

Multivariate Regression Analysis For Screening Process of Reliability Assessment

V.J. Kurian^{1,a}, M.C. Voon^{2,b}, M.M.A. Wahab^{3,c}, N.A. Iskandar^{4,d}, M.S. Liew^{5,e}

^{1,2,3,4,5}Civil Engineering Department, Universiti Teknologi PETRONAS, Tronoh, Malaysia

^akurian_john@petronas.com.my, ^bvoonminchin@gmail.com^b,

^cmubarakwahab@petronas.com.my, ^dhabilaiskandar@gmail.com, ^eshahir_liew@petronas.com.my

Keywords: Regression, Reserve Strength Ratio, Reliability.

Abstract

Reliability is one major concern in the Oil and Gas industry to date. Reliability issues due to aging and increasing environmental loading are common for jacket platforms in Malaysia. Authors have studied system reliability assessment on existing jacket platforms in Malaysian waters and concluded that the current reliability assessment in the industry practice is tedious. The scope of the study includes regression analysis using three different methods to generate regression equation which could serve as a simplified reliability prediction model. Regression equations generated in this study seek to improve the current reliability assessment of jacket platforms in Malaysia by providing means for screening process before proceeding to tedious reliability analysis.

Introduction

Life extension of offshore platforms has become one major concern in the Malaysian Oil and Gas industry because majority of the jacket platforms are operating beyond their design life. The marginal oil field or oil field that seems impossible to harvest ten years back, is now harvested with advancement in technology via the enhanced oil recovery method. Thus, there are increasing demands to extend the life of these platforms. Modifications or upgrading result in platforms being subjected to higher loading for which they may not be originally designed for [1]. An existing platform should undergo assessment process if there is addition of personnel, increased loading on structure, and damage found during inspection [2]. Pushover Analysis is incorporated in Risk Based Inspection (RBI) to provide Reserve Strength Ratio (RSR) value to proceed to Simplified System Reliability Analysis (SSRA) to obtain the probability of failure and reliability index of an assessed platform [3, 4]. RBI will categories platform based on probability of failure and recommends corrective plan accordingly. Instead of Pushover Analysis which requires a lot of technical competency and time, authors formulated a relationship between RSR and other variables to produce a simplified prediction model using regression analysis.

Regression Analysis

The regression analysis analyzes several variables to formulate relationship between a dependant variable and one or more independent variable(s) to arrive at a function that fits the data and returns an array of values that describes the function. The purpose of regression analysis to establish a formulation between the variables to produce an equation in which the parameters such as wave heights, initial air gap, water depth, and number of years after installation can be input to obtain the RSR value. The method using this simplified prediction model will be very useful to eliminate the lengthy process and costly work of conducting conventional reliability analysis [5]. Linear Regression has a curve fitting protocol carried out in a way that the data will be represented by a linear function. By utilizing the Regression Analysis provided in the Microsoft Excel's Data Analysis Tool, Linear Regression equation obtained from the regression analysis returns the best fit for linear function to represent the variables. The linear regression equation generated is in the form

shown in Equation 1, where y is the 'dependant' variable, RSR value and x is the 'independent' variable, input parameters [6]. Non Linear Regression could be implemented to fit a non linear function to the data for the data which are not described by the linear function. There are two Non Linear Regression approach that will be discussed in this paper, namely LOGEST and log. LOGEST is one of the regression analysis functions in Microsoft Excel to generate an exponential curve that fits the data [5]. The log approach is similar to LOGEST approach which generates an exponential curve with slight difference [7]. The LOGEST and log regression equation generated are as in the form shown in Equation 2 and 3, respectively;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (1)$$

$$y = \beta_0 * m_1^{x_1} * m_2^{x_2} * m_3^{x_3} * m_4^{x_4} \quad (2)$$

$$y = \beta_0 * x_1^{m_1} * x_2^{m_2} * x_3^{m_3} * x_4^{m_4} \quad (3)$$

Methodology

Prior to the regression analysis, the data ought to be sorted according to categories. 102 sets of RSR values accompanied by various other variable data such as numbers of years after installation, water depth, initial air gap, and wave height were obtained from the operating platforms. Since the data set is huge and varies from platform to platform, one regression equation generated is not possible to represent the whole set of data. Hence, the data was divided to 5 categories according to wave height. 92 data set was used for regression analysis while the 10 data set was used for sample testing. Hence, there were 2 testing samples for each category [8]. The methodology flow chart of the regression analysis is shown in figure 1.

