A linear regression model for marine propeller optimization, prototyping and design

Liu, P.; Wang, H.; Quinton, J.; Veitch, B.

Publisher’s version / Version de l'éditeur:

11th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery [Proceedings], 2006

Proceedings of ISROMAC 2006:
The Eleventh International Symposium on Transport Phenomena and Dynamics of Rotating Machinery
February 26 - March 2, 2006, Honolulu, Hawaii USA

ISROMAC2006-69

A LINEAR REGRESSION MODEL FOR MARINE PROPELLER OPTIMIZATION, PROTOTYPING AND DESIGN

Pengfei Liu *
National Research Council Canada
Institute for Ocean Technology
Box 12093, 1 Kerwin Place
St. John's, NL Canada A1B 3T5
Email: Pengfei.Liu@nrc-cnrc.gc.ca
Phone: 1-(709) 772-4575
Fax: 1-(709) 772-2462

Hong Wang
Department of Mathematics and Statistics
Faculty of Science,
Memorial University of Newfoundland
St. John's, Newfoundland
A1B 3X5

Justin Quinton
Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, Newfoundland
A1B 3X5

Brian Veitch
Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, Newfoundland
A1B 3X5

ABSTRACT
A multiple-variable linear regression direct solution model and a statistical model were developed for marine propeller design, optimization and prototype. Computing implementation for the direct solution model was made to create an integrated tool for the marine propeller development process. An error analysis for a simple case with only 4 independent variables was performed. This direct solution model was constructed to provide two functionalities: generation of a set of linear regression coefficients to establish a multiple-variable polynomial equation and interpolation of the multiple-variable data set that are generated by the polynomial equations. An application case was given using a set of data from a marine nozzle propeller series both to cover interpolation to produce curves and linear regression coefficients for interpolation, for both the direct solution model and the statistical model that was computed under a commercial software package. Though much higher than the statistical model, interpolation by the direct solution model showed an error of less than one-tenth of a percent for a group of nozzle propellers. The highly computing-efficient direct solution method showed its capability as a general-purpose linear regression tool which can be applied widely for optimal product prototyping and design.

Keywords: Linear Regression, Optimal Prototyping and Design, Marine Propeller

INTRODUCTION
In the marine propeller design and optimization process, a set of performance curves are required for each candidate model propeller. These performance curves are namely the thrust coefficient K_t, torque coefficient K_q and propulsive efficiency η, versus the advance coefficient J. For a special purpose propeller series, typical performance curves are usually given for a propeller series via cavitation tunnel or tow tank tests, in terms of geometric parameters. These parameters are mainly the propeller disk expanded area ratio, *EAR*, with the same blade sectional shape, i.e., the same sectional profile, camber and maximum thickness distribution, the nominal pitch diameter ratio, $p/D_{0.7R}$ with the same pitch distribution along the span of the blade, the number of blades, Z, each with the same blade planform contour. In propeller series model tests, a number of propeller models are manufactured to cover the possible range of geometric parameters for ship operation. A typical performance diagram is shown in Figure 1 as an example. The diagram is for a propeller with 4 blades (Z=4) and an expanded area ratio of 0.55 (*EAR*=0.55), It shows thrust curves versus the advance coefficient J for a range of pitch values from $p/D_{0.7R}$ =0.6 to 1.4, with an interval of $\Delta(p/D)$=0.2..

For a comprehensive propeller series test, a number of propeller models must be manufactured and tested. If the tests are to cover a range of EAR from, say, 0.4 to 1.2 with an interval of 0.2, and the number of blades is from Z= 3 to 6, there will be 5 EAR values and 4 blade number values. Along with 5 pitch values from $p/D_{0.7R}$ = 0.6 to 1.4, a total of 100 propeller models must be manufactured. Increasing the intervals will reduce the number of propeller models required.

