IDENTIFICATION ACCURACY OF ADDITIONAL WAVE RESISTANCE THROUGH A COMPARISON OF MULTIPLE REGRESSION AND ARTIFICIAL NEURAL NETWORK METHODS

doi:10.2478/mape-2018-0026

Date of submission of the article to the Editor: 03/2018
Date of acceptance of the article by the Editor: 06/2018

MAPE 2018, volume 1, issue 1, pp. 197-204

Assoc. Prof. Tomasz Cepowski
Maritime University of Szczecin, **Poland**

Abstract: The article presents the use of multiple regression method to identify added wave resistance. Added wave resistance was expressed in the form of a four-state nominal function of: “thrust”, “zero”, “minor” and “major” resistance values. Three regression models were developed for this purpose: a regression model with linear variables, nonlinear variables and a large number of nonlinear variables. The nonlinear models were developed using the author's algorithm based on heuristic techniques. The three models were compared with a model based on an artificial neural network. This study shows that non-linear equations developed through a multiple linear regression method using the author’s algorithm are relatively accurate, and in some respects, are more effective than artificial neural networks.

Keywords: identification, neural networks, regression, wave resistance

INTRODUCTION

Reaching the required service speed on board a ship is a serious hydromechanical question. A ship operating speed depends, among others on the parameters and operating conditions of the propulsion system and total hull resistance. Arribas (2007) and Bhattacharyya (1978) noticed that one component of total hull resistance is additional wave resistance, which is:

- ship motion at high sea,
- around 30-50% of the ship’s total resistance,
- a substantial reduction of service speed,
- dependent on hull dimensions and shape, as well as other factors.

Therefore, the prediction of any added resistance is a real challenge for the naval architects, who have to focus on any economic requirements connected to the selection of propulsion system parameters, fuel consumption and voyage time estimate. An assessment of added resistance in waves is also an essential element of various computer systems for the planning of a voyage. The effectiveness of such a system operation depends, first and foremost, on the accuracy of the approximation of various ship characteristics based on the simplified data, such as the main hull dimensions, state of loading, ship motion parameters and statistical wave parameters.

The added resistance of a ship in waves is difficult to determine and usually is predicted by the use of a model or numerical methods (Sigmund et al., 2018; Ji et al., 2017; Duan et al., 2013). An accurate prediction of added resistance requires a large amount of input data, such as vessel hull shape and dimensions, as well as wave conditions (Rawson et al 2001; Watson 1998) .

The mean added wave resistance usually is calculated using a function which initially describes the resistance contribution from a regular sine wave. Then, applying the superposition principle to the spectral distribution of the irregular wave, the mean resistance of an irregular (statistical) wave is calculated. The added resistance of a regular wave is calculated using various methods, most frequently the Gerritsma–Beukelman method (Gerritsma and Beukelman, 1972) or the Boese method.

Alternatively, simplified models that enable us to identify the added wave resistance are sometimes used. In these methods, the additional wave resistance values have linguistic form, eg "small resistance" - "high resistance". These methods mean we can estimate the added wave resistance level and can be used to develop a mathematical model based on observations carried out in real-life conditions on the ship. The advantage of this solution is the ability to develop a model to assess the added wave resistance without having to measure the phenomenon on a ship. This model could have practical application for navigation route optimization systems.

The author's article (Cepowski, 2007) presents an application of artificial neural networks to identify additional wave resistance depending on ship motion and wave parameters. This article elaborated on investigations of artificial neural networks which make it possible to identify additional wave resistance expressed in the form of a four-state nominal function equal to:

- thrust,
- lack of the resistance,
- small resistance value,
- large resistance value.

The artificial neural networks elaborated here were characterized by high accuracy within a wide range of ship motion and wave parameters values. Only 86 of the 2,646 cases were incorrectly identified. The disadvantage of this method was that it was complicated and difficult to interpret the mathematical models.

Other methods can also be used to identify any phenomenon occurrence, e.g. linear or logistic probability models.