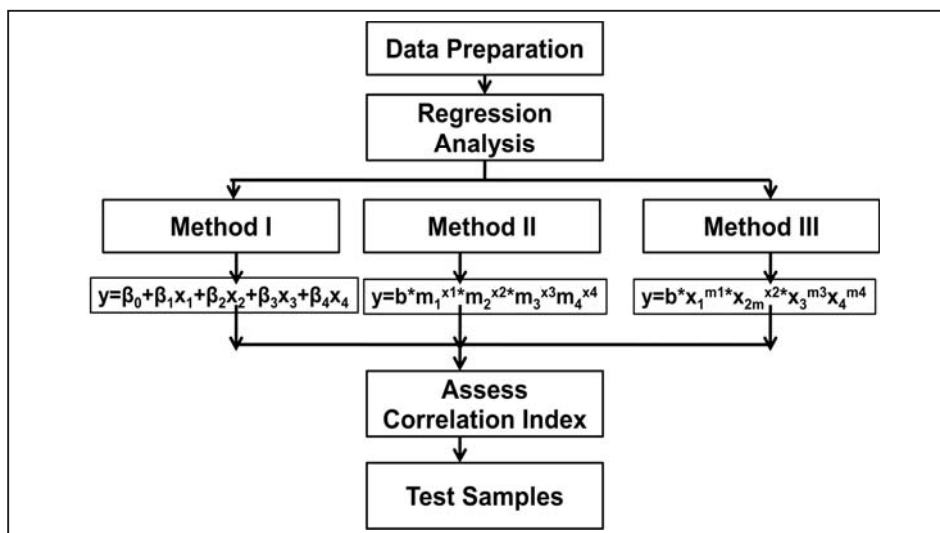


Fig. 1. The flowchart of the regression analysis methodology

Regression Analysis aims to generate formulae that will predict the RSR value based on the variable parameters. Linear and Non Linear Regression Analysis were used to generate the formulae representative of all jacket platforms in Malaysia. Correlation analysis was performed on 102 data sets and Table 1 shows the correlation coefficients for each input variable. Initial air gap and wave height are positively correlated with RSR, and have correlation coefficients of 0.139 and 0.295, respectively. The number of years installed and water depth are negatively correlated with RSR with correlation coefficients of -0.332 and -0.153, respectively. Generally, each variable contributes towards the predicting power of the regression equation generated and elimination of any variable results in lower predicting power. No matter how insignificant the correlation coefficient of each

variable, the inclusion of more variable to regression produces regression equation of higher predicting power, as long as the variables are correlated.

Table 1. Correlation Coefficients for Each Variables

CORRELATION ANALYSIS FOR 102 DATA SETS					
VARIABLES	No. of Years Installed	Water Depth	Initial Air Gap	Wave Height	RSR
No. of Years Installed	1.000				
Water Depth	-0.279	1.000			
Initial Air Gap	-0.412	0.182	1.000		
Wave Height	-0.414	0.409	0.539	1.000	
RSR	-0.332	-0.153	0.139	0.295	1.000

The correlation index or also known as r^2 value expresses the proportion of variance in the dependant variable explained by the independent variables. Increasing r^2 value towards 1 shows that the functions fit the data more accurate and decreasing r^2 value towards 0 means that the function is less accurate in representing the data [6]. Microsoft Excel 2007 was used to assess three different methods of performing regression in Excel. Method I was the Linear Regression data analysis package in Excel which produces a best fit linear function to represent the variables. Method II was the LOGEST function, also available in Excel, which could calculate an exponential curve to fit the data. The equation formulated was in a form of exponential function. Method III was the log approach where the regression analysis was performed using the logarithms of parameters as shown in Equation 5 and then transformed to format of Equation 6 [7].

$$\log R = \log A + x \log B + y \log C + z \log D \quad (4)$$

$$R = AB^x C^y D^z \quad (5)$$

The data was broken down into groups according to the wave heights as analysis conducted on full set of data may returned regression equation which has low prediction power. At the end of the analysis, formulae to predict the RSR value by input varying the independent variable such as number of years after installation, wave height, water depth and initial air gap is generated. The symbols representing of each variable in the formulae are shown in Table 2.