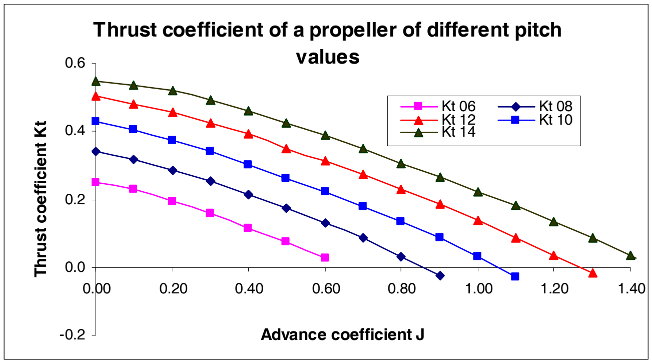

Figure 1. THRUST COEFFICIENT K_t OF A PROPELLER VERSUS ADVANCE COEFFICIENT *J*.

After the tests, a total of 20 propeller performance diagrams in the same form of Figure 1 can be plotted. The data obtained will be used to generate polynomial coefficients via linear regression and then establish a polynomial equation for propeller design and optimization with 3 independent variables, such as,

$$K_t = \sum_{i=0}^{I} \sum_{k=0}^{K} \sum_{l=0}^{L} (EAR)^i (p/D_{0.7R})^k J^l C_{i,k,l} \quad (1)$$

This work is to develop a numerical model to generate a set of cofficients for one or more independent variables in the form of $C_{i,k,l}$ in equation (1) and use the coefficients and values of independent variables to predict the hydrodynmic performance of an arbitrary canditate propeller within the series.

FORMULATION OF THE METHOD

To find a set of polynominal coefficients, for example with 4 independent variables, let the dependant variable:

$$y = f(x_1, x_2, x_3, x_4), \quad (2)$$

be represented by:

$$y = \sum_{i=0}^{i_m} \sum_{j=0}^{j_m} \sum_{k=0}^{k_m} \sum_{l=0}^{l_m} x_1^i x_2^j x_3^k x_4^l A_{i,j,k,l}. \quad (3)$$

For the simplicity of description, we set the values of the exponents i_m=1, j_m=1, k_m=1 and l_m=2. In practice, these values are often taken as 6-8, considering both accuracy and conservation of computing resources. An expanded form of equation (3) becomes

$$\begin{aligned} y = {} & A_{0,0,0,0}\, x_1^0 x_2^0 x_3^0 x_4^0 + A_{0,0,0,1}\, x_1^0 x_2^0 x_3^0 x_4^1 \\ & + \cdots + A_{1,1,1,2}\, x_1^1 x_2^1 x_3^1 x_4^2. \end{aligned} \quad (4)$$

Equation (4) has (i_m-0+1) (j_m-0+1) (k_m-0+1) (k_m-0+1) = (1-0+1)×(1-0+1)×(1-0+1)×(2-0+1) = 24 terms. It may be written in a matrix form, which can be solved for the polynomial coefficients directly:

$$[X][A]=[Y], \quad (5)$$

where [*Y*] is a vector storing 24 known values, which can be K_t, K_q or η values of the propeller performance curves, corresponding to the values of the independent variables raised to their respective powers and these values are stored in the square matrix [*A*].

Expanding equation 5, it gives:

$$\begin{bmatrix} 1 & x_{1,1}^0 x_{2,1}^0 x_{3,1}^0 x_{4,1}^0 \cdots x_{1,1}^1 x_{2,1}^1 x_{3,1}^1 x_{4,1}^2 \\ 1 & x_{1,2}^0 x_{2,2}^0 x_{3,2}^0 x_{4,2}^0 \cdots x_{1,2}^1 x_{2,2}^1 x_{3,2}^1 x_{4,2}^2 \\ \vdots & \vdots \quad \vdots \quad \vdots \\ 1 & x_{1,23}^0 x_{2,23}^0 x_{3,23}^0 x_{4,23}^0 \cdots x_{1,23}^1 x_{2,23}^1 x_{3,23}^1 x_{4,23}^2 \\ 1 & x_{1,24}^0 x_{2,24}^0 x_{3,24}^0 x_{4,24}^0 \cdots x_{1,24}^1 x_{2,24}^1 x_{3,24}^1 x_{4,24}^2 \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} A_{0,0,0,0} \\ A_{0,0,0,1} \\ \vdots \\ A_{1,1,1,1} \\ A_{1,1,1,2} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{23} \\ y_{24} \end{bmatrix}$$

Once [*A*] is obtained, the polynomial equation is defined.