The occurrence or absence of a predicted event are indicated in a logistic probability model. This model enables the user to calculate event probability and is often used to analyse surveys. The limitation of this method is a binary dependent variable, where the output can take only two values, "0" and "1". This method cannot be used for more than two various states, and in such cases, can be only applied to a linear probability model. Other advantages of the linear probability model are the interpretability of results and computing speed than compared to the logistic probability model.

A linear probability model often fits equally well, and is almost the same as logistic model. The probability linear model offers an unlimited number of the states. As shown in (Hellevik 2007), logistic and linear regression models often display almost identical results, but the logistic model estimates are much more complicated to interpret than the linear.

Consequently, the aim of this article was to develop an added wave resistance recognition model, simply by the use of regression methods alone and to also compare research results with those obtained using an artificial neuron theory model based on work presented in (Cepowski, 2007).

It was not possible to use a logistic regression method for this purpose as the added wave resistance took up 4 states. Therefore, a linear regression model was used to identify added wave resistance in these studies. Additionally, a heuristic algorithm to semi-automatically discover the best nonlinear equation was created by author for this purpose.

RESEARCH METHOD

To achieve the goals set out in the research, the same assumptions as in (Cepowski, 2007) have been made, i.e.:

- the investigations were performed for a B-517 bulk carrier in ballast loading condition with the following parameters:
  - Overall length: Lc = 198 m
  - Length between perpendiculars: Lpp = 185 m

  - Breadth: B = 24.4 m
  - Design draught: T = 11 m
  - Displacement: D = 18069 t
- the following operational parameters of the ship were taken into account:
  - ship speed V ranging from 0 to 15 knots, every 5 knots
  - wave encounter angle b = 0° (following waves), 15°, 30°, 60°, 75°, 90°, 105°, 120°, 150° 175°, 180° (heading waves)
  - significant wave height h ranging from 1 to 9 m, every second m
  - characteristic wave period p = 6 ÷ 20 s, every second s.
- added wave resistance was expressed in the form of a four-state nominal function R_w equal to:
  - "0" – thrust (for additional wave-generated resistance less than 0 kN)
  - "1" – zero resistance (for resistance values from 0 to 30 kN)
  - "2" – minor resistance value (for resistance values from 30 to 100 kN)
  - "3" – major resistance value (for resistance values exceeding 100 kN).

F functions which serve to identify added resistance, can be created in accordance with the following formula:

$$X(V,b,p,h) \xrightarrow{f} R_W \tag{1}$$

where:

X – set of input operational parameters such as ship speed *V*, wave encounter angle *b*, characteristic wave period *p* and significant wave height *h*

R_W – additional wave resistance expressed in the form of a four-state nominal function

f – a regression function search, enabled to identify added resistance R_W.

A Multiple Linear Regression, and a method developed by the author were applied to discover function f. The best combinations of the ship operational parameters were randomly searched through all their possible combinations in the author's method. The general algorithm scheme is shown in fig 1.

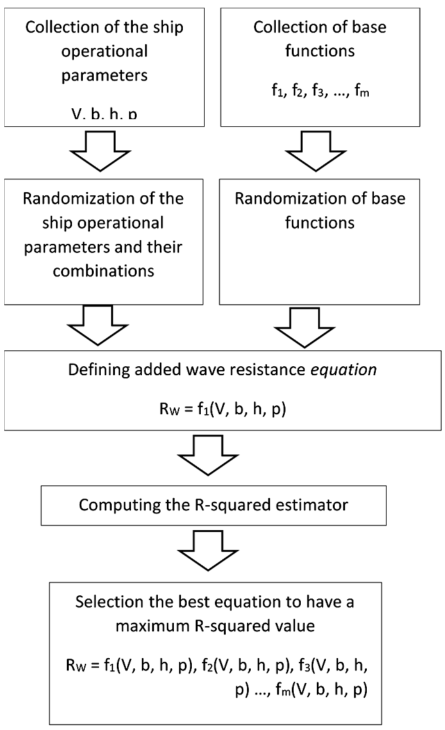

Fig. 1. The general algorithm scheme, where: R_W is estimated added wave resistance, V, b, h, t – operational parameters, f is a formula to calculate added wave resistance, m is the number of formulas

The base function collection included 360 arrays of nonlinear, exponential., power and logarithmic functions. NdCurveMaster software was developed on the basis of this algorithm by the author. The software (ndCurveMaster ver 3.2 2017) was applied to develop regression equations presented in this paper.

