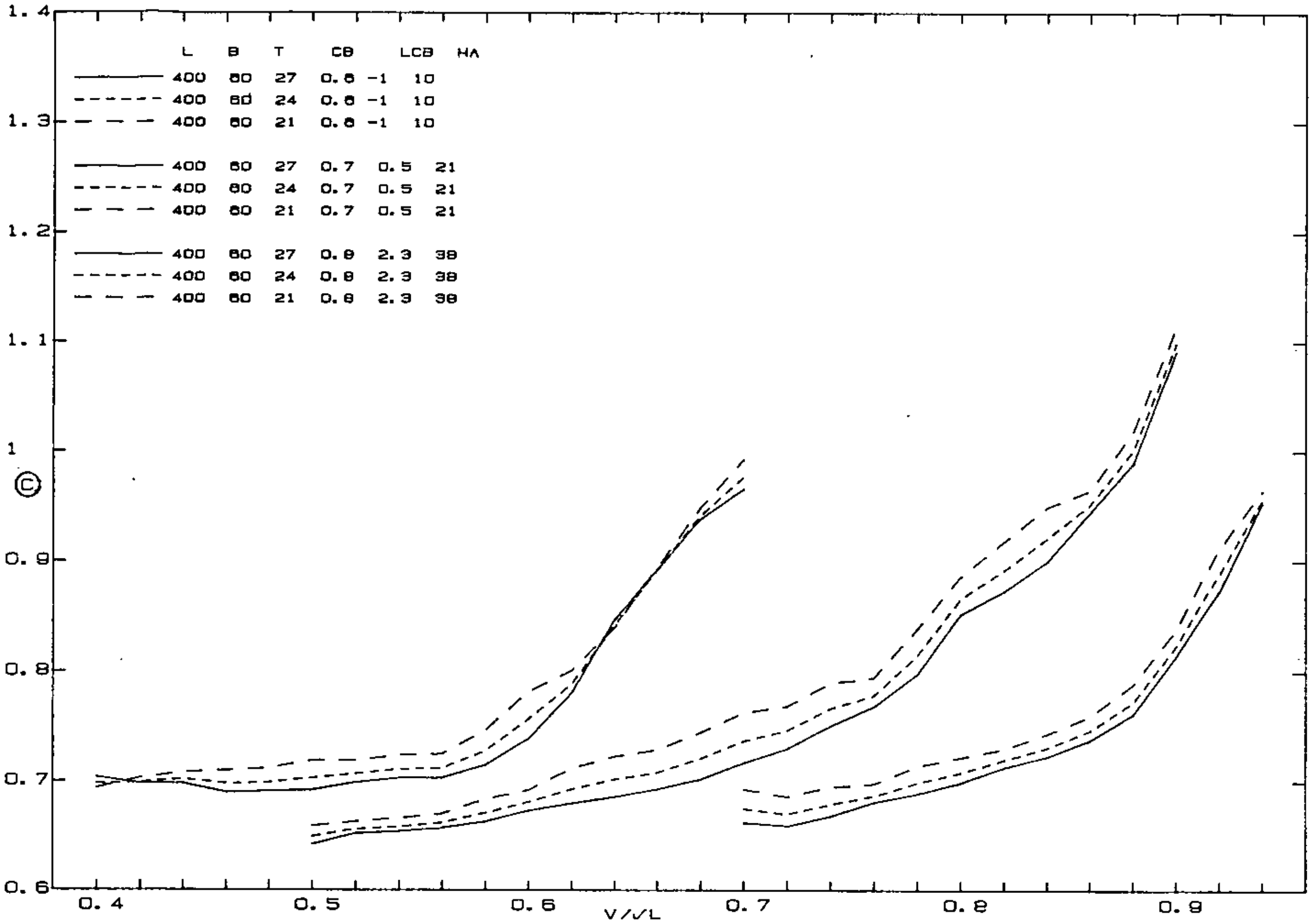

Fig. 2(a) CHANGES IN B/T (L/B=6.7)