Table 2. Parameters for Method I, II, and III

Symbol		Remarks
1	y	RSR
2	$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, b, m_1, m_2, m_3, m_4$	Constants/coefficients
3	X_1	Numbers of year after installation
4	X_2	Water Depth
5	X_3	Initial Air Gap
6	X_3	Wave Height

Result

The formulae generated by the regression analysis are shown in Table 3. Variables can be input to obtain the RSR value straightaway from the formulae. The function generated using method I is linear fit function and the function generated using method II and III are exponential fit functions. Correlation index among all the methods are compared to each other as in Table 4 to observe the method which best describes the relationship between the independent variables with the dependant variable.

Table 3. Formulas Derived (Method I, II, and III)

Method I	Method II	Method III
Wave Height < 8m	Wave Height < 8m	Wave Height < 8m
$y = 4.2626 - 0.1844X_1 - 0.02035X_2 + 0.02118X_3 + 0.8625X_4$	$y = 3.7657 * 0.9321^{x1} * 0.9935^{x2} * 1.01619^{x3} * 1.4048^{x4}$	$y = 516.02 * X_1^{-2.0822} * X_2^{-0.2168} * X_3^{0.0387} * X_4^{1.467}$
8m ≤ Wave Height < 10m	8m ≤ Wave Height < 10m	8m ≤ Wave Height < 10m
$y = 0.9857 - 0.1290X_1 - 0.01892X_2 - 0.1017X_3 + 0.7862X_4$	$y = 0.9909 * 0.9631^{x1} * 0.9928^{x2} * 0.9560^{x3} * 1.3233^{x4}$	$y = 56.17 * X_1^{-1.1449} * X_2^{-0.0861} * X_3^{-1.1164} * X_4^{0.4894}$
10m ≤ Wave Height < 12m	10m ≤ Wave Height < 12m	10m ≤ Wave Height < 12m
$y = -3.207 - 0.08005X_1 - 0.02193X_2 + 0.05564X_3 + 0.8343X_4$	$y = 0.4082 * 0.9726^{x1} * 0.9917^{x2} * 1.0209^{x3} * 1.3089^{x4}$	$y = 0.03956 * X_1^{-0.5061} * X_2^{-0.3257} * X_3^{0.1395} * X_4^{2.8603}$
12m ≤ Wave Height < 14m	12m ≤ Wave Height < 14m	12m ≤ Wave Height < 14m
$y = 2.7850 - 0.01571X_1 - 0.03179X_2 - 0.07725X_3 + 0.2044X_4$	$y = 3.3389 * 0.9954^{x1} * 0.9878^{x2} * 0.9722^{x3} * 1.0572^{x4}$	$y = 0.8983 * X_1^{-0.1171} * X_2^{-0.3585} * X_3^{-0.3348} * X_4^{1.3684}$
14m ≤ Wave Height	14m ≤ Wave Height	14m ≤ Wave Height
$y = -3.6447 + 0.01861X_1 - 0.01880X_2 - 0.01872X_3 + 0.4294X_4$	$y = 0.5458 * 1.0012^{x1} * 0.9956^{x2} * 0.9770^{x3} * 1.1261^{x4}$	$y = 0.02103 * X_1^{0.01932} * X_2^{-0.2107} * X_3^{-0.3549} * X_4^{2.2621}$

Table 4. Comparison Table for Correlation Index, r^2 for Two Methods

Regression Analysis			
Group	Method I	Method II	Method III
H < 8m	0.5166	0.4704	0.4923
8m ≤ H < 10m	0.3822	0.4291	0.3695
10m ≤ H < 12m	0.7777	0.7323	0.6965
12m ≤ H < 14m	0.4842	0.4926	0.3890
14m ≤ H	0.6198	0.5650	0.5751

Ten set of data samples were used to conduct a validation test using the regression equation generated by the three approaches. There were 2 test samples in each category. Table 5 shows the comparison of percentage of sample testing for all three methods studied. The results show that method III has the least percentage error compared to method I and II. Figure 2 shows the comparison of validated RSR with the estimated RSR of all three methods from the training data sets. Figure 3 shows the comparison of validated RSR with the estimated RSR of all three methods from the testing data sets. It can be observed that the regression equation predicts RSR towards the correct convergence of the RSR value.

