CHAPTER 1

INTRODUCTION

1.1 Background of Study

Corrosion under insulation (CUI) refers to the external corrosion of piping and vessels that occurs underneath the insulation as a result of the water ingress or moisture entrapped. Due to its nature, CUI tends to remain undetected until the insulation and cladding/jacketing is removed during inspection or when leak occur.

Therefore, the occurrence of CUI is seen as a major problem in refining, petrochemical, power industrial, onshore and offshore industries. There are a lot of cases that has been reported whereby CUI destroyed expensive industrial infrastructure and caused over timely-scheduled manufacturing plant and increase in facility downtime to conduct inspection. However, this problem can be reduced or eliminated by proper inspection control, proper installation and maintenance of insulation, or by proper selection, application, and maintenance of protective coatings [1].

General guidelines on proper CUI inspection is covered by American Petroleum Institute, API 581, the Base Resource Document Risk-Based Inspection program (RBI). This standard consists of guidelines to plan inspection for pressure vessels and piping system based on the risk calculated where risk is defined as the probability of failure and the consequences of the failure. Currently, the assessment of the probability of failure for CUI based on API 581 follows either the qualitative and semi-quantitative methods. The both methods are subject to many uncertainties and the results may not be optimized for inspection planning based on risk. Thus, predicting the probability of CUI using RBI seems to be inaccurate without quantitative data in which we will need another method to produce better assumption in determining the occurrence of it. Hence, this project aims to be one of the alternative methods by applying logistic regression model which can manipulate the qualitative data from inspection report to be some quantitative outcome for optimized inspection interval.

1.2 Problem Statement

The failure probability assessment of piping system subject to corrosion is based on measurable data, typically the wall thickness data, where data are statistically analyzed to establish the inspection intervals. However, the wall thickness data is not always available in the case of CUI, but the decision for planning an optimal inspection interval still need to be made in order to minimize the inspection and maintenance cost as well as to maintain the safety issue. At present, assessing the failure probability for CUI in RBI analysis is based on the American Petroleum Institute standard,(API 581) [1] where the likelihood analysis for CUI is predicted using the default corrosion rate , followed by a list of questionnaires without using the pipe wall thickness data, if available. This methodology is subjective and depends on the qualitative interpretation and judgment of the personnel involved resulting inaccurate failure probability prediction. A more objective methodology is needed to produce more accurate results.

Apart from that, historical data in CUI cases are so imperfect and most of the data are recorded without mentioning specific findings. General findings which are stated either CUI is observed or not observed. This type of data gives such a limitation as it cannot be used to predict the probability of failure using the conventional statistical techniques. Intuitively, the new approach which may serves in more quantitative way and also is able to cope with these existing circumstances is needed as to produce more accurate prediction for CUI inspection interval.

Therefore, in order to predict the appropriate inspection interval, a logistic regression method is proposed in this study since the model can be used for binary data (corrosion is observed and corrosion is not observed). Ideally, this research aims to point out the application of logistic regression model, a special method in linear regression in order to determine the probability of having CUI for RBI using CUI maintenance and inspection record. The response variables that characterize logistic regression as either binary go or not go (represented 0 or 1), are variables which do not lend themselves to analysis using traditional and conservative methods.

In addition to this project, there is a hypothesis made and need to be justified at the end of the project; states that the piping system which compromises of small bore and big bore will produce different probability of CUI occurrence. This hypothesis is important as this project will divide the piping system into two parts; small bore and big bore.

1.3 Significance of the Project

This project provides an alternative method in predicting the probability of CUI occurrence. It has been mentioned previously that the assessment of CUI is rather qualitative in RBI analysis. Therefore, the research work aims in establishing a more quantitative way in evaluating the failure probability in RBI by estimating the probability of CUI occurrence using a logistic regression model. This model will be used as a tool to predict the likelihood of having CUI given certain condition (i.e. operating temperature, year of service and etc.). Apart from that, this project can be a supplement for PETRONAS Risk-Based Inspection System (PRBI) that is currently applied in most PETRONAS subsidiary companies as a tool for planning inspection activities. CUI cases can be handled in a more effective and thus minimizing the maintenance budget.

1.4 Objectives and Scope of Study

The main objective of this study is to establish the probability of CUI failure using a logistic regression method. In order to achieve this objective, a few tasks and research work as indicated as the followings needs to be carried out:

- To develop a logistic regression model using MATLAB for small bore piping system and to validate the model.
- To develop a logistic regression model using MATLAB for big bore piping system and to validate the model
- To compare the results generated using the proposed model with the results generated from the standard API 581

CHAPTER 2

LITERATURE REVIEW AND THEORY

2.1 Corrosion under Insulation

Corrosion is defined as a chemical or electrochemical reaction between material, usually a metal, and its environment that produces a deterioration of the material and its properties [4]. In the most common usage of the word, this means a loss of an electron of metal reacting with either water or oxygen. There are several types of corrosion and one of them is corrosion under insulation (CUI).

CUI is one of major problems for refineries and petrochemical process industries. CUI is a severe problem because it results in staggering of maintenance costs for millions of dollars and can lead to the lost of production times as well as affecting plant integrity. Many chemical plants have experienced a variety of problems due to CUI [5]. For instances, according to PETRONAS Gas Berhad's financial year statement for year 2008/2009 [6], the CUI maintenance cost has achieved approximately RM5.6 million. The expenses which only covers for CUI maintenance process without including any other cost, for instance non-destructive testing, NDT (example of NDT are radiography test, ultrasonic etc.) has proved that CUI cases may increases the industry's expenditure.

CUI is a localized corrosion occurring at the interface of a metal surface and the insulation on that surface. This can be a severe form of corrosion particularly because the corrosion occurs beneath the insulation. The process starts when there is water being trapped in between the metal and the insulation. The confined environment of the insulation material over the pipe, tank or equipment creates conditions that encourage build up of moisture, resulting in corrosion. The corrosion is often more severe when the insulation restrict the evaporation process from occurring. In some cases the insulation acts as a carrier whereby moisture present in one area moves through the insulation to another area causing the corrosion to spread more rapidly as shown in Figure 1.