RESULTS

Three models of regression were developed to compare research results with those presented in (Cepowski) only by the use of:

- A multiple Linear Regression method,
- An algorithm developed by the author (presented in Figure 1),
- An algorithm developed by the author for a large number of independent variables

A significant level of alpha was equal to 0.05 in the investigations.

The most effective regression equations for additional added resistance identification were found to be as follows:

- with the use of Multiple Linear Regression method:

$$R_W = a_0 + a_1 \cdot b + a_2 \cdot V + a_3 \cdot p + a_4 \cdot b \cdot V + a_5 \cdot b \cdot h + a_6 \cdot b \cdot p + a_7 \cdot V \cdot h + a_8 \cdot h \cdot p + a_9 \cdot b \cdot V \cdot h + a_{10} \cdot V \cdot h \cdot p + a_{11} \cdot b \cdot V \cdot h \cdot p \quad (2)$$

- with the use of the author's algorithm:

$$R_W = a_0 + a_1 \cdot b^{4.7} + a_2 \cdot V^{1/5} + a_3 \cdot h^{-0.3} + a_4 \cdot p^{14} + a_5 \cdot b^{1.2} \cdot V^{1/5} + a_6 \cdot b^{2.2} \cdot h^{1/6} + a_7 \cdot b^{1.3} \cdot p^{2.1} + a_8 \cdot V^{1.2} \cdot h^{0.4} + a_9 \cdot V^{1.8} \cdot exp(p)^{-4} + a_{10} \cdot 1/4^h \cdot ln^6 p + a_{11} \cdot b^7 \cdot V^{0.7} \cdot h^{1.2} + a_{12} \cdot V \cdot h^{1/7} \cdot p^{0.4} + a_{13} \cdot b^{4.8} \cdot V^{0.6} \cdot h^{0.4} \cdot p^{1/2} \quad (3)$$

- with the use of the author's algorithm for a large number of independent variables:

$$R_W = a_0 + a_1 \cdot b^6 + a_2 \cdot V^{1/4} + a_3 \cdot ln^3 p + a_4 \cdot b^{1.6} \cdot V^{1/4} + a_5 \cdot b \cdot h^{0.6} + a_6 \cdot b^{1.2} \cdot p^{1.9} + a_7 \cdot V^{0.9} \cdot h^{2.6} + a_8 \cdot V^{0.9} \cdot p^{1/2} + a_9 \cdot ln^7 h \cdot ln^3 p + a_{10} \cdot b^{5.6} \cdot V^{0.8} \cdot h^{0.9} + a_{11} \cdot V^{1/21} \cdot 1/2^h \cdot p^{1.5} + a_{12} \cdot b^{3.5} \cdot V^{0.8} \cdot h \cdot p^{1/15} + a_{13} \cdot 1/2^V \cdot 1/2^h \cdot p^2 + a_{14} \cdot b^{3.8} \cdot V^{1/3} \cdot h^{1/9} + a_{15} \cdot b^{1.8} \cdot V^{0.8} \cdot h^{1.6} + a_{16} \cdot b^{1.5} \cdot V^{1/4} + a_{17} \cdot V^{0.8} \cdot h^{1/17} \cdot exp(p)^{-4} + a_{18} \cdot b^8 \cdot p^{1/17} + a_{19} \cdot h^{-6} + a_{20} \cdot b^2 \cdot h^{1/21} + a_{21} \cdot b^{1.2} \cdot h^{0.6} + a_{22} \cdot h^{-3} + a_{23} \cdot b^3 + a_{24} \cdot b^{0.9} \cdot p^{2.1} + a_{25} \cdot h^{3.2} \cdot p^{0.9} + a_{26} \cdot b^{12} \cdot V^{1.4} \cdot h^{1.6} \quad (4)$$

where:
Rw – added wave resistance in the form of a four-state nominal variable: '0' – resistance thrust, '1'– zero resistance, '2'– minor resistance, '3'– major resistance
V – ship's speed [kn],
b – wave encounter angle [deg],
p – characteristic wave period [s],
h – significant wave height [m],
$a_{1, ..., n}$ – coefficients shown in Tab. 4 – 6.