Table 5. Comparison of Percentage Error for All Three Methods

Validated RSR	METHOD I		METHOD II		METHOD III	
	RSR	% error	RSR	% error	RSR	% error
1.61	0.86	46.74	1.19	25.81	1.35	16.01
3.09	3.85	24.70	3.92	26.84	3.35	8.53
2.06	3.45	17.65	2.87	2.11	2.72	7.33
2.93	2.49	20.68	1.99	3.46	2.10	2.12
1.96	2.27	15.68	2.14	9.23	2.20	12.25

2.40	2.43	1.22	2.24	6.80	2.26	5.65
3.15	3.37	6.85	3.34	6.05	3.08	2.23
2.73	3.75	37.27	3.76	37.68	3.41	24.82
1.99	1.63	17.85	2.01	1.14	1.97	1.11
3.37	3.32	1.45	3.13	7.15	3.16	6.08
Average % error		19.01		12.63		8.61

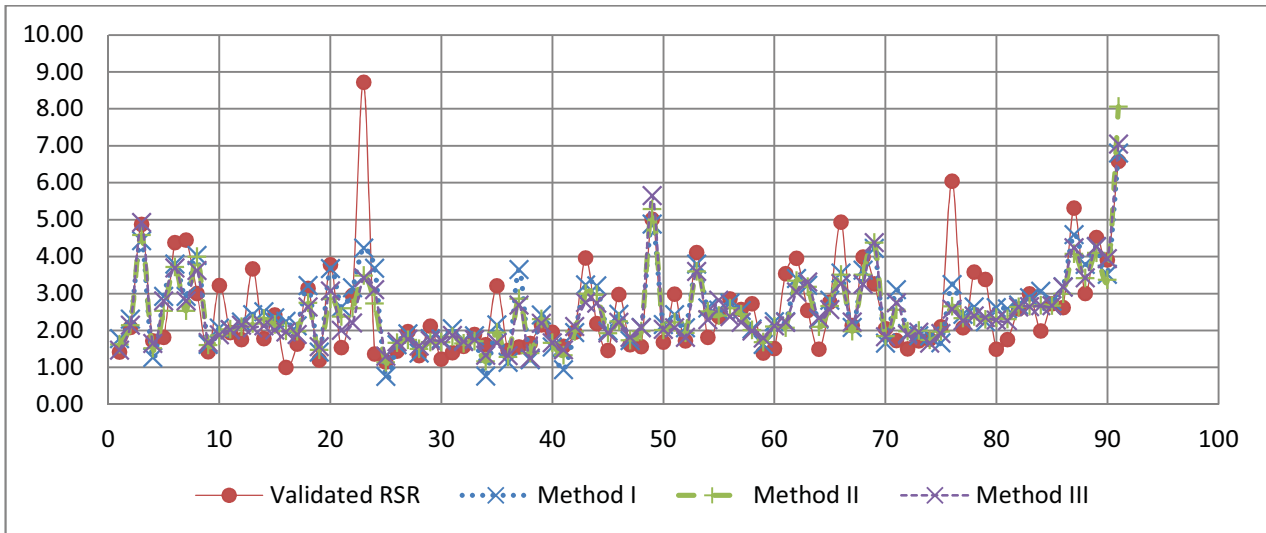


Fig. 2. The comparison graph between Validated RSR and Estimated RSR of training data set