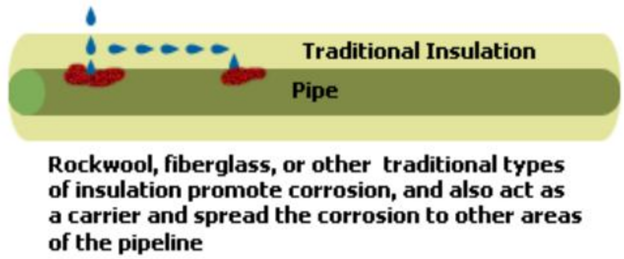


Figure 1.1: Mechanism of corrosion under insulation [2]

2.1.2 Factors that contribute to CUI

There are three major factors that are necessary for CUI to occur, namely, water, chemical content of water and operating temperature [7]:

Water

Water is the key point for corrosion to occur. Ordinarily, iron or steel corrode in the presence of both oxygen and water, and corrosion does not take place in the absence or either of these [8]. Water readily dissolves a small amount of oxygen from the atmosphere into solution and this may become highly corrosive. When the free oxygen dissolved in water is removed, the water is practically noncorrosive. If it is practically maintained neutral or slightly alkaline, it will be noncorrosive to steel [8]. However, most water that ingress inside which ingress inside insulation comes from rain which may contain chemical and acidic solution. In this case, normally, water can be introduced:

- During insulation storage and/or installation
- Through system leakages lack of maintenance and damage
- Due to heat transfer from warm to colder air

Chemical Content of Water

Carbon steel is not the only material that suffers from CUI due to water ‘‘containing’’ insulation. In many cases, austenitic stainless steel too cannot withstand water due to leaking insulation. Chlorides maybe introduced by rainwater or misty sea (or road salt) environments. Besides, traditional thermal insulation materials contain chlorides. If they are exposed to moisture, chlorides released may form a moisture layer on the pipeline surface, resulting in pitting/stress corrosion cracking. Therefore, the quality of the materials used for the insulation has to be controlled/specified in a way that these materials will not contain certain ‘‘acids’’ that can reduce the pH (pH below four (4) introduces acidic corrosion), this is especially important for carbon steel surfaces.

Operating Temperature

Operating temperature also contributes to CUI. In equipments or piping systems that operate in cyclic operating temperature; for example in regeneration process where the equipment operates at 300°C and during normal condition it operates at ambient temperature, it is most likely that CUI will be triggered. Here, the warm temperature normally results in more rapid evaporation of moisture and reduced corrosion rates. However, a surface that is covered with insulation will create an environment that holds in the moisture instead of allowing evaporation.

Besides, API 571 also specifies several critical criteria that contribute to CUI [9]:

1. Poor design and/or installations that allow water to become trapped will increase CUI.
2. Operating temperatures between the boiling point 100°C (212°F) and 121°C (250°F). CUI is particularly aggressive where operating temperatures cause frequent condensation and re-evaporation of atmospheric moisture

3. Equipment or piping systems that operate below the water dew-point tends to condense water on the metal surface thus providing a wet environment and increasing the risk of corrosion.
4. Plants located in areas with high annual rainfall or warmer, marine locations are more prone to CUI than plants located in cooler, drier, mid-continent locations.

2.2 State-of-Art to Failure Probability Assessment

Risk-Based Inspection (RBI)

Risk-based inspection (RBI) is an approach used for prioritizing inspection of static equipment based on risk. This approach estimates a risk associated with the operation of each equipment item based on a consistent methodology. RBI permits the shift of inspection and maintenance resources to provide a higher level of coverage on the high-risk items and an appropriate effort on lower risk equipment [1]. In this context, risk is defined as the product of the probability of failure and its consequence of failure whereby the methods to assess risk ranges from qualitative to quantitative manners. In the qualitative method, it is conducted based on the expert judgment with inert plant experiences. Meanwhile in the quantitative method, it is based on the statistical model developed typically using of the historical failure data.

In API 581, the quantitative assessment of the likelihood of failure begins with the database of generic failure frequency for onshore refining and chemical processing equipment. These generic frequencies are then modified by two terms, which are equipment modification factor (FE) and management modification factor (FM) to yield an adjusted failure frequency, as follow:

$$\text{Frequency}_{\,adjusted} = \text{Frequency}_{\,generic} \times \text{FE} \times \text{FM}$$

The equipment modification factor identifies specific conditions that can have a major influence on the failure frequency of that particular item. There are four sub-factors that lie under this element which are:

1. Technical module sub-factor (TMSF): TMSF examines materials of construction, the environment and the inspection program.
2. Universal sub-factor: conditions that affect all equipment items at the facility
3. Mechanical sub-factors: consideration
4. Process subfactor: influences that affect equipment integrity.

TMSF is the ratio of the frequency of the failure due to damage to the generic failure frequency times the likelihood of that damaged level to be present. For details to determine TMSF, refer **Appendix 3-3** for the quantitative analysis. TMSF requires the input of CUI corrosion rate. **Appendix 3-2** shows the flowchart to determine CUI corrosion rate for carbon steel material. The corrosion rate is determined given several factors such as the operating temperature, susceptible location (e.g. marine, potential cooling tower drift), types of coating or insulation being used, pipe support and others contributors. The generated corrosion rate is the estimated corrosion rate for such system.

The management modification factor (FM) covers mechanical integrity of facility or operating unit's management system that affect plant risk. The management evaluation system covers a wide range of topics whereby it involves questionnaires and interview for auditing process. The result from interview will be scored and evaluate to determine the level of risk, but again these techniques fall within quantitative methods which cause inaccurate prediction for any risk available [1].

Another way to analysis the failure probability quantitatively is through statistical analysis using failure data or degradation data. In country like Japan, they implements Risk Based Maintenance (RBM) for their fossil power plant as the solution in utility's requirement. Despite having RBM alone, the system is developed to perform the probabilistic risk analysis coupled with inspection system [10] in order to enhance the perfomance of analysis. Jovanovic (2003) has conducted a study on RBI program regarding risk-based inspection and maintenance in power and process plant in Europe. In his paper, he reviewed the current practices and trends using RBI by comparing European work and US work. He found out that making a successful RBI application is not an easy task. Thus, at the end of his paper report, he concluded that there are a lot of data, modeling, and software tools are needed, especially for the detailed quantitative analysis [3].

When the failure data is very limited, then the degradation data is used. For corrosion, the wall thickness data measured at the specified thickness measurement location (TMLs) can be used to assess the failure probability. However, in most cases, the data is also not available. Typically, the inspection data for CUI is inspected visually and this data can be translated as binary data (CUI found = 1 and CUI not found = 0). This data can be modeled using a logistic regression model.