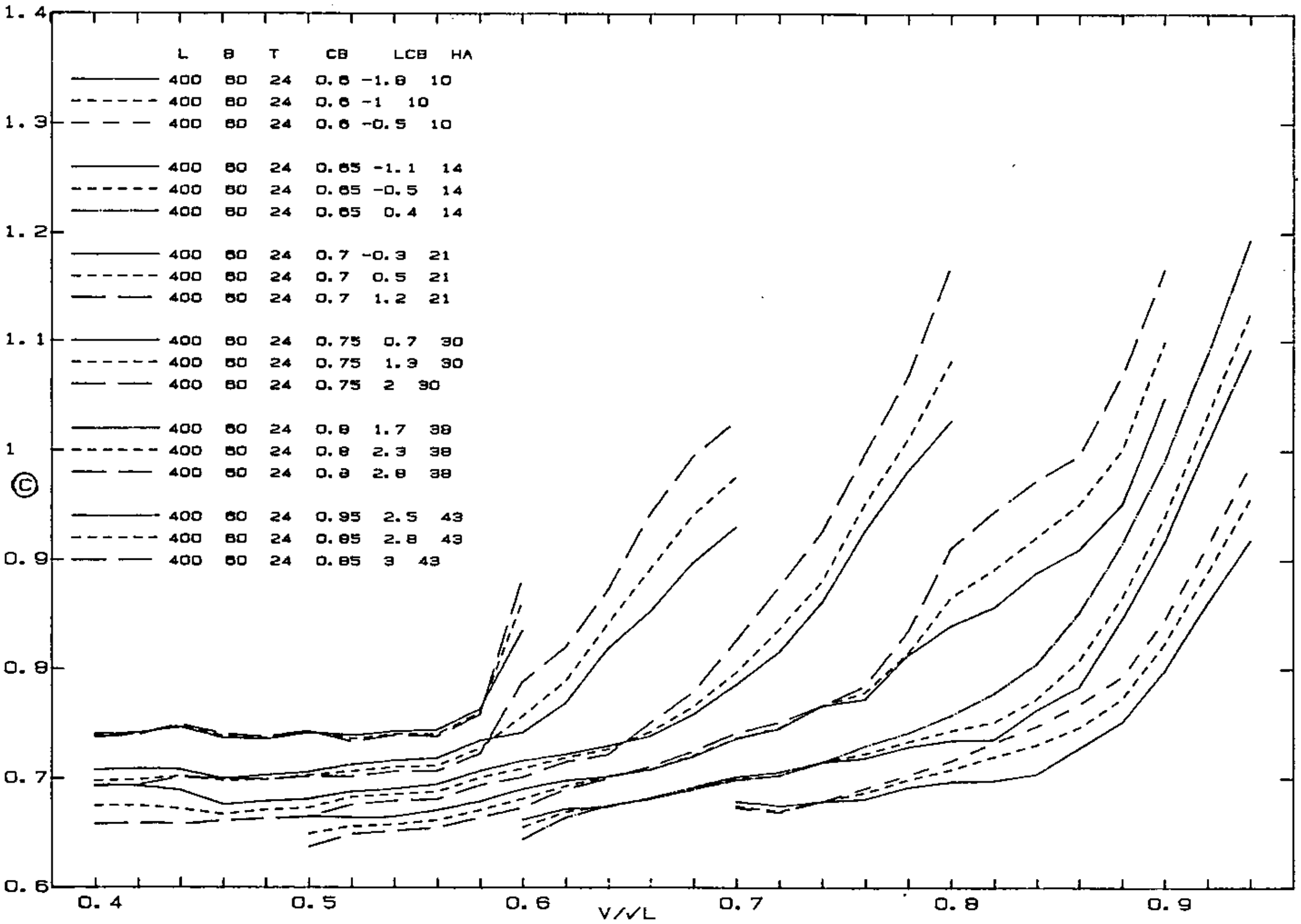

Fig. 2(b) CHANGES IN LCB