RESULTS AND DISCUSSION

Application case for the newly developed model

To test the method and its implementation, a set of propeller propulsive performance data was used [Yossifov et al. 1989]. Figure 2 shows a nozzle propeller in three different surface modeling approaches: hidden line, solid modeling and wire frame (Liu 2002 and Liu et al. 2002).

While in the figure only one combination of the propeller is shown, this propeller series has three different nozzles, N=1, 2 and 3, four nominal pitch values of $(p/D)_{0.7R}$ =1.0, 1.1, 1.2 and 1.3, and three *EAR* values of 0.5, 0.6 and 0.7. The dependent variable is either K_t, K_q or η and the four independent variables are: *N*, $(p/D)_{0.7R}$, *EAR* and advance coefficient *J*.

Using the currently developed linear regression model, a set of polynomial coefficients can be obtained. These coefficients defined the polynomial equation which is,

$$K_t = f(J, N, p/D_{0.7R}, EAR), \quad (7)$$

in a form of

$$K_t = \sum_{i=0}^{i_m} \sum_{j=0}^{j_m} \sum_{k=0}^{k_m} \sum_{l=0}^{l_m} N^i (EAR)^j (p/D_{0.7R})^k J^l C_{i,j,k,l}.$$

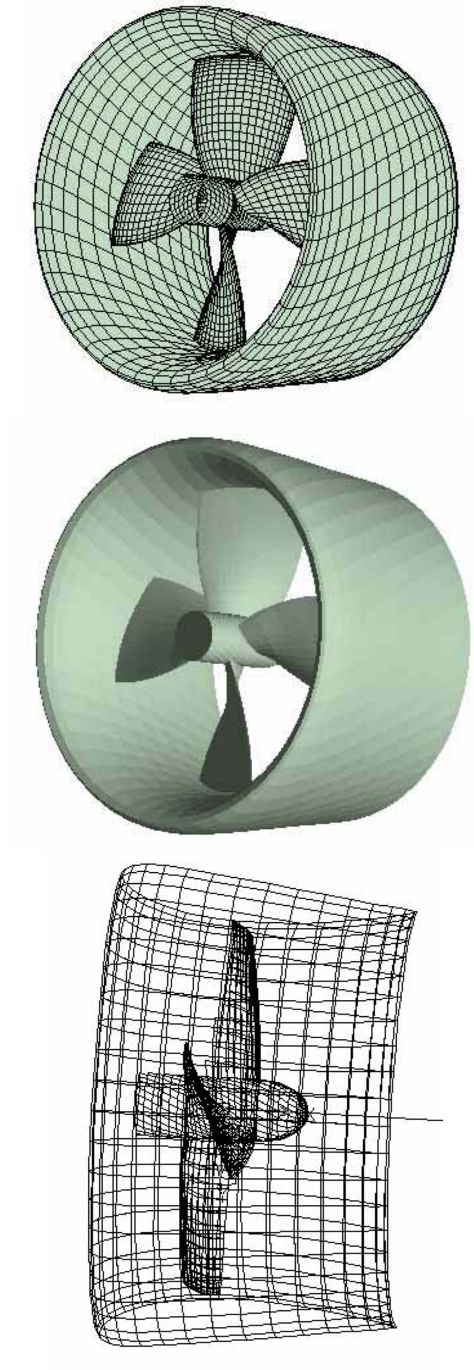

Figure 2. APPLICATION PROPELLER SERIES GEOMETRY.

The values of i_m, j_m, k_m and l_m are usually chosen at a small value less than 10. For an 8-variable problem, if the value of the power is set at 8, the matrix size will be $(9\times9\times9\times9\times9\times9\times9\times9)^2$ $=43046721^2$ which needs about 1,380,000 GB of computing memory. For too large a matrix size, an iterative solver should be used and the solver might need to be run under a parallel computing environment [Liu and Li 2002].