2.3 State of Art-Estimation of likelihood

Generally, the likelihood probability can be estimated through several methods, depending on the condition of the problems. Historically, people may choose to apply other methods which are Poisson Regression and Fist Order Reliability Method (FORM) for this purpose. As an example, based on previous work, Madanat et. al (1995) introduced an ordered probit model for estimating the likelihood probability from inspection data whereby the model assumes the existence of an underlying continuous variable in the facility and infrastructure. Here, the model applied Poisson regression analysis in order to capture the infrastructure performance and deterioration [13]. However, applying Poisson regression to CUI may cause constraint since it seeks to model a count which means it counts for any deterioration point in order to determine the failure probability. In CUI, considering the nature of the process and subjective way it may possess, counting model is not applicable since it is not appropriate to open the whole insulation for counting process.

FORM may not also suitable to be used for CUI deterioration cases since this model requires wall thickness in order to predict the likelihood of failure. Normal external corrosion may be applicable for this model because the wall thickness can be easily found, but for CUI cases, the wall thickness is too limited.

2.4 Logistic Regression Framework

Regression methods have become an essential component of any data analysis when the relationship between a dependent variable and one or more independent variables are to be established. Prior to engaging in a study of logistic regression modeling, it is important to understand that the goal of an analysis using this method is the same as that of any model-building technique used in statistics, that is, to find the best fitting model [14].

Like a linear regression model, a logistic regression model describes the relationship between a dependent variable and a set of independent variables. The techniques employed in logistic regression analysis also follow the same general principles used in linear regression. However, the difference between these two models is that the dependent variable in logistic regression is binary or dichotomous and assumes a Bernoulli distribution.

In the scope of this study, typically what is available in inspection/maintenance data is the inspection result which stated whether corrosion was found and treated or corrosion was not seen. These types of data are classified as binary responses with 0 representing ‘‘corrosion was not seen’’ and 1 representing ‘‘corrosion was found’’, thus, can be modeled using a logistic regression to establish the relationship between the inspection result and one or more independent variables [15].

To explain the concept of logistic regression, the logistic function that describes the mathematics behind this regression should be defined clearly. The logistic regression function is [15]:

$$f(y) = \frac{1}{1+e^{-y}} \quad \text{where} \quad -\infty < y < +\infty \tag{1}$$

Where the y is model input and $f(y)$ is the model output, or the probability of occurrence for an event as illustrated in Figure 2.1. From the function, it can be observed that the logistic function can take an input having value from negative infinity to positive infinity whereas the output is confined to values between 0 and 1. The parameter y can be written as the linear sum of the independent variables as follows:

$$y(x) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} +_{....} + \beta_K X_{i,K} \tag{2}$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ are unknown regression coefficients and $X_1, X_2, \dots, X_k$ are independent variables.

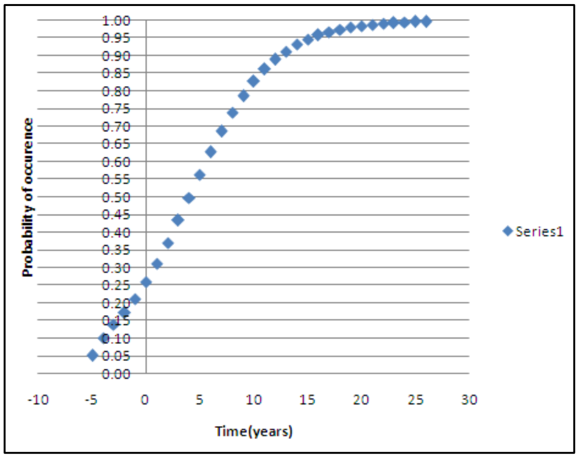


Figure 2.1: Graph logistic function

The intercept is the value of y when the value of all risk factors is zero. Each of the regression coefficients describes the size of the contribution of that risk factor. A positive regression coefficient means that that risk factor increases the probability of the outcome, while a negative regression coefficient means that risk factor decreases the probability of that outcome; a large regression coefficient means that the risk factor strongly influences the probability of that outcome; while a near-zero regression coefficient means that that risk factor has little influence on the probability of that outcome [13].

2.4.1 Distribution Assumption

Before the technique to estimate the model parameter is discussed, the assumption of the dependent variables following the Bernoulli distribution is explained. Binary data follow a Bernoulli probability mass function rather than a normal distribution because the output variable is binary and corresponds with the conditions listed below [16]. For observations on a categorical variable with two categories, the binomial distribution applies to the sum of the outcomes when the following three conditions hold true:

- For a fixed number of observation n, each falls into one of two categories.
- The probability of falling in each category, $(\omega 1)$ for the first category and $(1-\omega 1)$ for the second category, is the same for every observation.
- The outcomes of successive observations are independent; that is, the category that occurs for one observation does not depend on the outcomes of other observations.

2.4.2 Formulation Logistic Regression

2.4.2.1 Deriving the logistic regression formulation

Logistic regression is used to analyze the dependence of a binary response variable *y* on a set of *K* independent explanatory variables [35]:

$$log\left(\frac{P_i}{1-P_i}\right) = \log(odds) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + _{....} + \beta_K X_{i,K} \qquad (3)$$

Equation (1) may be equivalently rewritten to yield the predicted probability of occurrence satisfying the constraint which is $0 < P_i < 1$ [35]:

$$P_i\,(y_i = 1|X_i) = \frac{e^{\beta_0 + \beta . X_i}}{1 + e^{\beta_0 + \beta . X_i}} = \frac{1}{1 + e^{-\beta_0 + \beta . X_i}} \qquad (4)$$

P_i is the predicted probability of the occurrence (y_i=1) for the *i*th observation (*i*=1…N). 1-P_i is the probability of non-occurrence (y_i=0). β is a (K+1) column vector unknown parameters to be estimated including the intercept term. X_i is the (K+1) row vector of explanatory variables accounting for the *i*th observation. The explanatory variable can be continuous, categorical or both. The *odds* is defined as the ratio of the probability of occurrence over the probability of non-occurrence.

In this case, linear regression based on (1) cannot be used for the following reasons:

1. The response y_i is either 0 or 1 so the left hand size of (1) cannot be evaluated.
2. The response variable is a discrete binary data and it cannot be assumed to be normally distributed.
3. The predicted response may fall outside the (0-1) range, thus yielding meaningless result.