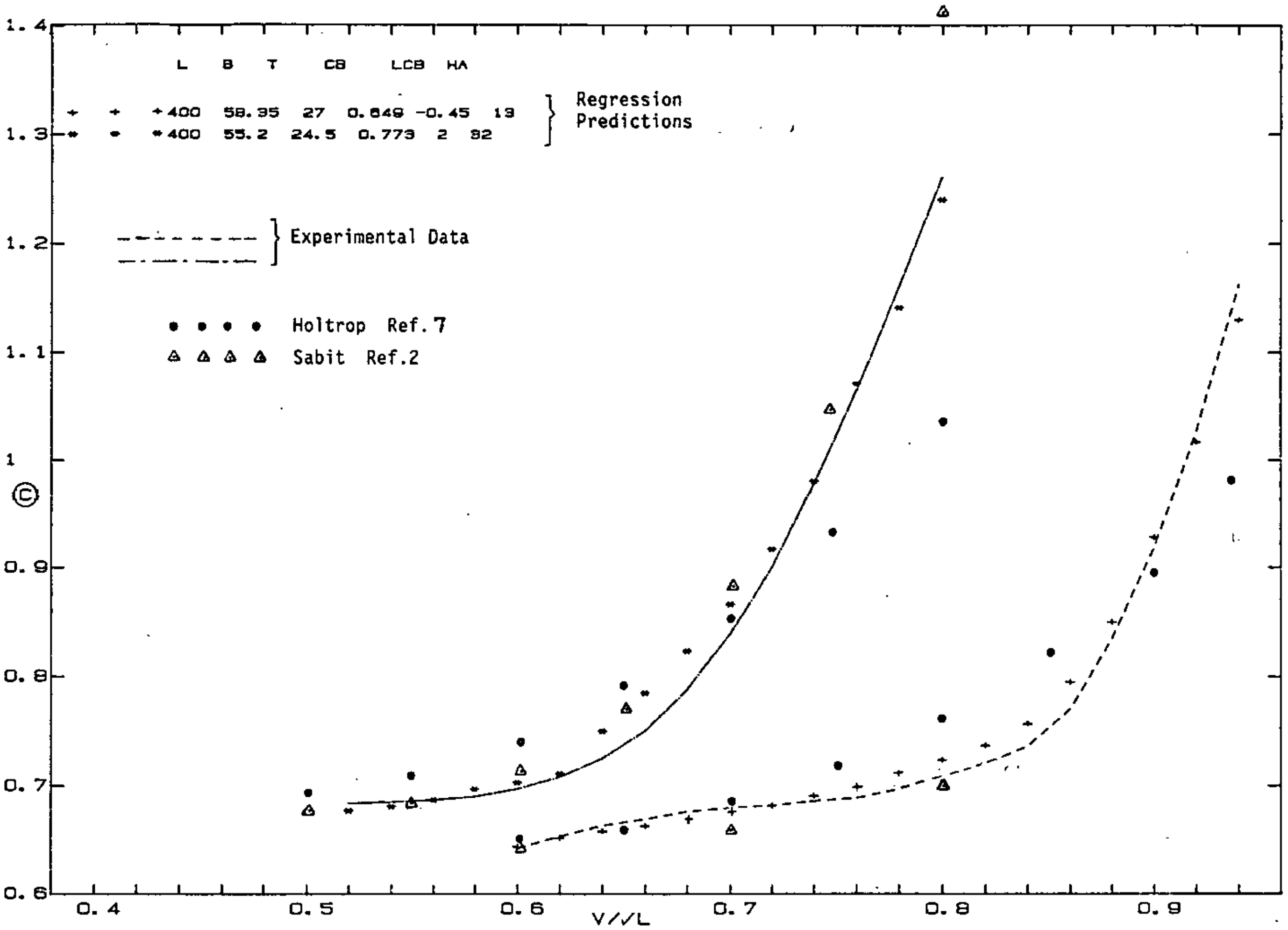

Fig. 3(a) COMPARISONS WITH OTHER REGRESSION FORMULAE