Tables 1- 6 present variance and regression analysis of the above equations. The values of standard SE and the R-squared errors relating to elaborated relationships (2) – (4) are given in Tab. 7. An analysis of variance presented in Tables 1-3 shows that all equations are statistically significant. Regression analysis presented in Tables 4-6 show that all variables in equations (2) - (4) are also statistically significant. Table 7 shows that the equation (2) is characterized by satisfactory accuracy, while equations (3) and (4) are characterized by good accuracy.

Table 1
Analysis of equation variance (2), where: df – degrees of freedom, SS – sum of squares, MS - mean square

	df	SS	MS	F test	p-value
Regression	11	1619.297	147.2088	585.0595	5.55E-16
Errors	2452	616.956	0.251613		
Total	2463	2236.253			

Table 2
Analysis of equation variance (3), where: df – degrees of freedom, SS – sum of squares, MS - mean square

	df	SS	MS	F test	p-value
Regression	13	1910.69084	146.976218	1106.06194	5.55E-16
Errors	2450	325.561996	0.1328824		
Total	2463	2236.25284			

Table 3
Analysis of equation variance (4), where: df – degrees of freedom, SS – sum of squares, MS - mean square

	df	SS	MS	F test	p-value
Regression	26	1981.717	76.21989	729.7518	5.55E-16
Errors	2437	254.5357	0.104446		
Total	2463	2236.253			

Table 4
Regression analysis for the equation (2)

	Value of a	Std Error	t-Value	P > |t|
a0	-2.41E-01	0.083968	-2.87E+00	0.004109
a1	0.012815	0.000928	13.80582	8.27E-42
a2	0.047987	0.006821	7.035463	2.57E-12
a3	0.040385	0.006559	6.157143	8.63E-10
a4	-2.23E-04	6.68E-05	-3.34E+00	0.00084
a5	0.00176	9.12E-05	19.29056	2.60E-77
a6	-6.87E-04	5.54E-05	-1.24E+01	2.65E-34
a7	0.010699	0.001609	6.64953	3.61E-11
a8	-4.85E-03	0.000668	-7.25E+00	5.43E-13
a9	-1.11E-04	1.64E-05	-6.74E+00	1.93E-11
a10	-4.91E-04	0.000105	-4.67E+00	3.2E-06
a11	6.4E-06	9.62E-07	6.651474	3.57E-11

Table 5
Regression analysis for the equation (3)

	Value of a	Std Error	t-Value	P > |t|
a0	0.2490437	0.0615499	4.0462083	0.0000537
a1	-1.01E-10	2.32E-12	-4.36E+01	0.00E+00
a2	0.9500981	0.027393	34.6839572	8.59E-215
a3	-6.58E-01	0.0856514	-7.68E+00	2.28E-14
a4	6.12E-20	1.86E-20	3.2891706	0.0010191
a5	-2.76E-03	0.000082	-3.37E+01	9.57E-205
a6	0.0000614	8.84E-07	69.4327879	0.00E+00
a7	-4.65E-06	1.46E-07	-3.19E+01	4.52E-187
a8	0.0138254	0.0018719	7.3857777	2.07E-13
a9	-4.24E+07	9.72E+06	-4.36E+00	0.0000135
a10	0.004668	0.0003298	14.1552743	8.78E-44
a11	-5.70E-18	3.31E-19	-1.72E+01	6.65E-63
a12	-1.80E-02	0.0020314	-8.87E+00	1.34E-18
a13	1.40E-12	6.42E-14	21.8819682	4.26E-97

Table 6
Regression analysis for the equation (4)