As logistic regression take place in Bernoulli distribution, there are two class of problem where P (y_i=1) = ω_1 and P (y_i=0) = 1- ω_1 may occur. These can be derived to yield equation (1) and (2) whereby equation (1) maybe transformed by taking the natural logs as follow [35]:

$$\ln\left(\frac{P(\omega 1|x)}{P(1-\omega 1|x)}\right) = \beta_0 + \beta.X_i \tag{5}$$

$$\ln\left(\frac{1-P(1-\omega 1|x)}{P(1-\omega 1|x)}\right) = \beta_0 + \beta.X_i \tag{6}$$

$$\frac{1-P(1-\omega 1|x)}{P(1-\omega 1|x)} = e^{\beta_0+\beta.X_i} \tag{7}$$

$$\frac{1}{P(1-\omega 1|x)} - 1 = e^{\beta_0+\beta.X_i} \tag{8}$$

$$\frac{1}{P(1-\omega 1|x)} = 1 + e^{\beta_0+\beta.X_i} \tag{9}$$

Rearrange the above equation gives:

$$P(x|\,\omega 1) = \frac{e^{\beta_0+\beta.X_i}}{1+e^{\beta_0+\beta.X_i}} = \frac{1}{1+e^{-\beta_0+\beta.X_i}} \tag{10}$$

The vector β is estimated by maximizing the likelihood function which will be explained further in likelihood function.

2.4.2.2 Likelihood function for Logistic regression

Since logistic regression predicts probabilities, it is necessary to consider the probability of CUI case was either P_i, if y_i=1 (corrosion was observed), or 1-P_i, if y_i=0 (corrosion was not observed). Then the likelihood function is then:

$$L(\beta, y) = \prod_{i=1}^{n} P_i^{y_i} (1 - P_i)^{(1-y_i)} \tag{11}$$

$$L(\beta, y) = \prod_{i=1}^{n} \frac{1}{1 + e^{-\beta_0 + \beta X_i}}^{y_i}_{i} (1 - \frac{1}{1 + e^{-\beta_0 + \beta X_i}})^{(1-y_i)} \tag{12}$$

Equation (#) maybe transformed, by taking the natural logs, to yield the following maximization problem:

$$f(\beta, y) = \sum_{i=1...N} \ln \left(\frac{1}{1 + e^{-\beta_0 + \beta X_i}}\right)^{y_i} + ln\left(1 - \frac{1}{1 + e^{-\beta_0 + \beta X_i}}\right)^{(1-y_i)} \tag{13}$$

Where *f(β)* is the log-likelihood function and maybe rewritten as:

$$f(\beta, y) = y'X.\beta - \sum_{i=1...N} \ln \left(1 + e^{\beta_0 + \beta X_i}\right) \tag{14}$$

Where X is (N) x (K+1) matrix corresponding to the N observations of the K-explanatory variables including a column vector of ones for the intercept. At this stage, it is need to find the maximum likelihood estimator for *β* and, here Newton-rahpson iterative method is applied.

2.4.2.3 Newton-Rahpson Iterative method

In this study, Newton –Raphson is like an iterative method, used to obtain parameter estimation for maximum likelihood. For this concept, theoretically, it will choose initial estimates of the regression coefficients, such as b_0=0. At each of iteration *t*, it will update the coefficient [20].

$$\beta_{i+1} = \beta_i - \left(\frac{\partial^2 l(\beta)}{\partial\beta\, \partial\beta^T}\right)^{-1} \frac{\partial l(\beta)}{\partial\beta} \tag{15}$$

Where the $\frac{\partial f(\beta)}{\partial(\beta)}$ denotes a (K+1) vector of partial derivatives of the function *f(β)* and is given by:

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1\ldots N} (y_i - \frac{e^{\beta_0+\beta.X_i}}{1+e^{\beta_0+\beta.X_i}}) .X_i \tag{16}$$

This can be rewritten in matrix form as:

$$\frac{\partial l(\beta)}{\partial \beta} = X_{i'}.(y - \frac{e^{\beta_0+\beta.X_i}}{1+e^{\beta_0+\beta.X_i}}) \tag{17}$$

Then, for second derivatives, $\frac{\partial^2 f(\beta)}{\partial(\beta)\partial(\beta')}$ denotes a (K+1) x (K+1) square symmetric matrix of second order derivatives and the function *f(β)* is given by:

$$\frac{\partial^2 f(\beta)}{\partial\beta\,\partial\beta'} = -\sum_{i=1\ldots N} - \frac{(1+e^{\beta_0+\beta.X_i}).e^{\beta_0+\beta.X_i} - e^{\beta_0+\beta.X_i}(e^{\beta_0+\beta.X_i})}{(1+e^{\beta_0+\beta.X_i})^2}. X_i'.X_i \tag{18}$$

The equation above can be rewritten in matrix form as:

$$\frac{\partial^2 f(\beta)}{\partial\beta\,\partial\beta'} = -X_i'.D.X_i \tag{19}$$

Where D is an (N) x (N) diagonal matrix:

$$D_{i,i} = \frac{(1+e^{\beta_0+\beta.X_i}).e^{\beta_0+\beta.X_i} - e^{\beta_0+\beta.X_i}(e^{\beta_0+\beta.X_i})}{(1+e^{\beta_0+\beta.X_i})^2} \tag{20}$$

The iteration procedure starts with an initial guess set to:

$$\beta_0 = (X'X)^{-1}.X'.y \tag{21}$$

This iteration will stop when the percentages of error is decrease to the smallest value which approximately becomes zero.

2.5 Two-tailed p-value

The p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. A *small* p-value indicates that the observed value of the test statistic lies far away from the hypothesized value of mean. In other words, it shows the *less* likely the result is if the null hypothesis is true, and consequently the *more* "significant" the result is, in the sense of statistical significance [33].

One often rejects a null hypothesis if the p-value is less than 0.05 or 0.01, corresponding to a 5% or 1% chance respectively of an outcome at least that extreme, given the null hypothesis. In order words, the p-value is the observed significance level of a hypothesis test made. When test uses two tailed p-value, using a significance level α=0.05, it is testing half of the alpha (α =0.025) in one direction and half of the other alpha to test statistical significance in other direction. Here, the test is conducted for the relationship in both directions.

As for one case where there is a need to compare the mean of a sample to a given value *x* using a t-test. The null hypothesis is that the mean is equal to *x*. A two-tailed test will test both if the mean is significantly greater than *x* and if the mean significantly less than *x*. The mean is considered significantly different from *x* if the test statistic is in the top 2.5% or bottom 2.5% of its probability distribution, resulting in a p-value less than 0.05.