For this application case, for simplicity, we set $i_m = 2$, $j_m = 2$, $k_m = 3$ and $l_m= 6$. The matrix size is then 252^2. A total of 252 data points were prepared as in Table 1:

Table 1. INPUT DATA FORMAT AND LIST

	N	EAR	PD	J	Kt
1	1	0.5	1.0	0.00	0.518087
2	1	0.5	1.0	0.14	0.433046
3	1	0.5	1.0	0.28	0.353132
.	.	.	.	.	. . .
252	3	0.7	1.3	1.14	-2.464862

The current model was used to generate a set of linear regression coefficients by solving for $C_{i,j,k,l}$ to produce a polynomial

$$K_t = \sum_{i=0}^{2} \sum_{j=0}^{2} \sum_{k=0}^{3} \sum_{l=0}^{6} N^i (EAR)^j (p/D_{0.7R})^k J^l C_{i,j,k,l}. \quad (8)$$

Plugging the values of these independent variables into equation (8) for the 252 data points, the K_t values are obtained by the polynomial equation. Error estimation can then be done by comparing the K_t values from the polynomial equation with those from Table 1 in terms of

$$\%E \quad = \quad \left| \frac{K_{t1}(J) - K_{t2}(J)}{K_{t1}(J=0)} \right| \times 100\%. \qquad (9)$$

The percent error for the 252 data points is plotted in Figure 3.

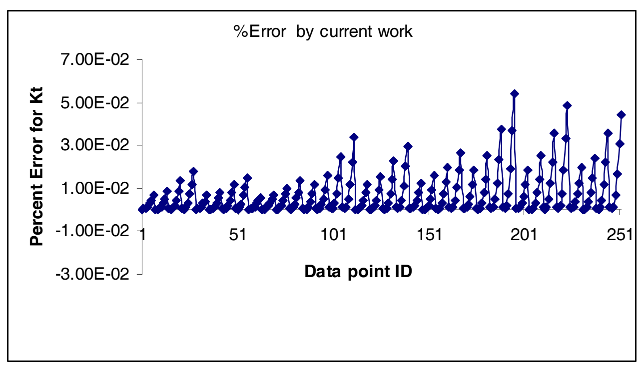

Figure 3. ERROR IN PERCENT BY THE CURRENT NUMERICAL MODEL, DGSI.

It is shown in Figure 3 that the maximum percentage error is about 0.05%. This accuracy is sufficient for engineering design. A much higher accuracy may possibly be obtained when the value of the exponents is increased, though increased computing resources are also required. To generate these coefficients for a one-time execution, the implemented code

DGSI, took about a couple of minutes of CPU time on a 3.0 GHz PC.

Application Case for the Statistics Commercial Package SPlus and Comparison

Statistics model

The statistical method applied here is the traditional regression analysis based on multiple predictors that are identified as *N, EAR,* $(P/D)_{0.7R}$ and *J* with a response variable K_t . First, we establish a probabilistic model denoted by

$$K_t = \sum_{i=0}^{2} \sum_{j=0}^{2} \sum_{k=0}^{3} \sum_{l=0}^{6} N^i (EAR)^j (p/D_{0.7R})^k J^l C_{i,j,k,l} + \varepsilon. \quad (10)$$

where ε is the random factor having an approximately normal distribution with zero mean and constant variance. In this topic, this assumption is not critical because the data collected here only contain record errors.

Base on the observed data, we first form a 252×252 matrix for the factors N^i, $(EAR)^j$, $(P/D)^k$ and J^l when *i, j, k* and *l* vary, so that the regression formula (10) becomes

$$K_t = C \times A' + \varepsilon, \quad (11)$$

where *C* is a 1×252 vector and *A'* is the transpose of the 252×252 matrix.

The Least-square Method (LM) is employed to estimate *C* when *K* and *A* are observed. The basis of the LM method is to minimize the sum of squared errors:

$$SSE = \sum (K_{t\_original} - K_{t\_estimated})^2. \quad (12)$$

Since there are 252 variables to be estimated in vector *C* with 252 observations, one will observe some singular terms in the process of minimization, which will yield zero estimates for those $C_{i,j,k,l}$ coefficients. This is not a surprising result in this kind of estimation process.