	Value of a	Std Error	t-Value	P > |t|
a0	0.447867	0.063296	7.075788	1.93E-12
a1	-1.10E-12	4.95E-14	-2.22E+01	6.19E-100
a2	0.644304	0.029564	21.79334	2.38E-96
a3	-1.62E-02	0.002715	-5.97E+00	2.80E-09
a4	-1.19E-02	0.000767	-1.55E+01	1.47E-51
a5	-1.81E-02	0.001332	-1.36E+01	1.08E-40
a6	-5.83E-05	3.7E-06	-1.56E+01	1.76E-52
a7	0.000243	2.38E-05	10.21183	5.30E-24
a8	-1.37E-02	0.001193	-1.15E+01	1.21E-29
a9	0.000199	6.78E-05	2.93505	0.003366
a10	-9.28E-14	5.59E-15	-1.66E+01	1.23E-58
a11	0.024201	0.001563	15.48334	1.19E-51
a12	3.44E-09	1.73E-10	19.93332	5.07E-82
a13	0.004845	0.000509	9.513478	4.26E-21
a14	1.16E-08	6.94E-10	16.7592	9.45E-60
a15	-2.51E-06	1.34E-07	-1.87E+01	4.39E-73
a16	0.017749	0.00119	14.91465	3.32E-48
a17	-4.95E+08	1.02E+08	-4.87E+00	1.2E-06
a18	1.63E-17	7.50E-19	21.79234	2.42E-96
a19	6.074821	0.927909	6.546787	7.14E-11
a20	-2.80E-04	2.94E-05	-9.53E+00	3.77E-21
a21	0.008725	0.000476	18.32714	2.17E-70
a22	-6.69E+00	0.968204	-6.91E+00	6.17E-12
a23	4.6E-06	2.81E-07	16.41442	1.64E-57
a24	0.000103	9.2E-06	11.20518	1.88E-28
a25	-8.60E-05	2.83E-05	-3.04E+00	0.002413
a26	2.14E-30	2.61E-31	8.201639	3.79E-16

Table 7
Statistics of the elaborated relationships (2) – (4)

	Equation (2)	Equation (3)	Equation (4)
R-Squared	0.724112	0.854416	0.886178
Adjusted R-Squared	0.722874	0.853644	0.884963
Multiple R-Squared	0.850948	0.924346	0.94137
Standard error SE	0.501611	0.36453	0.323182

An identification of added wave resistance was carried out by rounding the actual values determined by means of equations (2) - (4) to integer values: 0, 1, 2, 3.
Identification results were compared with reference values and artificial neural network statistics presented in (Cepowski, 2007) and shown in Fig. 2.

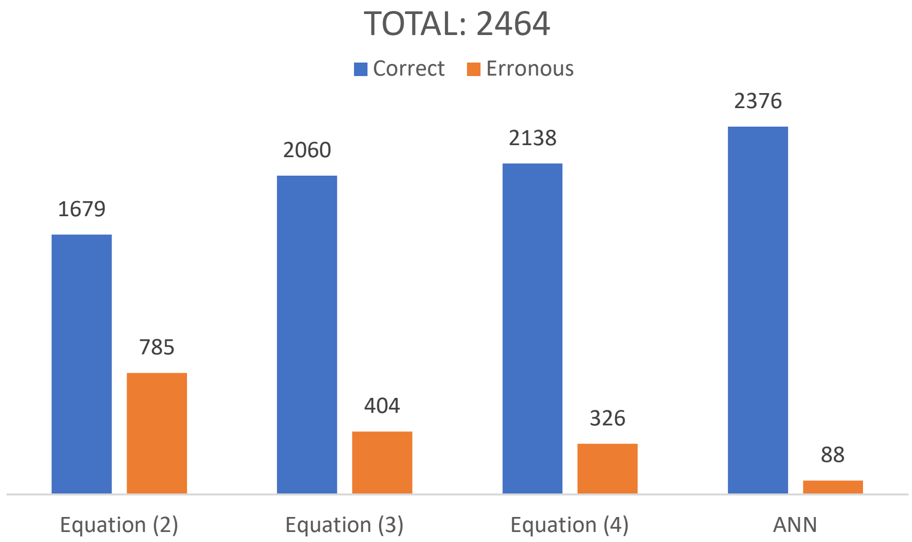

Fig. 2. Statistics applied to classification problems of the artificial neural network and regression (2) – (4) identifying the additional wave-generated resistance

CONCLUSIONS

The presented research proves that it is possible to carry out an identification simply by the use of a multiple linear regression method. But, regressions developed simply by the use of multiple linear regression methods are not particularly accurate. Only adding non-linear variables to the model, and increasing their number, leads to more accurate results. Heuristic techniques developed by the author for this purpose were found to be effective.