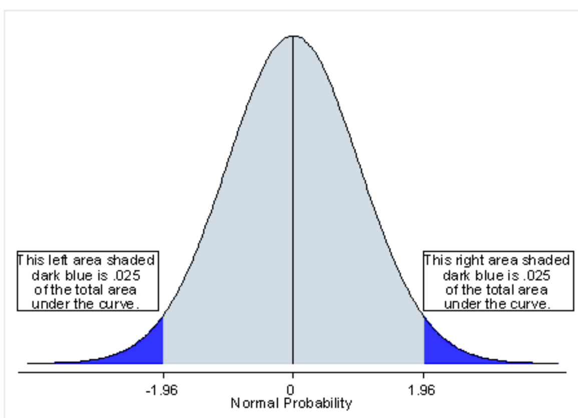


Figure 2.2: Two-tailed p-value [33]

CHAPTER 3

METHODOLOGY

3.1 Project Framework

The project work is focusing on development of logistic regression model using MATLAB software. Apart from that, data analysis which is based on the data collection from PETRONAS Gas Berhad (PGB), Plant Operation Division, Kerteh is divided into small bore and big bore group. Flowchart below describes details the process involved throughout this project.

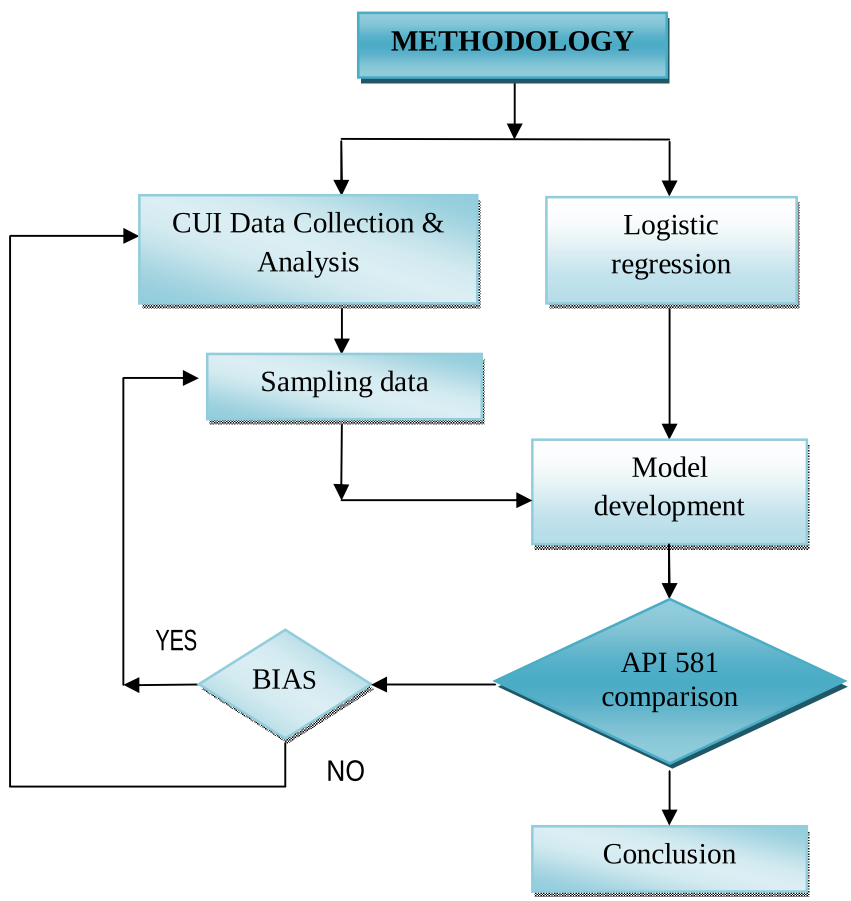


Figure 3.1: Flow chart for project work FYP2

3.1.1 Data Acquisition

The data acquisition phase included gathering both physical and condition information regarding CUI deterioration for piping system. There are 2 groups sample data taken which are from small-bore and big bore piping system. Small bore is for pipe size which has nominal pipe size (NPS) less that 2 inch while big bore has NPS more than 2 inch. From these samples, several factors which contribute to CUI have been considered. At the beginning of the research, there are 5 types of factors being considered as variables which are:

- Operating temperature
- Age (year of service),
- Process service
- Type of insulation process service
- Type of material.

However, since CUI is an external corrosion and the process start to accumulate on the top of pipe surface, therefore process service do not contribute and affect the mechanism of corrosion. The process service factor will be useful and significant for internal corrosion like erosion corrosion. Besides, the type of material is not a typical factor to be considered for CUI because it limits to carbon steel and stainless steel only. As this project is only focusing on carbon steel piping system, then the material type will be considered as constant for the whole sample data.

After the elimination of 2 factors, there are only three factors or explanatory variables (operating temperature, types of insulation and age) that contribute directly to the CUI deterioration and then are selected for the model development.

3.1.2 Variable Definitions

In this part, the definition for the variable will be explained details including using dummy variable as the input data into MATLAB.

Binary Response Variable

The term binary response refers to any variable having only two possible outcomes. In CUI deterioration cases, as previously mentioned, corrosion status is a binary response variable and may be classified as either corrosion observed or corrosion not observed.

Continuous variable

This continuous variable is one for which can take any value within the limits of variable ranges. In this model, age or year of service is considered under continuous variable (6, 10 and 15 years of services).

Covariates

A regression can simultaneously handle both quantitative and qualitative explanatory variables. In this case, the model combines elements of standard regression analysis, for which the predictors are quantitative, and analysis of variance where the predictors are qualitative. In the logistic regression model, the response variable is a binary variable; all explanatory variables are considered as covariate. For CUI deterioration cases, all explanatory variables consist of two types of variables: quantitative and qualitative. The quantitative variable such as operating temperature and the qualitative variable such as types of insulation are considered as categorical covariates.

However, due the weakness of categorical variable in regression model whereby it cannot be meaningfully interpreted, some other method of dealing with the information has been used. In order to cater this problem, process of transforming from a single categorical variable into dichotomous variable is used which is later known as dummy variable [30].

Dummy Variables

Dummy variables are artificial explanatory variables in a regression model whereby the dummy codes are a series of numbers assigned to indicate group membership in any *mutually exclusive and exhaustive* category. As mentioned previously in Data Acquisition, dummy variable need to be used in order to overcome the weakness of categorical variable as it cannot be meaningfully interpreted in regression model. In dummy variable, it will be dichotomous variable as each variable is assumed one of two values, 0 or 1, indicating whether an observation falls in a particular group.