After the estimation procedure is carried out (here by S-Plus, an advanced statistical software package), the vector *C* will be estimated with (252 less the number of singulars) non-zero numerical values and the zero values will be assigned to those $C_{i,j,k,l}$ that are considered singular. After this process, a 1×252 vector *C* is generated. This can be viewed as a filtering procedure which will single out those N^i, *(EAR)*, $(P/D)^k$ and J^l terms that are not significant to the probabilistic model in equation (10).

Furthermore, more information in the regression analysis can be obtained such as the standard error, (*t*-value) and *p*-value (*Pr((test statistic)* >|*t*| *)*) for each $C_{i,j,k,l}$ estimated. A sample output of this information is listed Table 2:

Table 2. A SAMPLE OUTPUT FOR REGRESSION ANALYSIS

	Std. Error	t value	Pr(>|t|)
1	0.046340930	-16.613806360	1.154632e-14
2	0.267453190	1.421098450	1.681520e-01
3	1.079734330	0.470970110	6.419148e-01
. . . .			
26	0.981486080	-0.597363020	5.558597e-01
27	0.443841450	1.474089110	1.534547e-01
28	NA	NA	NA
29	0.156353330	7.597914470	7.765376e-08
30	0.851771110	-1.258532740	2.203047e-01
. . . .			
247	0.044432390	-0.704020400	4.881953e-01
248	NA	NA	NA
249	NA	NA	NA
250	NA	NA	NA
251	NA	NA	NA
252	NA	NA	NA

where the NA terms correspond to those zero (singular) valued $C_{i,j,k,l}$ coefficients.

Since the critical *p*-value is provided in the output, the individual calculation formulas for standard error (and *t* value) become less important. One can find them in the text book. The use of the *p*-value for each term is to determine the significance of each $C_{i,j,k,l.}$ estimated. For example, if the *p*-value shown is 7.765376e-08, the corresponding $C_{0,1,0,0}$ is highly significant. To ensure maximum numerical accuracy, instead of abandoning all non-significant terms (such as the one shown with a *p*-value of 6.419148e-01, which is not statistically significant), we keep those very small but non-zero coefficients $C_{i,j,k,l}$ so that those highly dependent predictors remain in the polynomial in the deterministic component in equation (10).

Discussion of Regression Analysis Results:

Based on the propeller shaft torque (K_q) data set, which is similar to the K_t data, we find that at *95%* confidence level, the difference (Error) between K_q and $K_{q\_estimated}$) is estimated within the confidence interval (1.529957e-05 2.228456e-05), which is close enough to zero. However, we could not conclude that the error is zero since the *p*-value for this test is 2.2e-16, which will reject the hypothesis that K_q-$K_{q\_estimated}$=0 at almost any level of significance. Moreover, we have to admit that the K_q is over estimated by the regression model defined in equation (10), even if it is by not that much.

Comparison between the Current Model (DGSI) and the Regression Model

Being given a data set like the propeller shaft torque (K_q) data set, DGSI used a different approach to achieve the determination of the polynomial coefficients $C_{i,j,k,l}$. Based on this single data set, we found that DGSI does a very good job to resolve the estimated K_q by those $C_{i,j,k,l}$ coefficients resolved by it.

It might be affected by the algorithm used in the DGSI procedure, but one has to notice that neither the estimated coefficients $C_{i,j,k,l}$ nor the resolved $K_{q\_estimated}$ are consistent. In other words, they are independent sets of estimates and evaluations.

One also needs to notice that at 95% confidence level, DGSI produces a wider confidence interval for the paired difference (K_q-$K_{q\_estimated}$) as (0.0001111199 0.0001467610) which is a shift even more to the right. However, the mean of the differences is only 0.0001289405 which is again very close to zero. Not surprisingly, we will have to conclude that no evidence supports that the difference (K_q-$K_{q\_estimated}$) is zero. In addition, we have to make clear that the above discussions in this section are based on a single set observations. A percentage error diagram for the statistical model via S-Plus is plotted and shown in Figure 4.