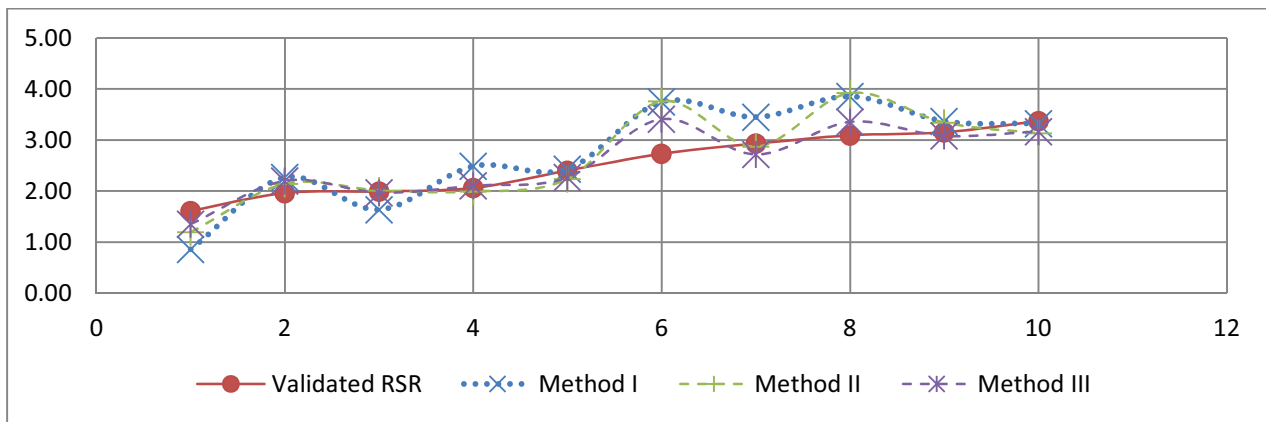


Fig. 3. The comparison graph between Validated RSR and Estimated RSR of testing data

Discussion

Table 4 shows the correlation indexes for the 3 methods studied. It was observed that the correlation index is not near to 1.0. This may be due to a lot of uncertainties contributing to a platform's RSR, instead of just the four variables which are number of years after installation, water depth, initial air gap, and wave height. However, the regression equations obtained show promising predicting power as seen in Figure 3 and Table 5 where the predicted RSR do not differ much from the actual validated RSR value. This means that by utilizing the four variables, reliability engineers can use the regression equations provided to estimate the RSR value. This however does not confirm reliability of a jacket. If the RSR estimated shows value within the range of 1.0-2.0, reliability engineers can go ahead and conduct pushover analysis to determine accurate RSR values. It is noted that the data used for regression consists of 102 validated RSR results and available parameters from platforms in Malaysian waters. The regression equation generated displays significant prediction power, especially method III which gives percentage error at 8.61%. It is recommended that for further

study, more numbers of platforms and other parameters such as current velocity, wind speed, deck loading, and etc can be considered. More numbers of platform considered will improve on the data size hence it can represent the Malaysian jacket platform more in general. Besides, inclusion of more parameters may increase the prediction power of the regression equation produced as well.

Conclusion

The following conclusion can be made using the results provided from above:

1. Regression equation generated shows significant prediction power for preliminary reliability screening process.
2. Regression equation produced using method III which is a Non Linear regression produces the best result among the three methods used.

References

- [1] N.W. Nichols, Petronas Carigali Sdn Bhd; T.K.Goh, Petronas Research & Scientific Services Sdn Bhd; H. Bahar, Petronas Carigali Sdn Bhd, "Managing Structural Integrity for Aging Platform", Proceedings, SPE Asia Pacific Oil and Gas Conference and Exhibition, Adelaide, Australia, No SPE101000, September 2006
- [2] S. A. Kheiri and M.R. Bahaari, "Evaluation of Ultimate Strength of Jacket Type Structures under Marine Loading", Proceedings, 7th International Conference on Coasts, Ports and Marine Structures
- [3] Narayanan S.P. and M. Kabir B.M.A, "Structural Integrity Management for Fixed Offshore Platforms in Malaysia", World Academy of Science , Engineering and technology 58 2009
- [4] Group Technical Solution (GTS) (2011), "Global Ultimate Strength Analysis for SBO Platform: SMJT-E", Report ID: RD-PLE-CSP-RBI-S309
- [5] Pattaradanai Pueksap-anan, "Sensitivity Analysis for RSR Offshore Steel Type Platforms", M. Sc. Thesis, Asian Institute of Technology, School of Engineering and Technology, Thailand, August 2010.
- [6] A.M. Brown, "A Step-by-step Guide to Non-linear Regression Analysis of Experimental Data Using a Microsoft Excel Spreadsheet", Computer Methods and Programs in Biomedicine 65 (2001) 191-200
- [7] C.Y. Ng, V.J. Kurian, and M.A. Yassir, " Regression Analysis for the Dynamic Responses of Semi-Submersible Platform ", 7th International Conference on Steel & Aluminium Structures, Kuching, Sarawak, Malaysia July 13-15, 2011.
- [8] J.C. Lui, M.L. Sue, and C.H. Kou, " Estimating the Strength of Concrete Using Surface Rebound Value and Design Parameters of Concrete Material ", Tamkang Journal of Science and Engineering, Vol. 12, No.1, pp.1-7 (2009).