For dummy variable to be used, if there are *K* level of purposes of the research (*K* groups), we need to have *K-1* dummy variables to represent *K* groups. Let say, if there are 6 temperature groups, we need to have 5 dummy variables to represent the group which one of the groups will not be represented as dummy variable. It will be considered as a reference to which each of the group should be compared.

3.1.3 Choosing a Reference Group of Dummy Variable

In dummy coding, one group is designated as the reference group and is assigned a value of 0 for every code variable. The choice of the reference group is statistically but not substantively arbitrary. Hardy (2003) suggested three practical considerations to guide the choice [32].

1. The reference group should serve as the useful comparison (eg. a control group; the group expected to score highest or lowest Y; on a standard treatment).
2. The clarification of interpretation of the results, whereby the reference group should be well defined and not as “waste-basket” category (eg: “other” for religion).
3. The reference group should not have a very small sample size relative to the other groups. This consideration enhances the likelihood of replication of individual effects in future research.

In this analysis, for operating temperature group, Operating temperature group 6, highest operating temperature (more than 121°C) is chosen as the reference. Meanwhile for type of insulation, Insulation Type 2 (cellular glass) is chosen as reference. When we stick to choose one group as the dummy variable, it will not affect the end result of the analysis if at the same time other statistician might choose another group to be represented as dummy. According to Hardy (2003);

> Regardless of which category is chosen as the reference group, the absolute difference in outcome will be the same [32].

To prove this statement, as a example for this case, we choose Insulation type 2 as reference, and it yield this linear function of independent variables (age, operating temperature and type of insulation);

$$y(x) = -4.1067 + 0.2365\beta 1 + 1.9212\beta 2 + 1.7926\beta 3 + 1.5733\beta 4 + 1.6528\beta 5 + 1.3735\beta 6 + 0.1274\beta 7 \quad (22)$$

Where β_0 = Intercept; β_1= age (year of service); $\beta 2$- $\beta 6$=dummy variable for operating temperature group; and $\beta 7$= dummy variable for Insulation Type 2.

For the case when a 2- inch pipe with operating temperature group 1, insulation type 1and already in service for 10 years, the estimated $y(x)$ = -2.0581+ 0.2365 (10) = 0.3069

In contrast, if we choose Insulation type 1 as reference, it will yield different intercept and coefficient for that particular beta ($\beta 7$) which is given as:

$$y(x) = -3.9793 + 0.2365\beta 1 + 1.9212\beta 2 + 1.7926\beta 3 + 1.5733\beta 4 + 1.6528\beta 5 + 1.3735\beta 6 - 0.1274\beta 7 \quad (23)$$

Where β_0 = Intercept; β_1= age (year of service); $\beta 2$- $\beta 6$=dummy variable for operating temperature group; and $\beta 7$= dummy variable for Insulation Type 1.

Using the same pipe with the same conditions in previous case, it will yield estimated linear function of independent variable, $y(x)$ = -2. 0581+ 0.2365(10) = 0.3069.

Here, we can see the end result for both equations will produce the same estimated $y(x)$.

3.2 Logistic Regression Development Model

In developing the logistic regression model, several steps have been employed as shown Figure 3.2.

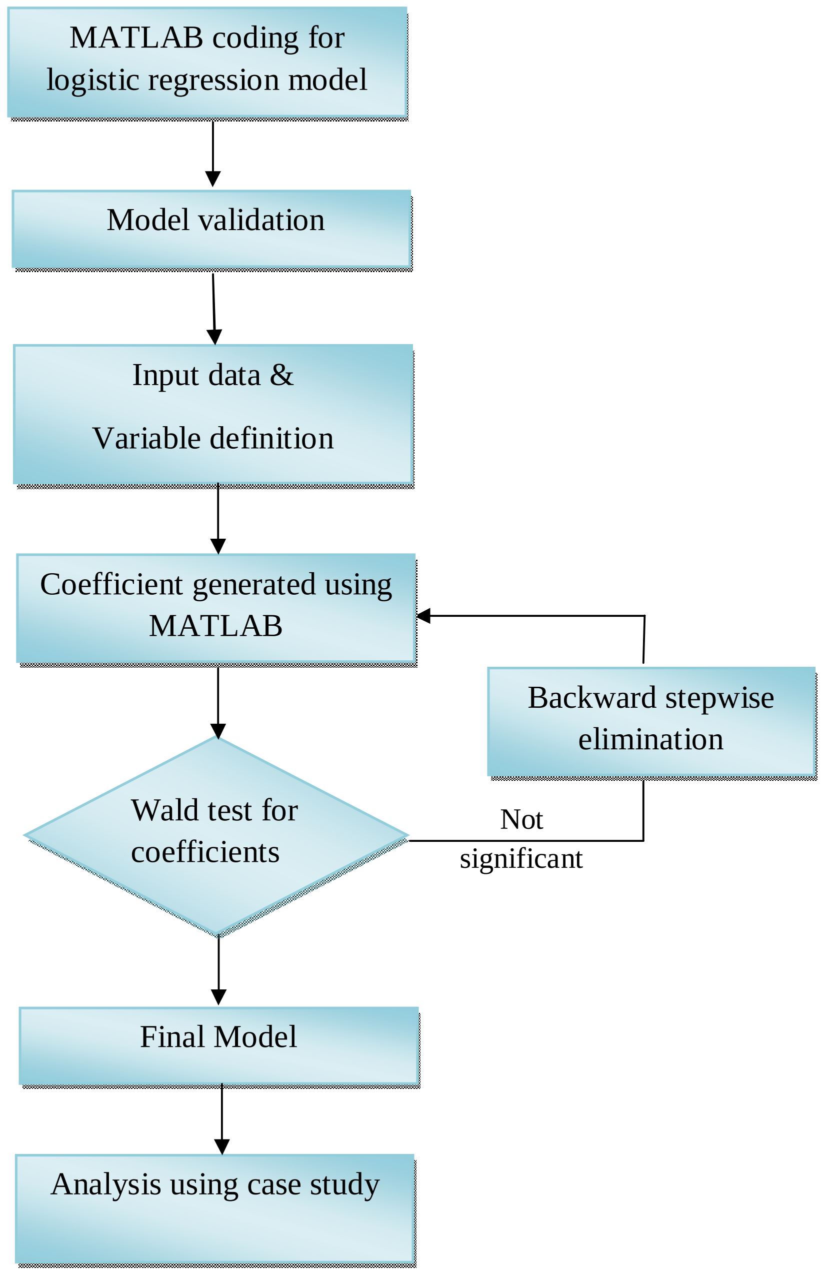


Figure 3.2: Formulation of logistic model development

3.2.1 MATLAB Functions

The logistic model is developed in MATLAB R2009a using two main functions; *glmfit* and *glmval* functions. In MATLAB, these functions are used in such a way [28]:

```
b = glmfit(x,y,distr)
```

The coding above returns as output of a *p*-by-1 vector b of coefficient estimates for a generalized linear regression of the responses in y on the predictors in **x**, using the distribution ***distr***. The distr can be any of the following strings, 'binomial', 'inverse Gaussian', 'normal' and 'Poisson'.