A logistic regression method only enabled the identification of a binary variable where the output can take only two values, "0" and "1". In contrast, the MLR method allows the identification of a multi-state variable where the output can take a number of values. In this study a four-state variable was successfully identified using a MLR-based method.

Classification statistics of linear regression equations are less effective than the artificial neural network statistics. The most accurate regression equation found 326 erroneous solutions, while the neural network error only displayed 88 solutions (almost three times less). Which offers vast advantages using artificial neural networks over a multiple linear regression method.

Without a doubt, the form of regression equation developed through the use of multiple linear regression is clear, and it is simple to interpret the influence of individual variables in regression equations. The interpretation of regression equation coefficients is less difficult than the coefficients of the artificial neural network. Through these methods, discovering new phenomena is possible on the basis of regression equations.

A regression equation has a simpler form, fewer elements and is faster to calculate than the neural network, and this is vitally important from point of view of numerical calculations.

For example, the presented ANN contains almost 140 coefficients, while equation (4) has only 27.

The probability of overfitting is less likely during the development of regression equations. There was no need to control and prevent overfitting, because the number of data was found to be larger in the presented studies.

However, models developed using artificial neural networks always require the occurrence of overfitting. About 25% of the data is lost due to the need to verify the occurrence of overfitting.

These investigations confirm (Hevelik, 2017), that identification accuracy depends on data model fitting in a logistic or linear regression method.

Equations (2) and (3) have a similar number of elements, ie equation (2) - 12 and equation (3) - 14 coefficients. Independent variables in equation (2) are linear, while in equation (3) they are non-linear. This non-linearity meant we were able to obtain practically twice as much identification accuracy, which is shown in Fig. 2. The main reason for this is the non-linear influence of operating parameters on additional wave resistance.

REFERENCES

Arribas, FP. (2007). Some methods to obtain the added resistance of a ship advancing in waves. Ocean Engineering, 34 (2007), pp. 946–955.

Bhattacharyya, R. (1978). Dynamics of Marine Vehicles. John Wiley & Sons Inc, 1978.

Cepowski, T. (2007). Application of artificial neural networks to approximation and identification of sea-keeping performance of a bulk carrier in ballast loading condition. POLISH MARITIME RESEARCH 4(54) Vol 14; pp. 31-39 DOI: 10.2478/v10012-007-0037-6.

Cepowski, T. *ndCurveMaster vr. 3.2* [online copy] Available from: https://www.ndcurvemaster.com [Accessed: 4 Apr. 2017]. Szczecin, Poland.

Duan, WY, Li, CQ. (2013). Estimation of Added Resistance for Large Blunt Ship in Waves. JOURNAL OF MARINE SCIENCE AND APPLICATION, Volume: 12 Issue: 1, pp: 1-12, DOI: 10.1007/s11804-013-1177-6.

Gerritsma, J., Beukelman, W. (1972). Analysis of the Resistance Increase in Waves of a Fast Cargo-ship. International Shipbuilding Progress, No 18(217).

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy, International Journal of Methodology, Quality & Quantity, Volume 43, Issue 1, pp 59–74.

Ji, ZB; Lei, N (2017) Study on Calculation Method of Added Resistance of Ships in Irregular Waves. NAVAL ENGINEERS JOURNAL. Volume: 129 Issue: 4, pp. 123-134.

Rawson, K.J. and Tupper, E.C. (2001). Basic Ship Theory. Ship Dynamics and Design. Volume 2. Fifth edition. Butterworth-Heinemann.

Sigmund, S, el Moctar, O. (2018). Numerical and experimental investigation of added resistance of different ship types in short and long waves. OCEAN ENGINEERING, Volume: 147, pp.: 51-67.

Watson, D.G.M. (1998). Practical Ship Design, Volume 1, Elsevier Science.