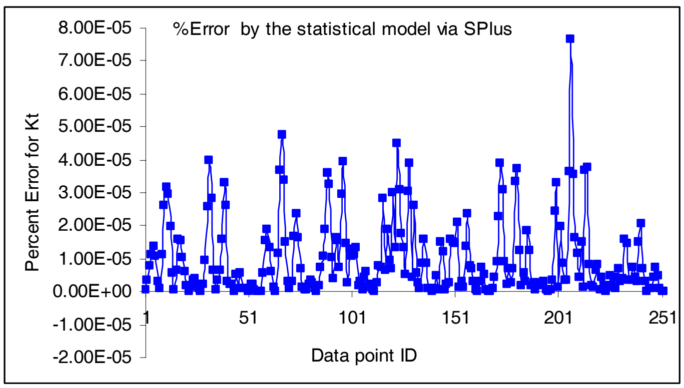

Figure 4. PERCENT ERROR FOR THE ESTABLISHED STATISTICAL MODEL AND COMPUTED BY S-PLUS.

The direct solution method for linear regression coefficient developed in this work produced much larger percentage errors (0.005% versus 0.00007%) than the statistical model implemented and computed under S-Plus. A substantial accuracy improvement can be made if the direct solution model employed some kind of iteration process that requires more CPU time. However, the percentage error produced by the direct solution method without iteration is small enough for engineering design. Each run of the statistical model under S-Plus required about 4 hours (this includes the data format time on a P3 850MHz computer with 1GB RAM) of CPU time versus a couple minutes for the direct solution model implemented in DSGI (100:1). For the 4-variable problem, the saving of CPU by DSGI is not very significant. When the data points become larger and the power of the exponents is around 10, CPU time and memory could become prohibitive. Therefore, the current direct solution model is a good alternative for engineering design applications, especially for a large number of independent variables and high exponential values.

CONCLUSION

A multiple-variable linear regression direct solution model and a statistical model were developed for marine propeller design, optimization and prototyping. Computing implementation for the direct solution model was made to create an integrated tool for the marine propeller development process. The direct solution model, without an iteration process, has a much larger error in percentage (0.005% versus 0.00007%) but is small enough for engineering design and computations. The statistical model via a commercial software package has a smaller percentage error with a requirement of much longer CPU time (100:1). For a linear regression task with a large number of independent variables and high order of exponent, the current developed model could be a better alternative.

ACKNOWLEDGMENTS

The authors would like to thank National Research Council Canada, Natural Sciences and Engineering Research Council Canada and Transport Canada for their support. We are also grateful to Mr. Mohammed Fakhrul Islam, for the assistance in running the direct solutions model code and gathering the data. Mr. Derek Yetman is also acknowledged for the proofread of the manuscript.

REFERENCES

[1] Yossifov, K., Staneva, A. and Belchev, V., 1989. “Equations for Hydrodynamic and Optimum Efficiency Characteristics of the Wageningen K_c Ducted Propeller Series”. *Fourth International Symposium on Practical Design of Ships and Mobile Units,* pp. 1-15.

[2] Liu, P., 2002. "Design and Implementation for 3D unsteady CFD Data Visualization Using Object-Oriented MFC with OpenGL". *Computational Fluid Dynamics Journal of Japan (CFDJJ)*, **11**, no. 3, pp. 335-345.

[3] Liu, P., Colbourne, B. and Chin, S., 2005. "A time-domain surface panel method for a flow interaction between a marine propeller and an ice blockage with variable proximity". *Journal of Naval Architecture and Marine Engineering*, 1(2005) pp.14-19

[4] Liu, P. and Li, K., 2002. “Programming the Bi-CGSTAB Matrix Solver for HPC and Benchmarking IBM SP3 and Alpha ES40”. *16th International Parallel & Distributed Processing Symposium,* Ford Lauderdale, Florida, pp. 269-317.