In this CUI case ***dist*** refer to binomial distribution. In most cases, ***y*** is an *n*-by-1 vector of observed responses whereby for this case, the response is either CUI is observed (1) or CUI is not observed (0). For the binomial distribution, ***y*** can be a binary vector (0 or 1) indicating success or failure at each observation, or a two column matrix with the first column indicating the number of successes for each observation and the second column indicating the number of trials for each observation [28].

In this project, the code above is used as:

```
b=glmfit(x,[y ones(339,1)],'binomial','link','logit');
```

Where:

- `'ones'` create array of all ones
- `'logit'`, default for the distribution `'binomial'` and represent $\log(\mu/(1-\mu)) = \beta x$
- `'link'` to link the anchor or pointer to the report

Next function used is `glmval'

```
fitted = glmval(b,x,'logit')
```

Where computes the predicted distribution parameters for observations with predictor values ***x*** using the coefficient vector β and link function '*link*'. Typically, $\boldsymbol{\beta}$ is a vector of coefficient estimates computed by the *glmfit* function. Here the code ‘link’ represents the logit function used in previous *glmfit* function whereby the value must be the same. Instead of producing coefficient estimates for the beta, the powerful software also provide several output using coding below:

```
[b,dev,stats] = glmfit(...)
```

It returns deviance (*dev*) which is also known as the generalization of the residual sum of squares. Apart from that, stats returns those several output which are listed below:

- `beta` — Coefficient estimates `b`
- `dfe` — Degrees of freedom for error
- `se` — Vector of standard errors of the coefficient estimates `b`
- `t` — *t* statistics for `b`
- `p` — *p*-values for `b`

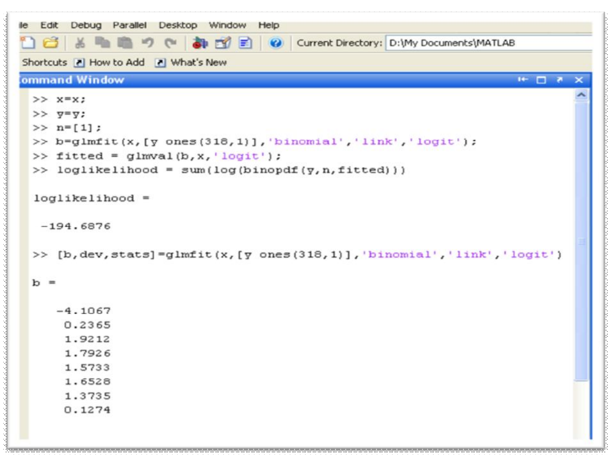


Figure 3.3: Command window for MATLAB coding for small bore piping system

For the whole MATLAB coding, please refer **Appendix 4-1.**

3.2.2 Model Validation

Once the final model is obtained, it is necessary to validate it with sample data from others. In this case, the validation model will be carried out using sample data based on the Evans Country data set. The Evans Country study is a cohort study of men, such that the number of individual at an exposure level without disease and the number with disease due to the simultaneous effect of catecholamine and cigarette smoking [36, 37]. Kevin Sullivan applied logistic regression model to find the probability of person might have coronary heart disease (CHD) due to these parameters.

The original result obtain by Kevin Sullivan using Java Script coding will be compare with result generated using MATLAB.

3.3 CUI Logistic Regression Model

3.3.1 Input data

All sample data are inserted into MATLAB based on the type of variable used. In the current model, operating temperature group 6 is referred as to as the base case and hence, 5 additional dummy variables are defined with respect to the base as follow:

- Op. temperature G1: 1 when temperature in range 49°C to 93°C, 0 otherwise;
- Op. temperature G2: 1 when temperature in range -12°C to 16°C, 0 otherwise;
- Op. temperature G3: 1 when temperature in range 16°C to 49°C, 0 otherwise;
- Op. temperature G4: 1 when temperature in range 93°C to 121°C, 0 otherwise;
- Op. temperature G5: 1 when temperature less than -12°C, 0 otherwise;
- Op. temperature G6: 1 when temperature is more than 121°C, 0 otherwise;

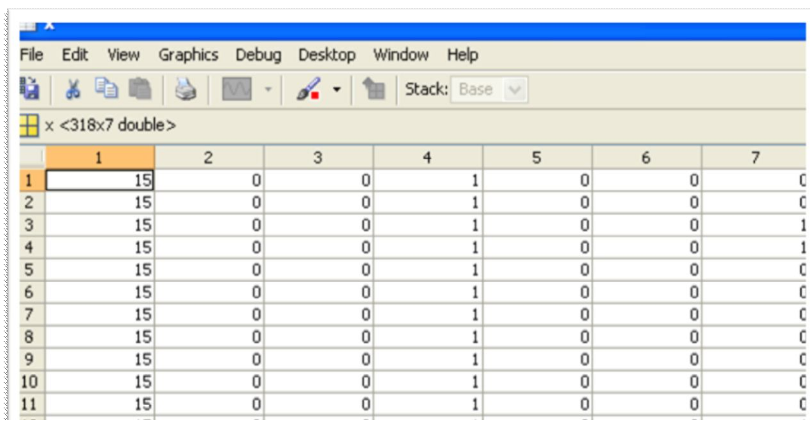
As for insulation type, there are two types of insulation involved. Considering insulation type 2 as the reference, there is only one dummy variable for insulation type as;

- Type 1 insulation: 1 when insulation used is calcium silicate, 0 otherwise;

In this analysis, age of pipes (i.e. year of service), is considered as the continuous variable. Table 1 shows a hypothetical sample of data which represents each type of the variables.