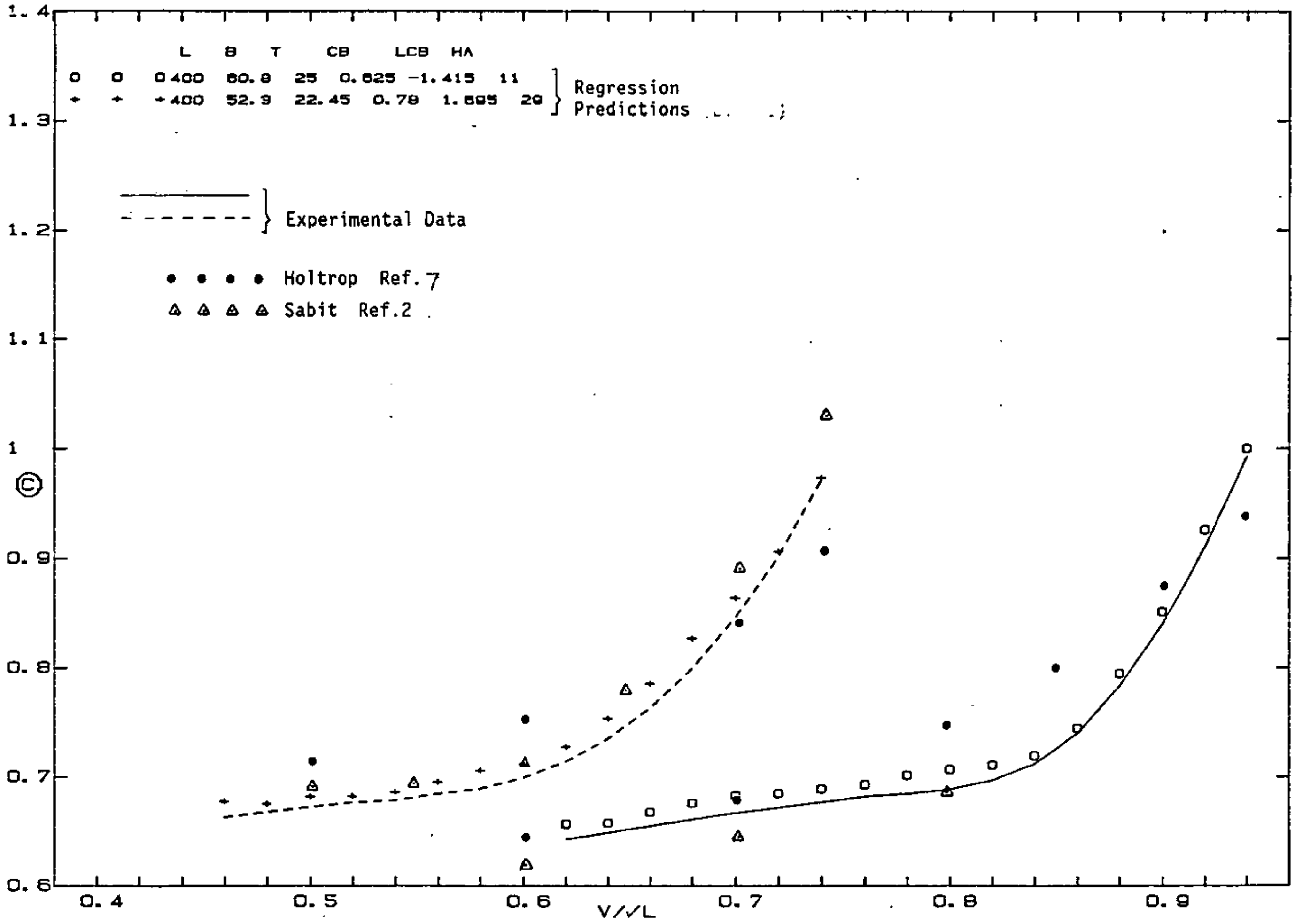

Fig. 3(b) COMPARISONS WITH OTHER REGRESSION FORMULAE