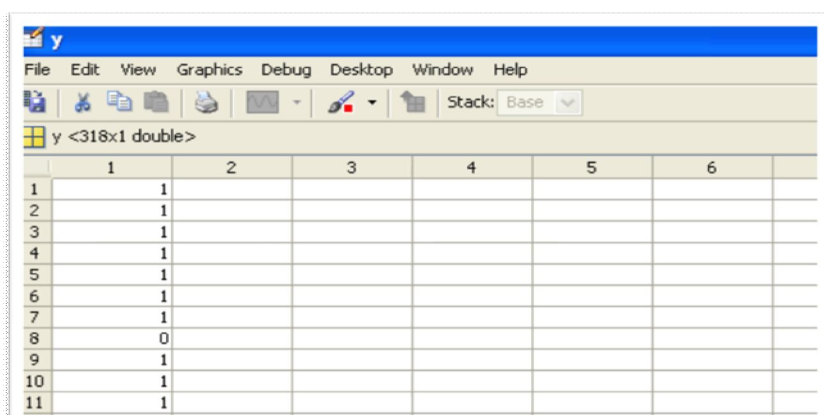
Table 3.1: Sample data for continuous, dummy and response variable for CUI cases.

Age	Op.Temp G1? (Yes=1)	Op.Temp G2? (Yes=1)	Op.Temp G3? (Yes=1)	Op.Temp G4? (Yes=1)	Op.Temp G5? (Yes=1)	Insulation G1? (YES=1)	Response
10	1	0	0	0	0	1	1
10	0	1	0	0	0	0	1
10	0	0	1	0	0	1	0
15	0	0	0	1	0	0	1



	1	2	3	4	5	6	7
1	15	0	0	1	0	0	0
2	15	0	0	1	0	0	0
3	15	0	0	1	0	0	1
4	15	0	0	1	0	0	1
5	15	0	0	1	0	0	0
6	15	0	0	1	0	0	0
7	15	0	0	1	0	0	0
8	15	0	0	1	0	0	0
9	15	0	0	1	0	0	0
10	15	0	0	1	0	0	0
11	15	0	0	1	0	0	0

Figure 3.4: Sample table for independent variables (*x*) in MATLAB



	1	2	3	4	5	6
1	1					
2	1					
3	1					
4	1					
5	1					
6	1					
7	1					
8	0					
9	1					
10	1					
11	1					

Figure 3.5: Sample table for response variable (*y*) in MATLAB

3.3.2 Result analysis

Once logistic regression has been developed, model selection method will be applied. Model selection is applied to analyze whether each of the variables used will be significant or insignificant to deterioration of CUI. Model selection consists of two steps:

- Examination of the significance of each parameter by performing Wald Test using standard error and p-value;
- Sensitivity analysis using Kruskal-Wallis methods.

3.3.2.1 Wald test, standard error and p-value

The Wald tests are based on chi-square statistics that tests the null hypothesis that a particular variable has no significant effect given that the other variables are included in the model. In the formulation of the logistic model, the Wald test was preformed first on each variable or model parameter to investigate its significance [27].

This Wald test is carried out by using the coefficient divided by its standard error whereby p-value is representing the probability of seeing a result as extreme in a collection of random data in which the variable have no effect.

3.3.2.3 Sensitivity analysis

The objective of the sensitivity analysis is to test the reliability of the model by evaluating its sensitivity to minor changes in the data set. To support this objective, and show that there is no difference among these three models from statistical point of view, the Kruskal- Wallis test was performed. The new logistic models are developed using 80% and 90% and 100% of the set of data.

Kruskal-Wallis test is used to evaluate differences between three or more treatment conditions (or populations) using data from an independent-measure design. It is used under the following hypotheses:

- H_0 : The models are equal (no significant difference between model)
- $H_{a:}$ The models are different.

In addition, Kruskal-Wallis test will be performed to test the two hypotheses. The test procedure is as follow:

1. To prepare the data for Kruskal-Wallis test, the complete set of original scores is rank-ordered using the standard procedure for ranking tied [ref]. Thus, combine all three samples into one large sample, before ranking the scores..
2. Find r_i, the sum of the ranks of the observation in the ith (i=1, 2, 3) sample.
3. Compute the test statistic Kruskal-Wallis using equation (24)

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^{3} \frac{r_i^2}{n_i} - 3(N-1) \tag{24}$$

   Where N= total sample size and ni= sample size for group *i*
4. Under H_0, Kruskal-Wallis follows an approximate chi-square distribution with k-1 degrees of freedom.
5. Reject the null hypothesis that all three models are the same if KW> $X^2_{\alpha,2}$.

3.3.3 Backward Stepwise Elimination

Backward elimination is an iterative variable-selection procedure where it begins with a model containing all the independent variables of interest. Then, at each step the variable with smallest F-statistic is deleted (if the F is not higher than the chosen cutoff level).

This step will be conducted if the coefficients obtained from result analysis are not significant. Here, the insignificant variable will be excluding from the input data for the next iteration in MATLAB.

3.4 Case study

Case study will be conducted based on likelihood analysis in Section 8 in order to evaluate frequency adjusted for 3 types of operating temperature.

$$Frequency_{adjusted} = Frequency_{generic} \times F_E \times F_M$$

Frequency $_{generic}$ is based on a compilation of available records of equipment failure histories. F_e which refers to the equipment modification factor identifies the specific conditions leads to major influence on the failure frequency. The last component given as F_M stands for management system evaluation factor.

CHAPTER 4

RESULT ANALYSIS

4.1 Model validation

Logistic regression model has been used previously for other cases like deterioration pavement crack, determine inspection interval for airplane and some other cases. A sample data is taken from Evans Country case study to validate logistic regression model [37]. Table below shows comparison between original coefficients generated via logistic regression model using Java Script codes and coefficients generated using MATLAB from this analysis.

Table 4.1: Comparison between original coefficient and coefficient using MATLAB

Parameters	**Original result from logistic regression model via Java Script by Kevin Sullivan [36]**	**Result from logistic regression via MATLAB**
Intercept	-2.9266	-2.9267
Variable 1 cateholamine category	1.3952	1.3953
Variable 2 smoking category	0.8653	0.8653
Variable 3 interaction category	-0.4498	-0.4498
Log-likelihood	417.8980	417.8980

From the table, it shown that logistic regression model developed in MATLAB may produce the same coefficients as original result by Kevin Sullivan. Thus, it is proved that the logistic regression model develop is acceptable.

4.2 Small bore

4.2.1 Backward stepwise elimination method

From backward stepwise method, the initial coefficients are generated from MATLAB as below:

Table 4.2: Initial coefficients generated from Matlab

Parameter	Coefficient	Standard error
Intercept	-4.1067	0.7491
Age (year of service)	0.2365	0.0410
Temperature group		
Op. Temp G1	1.9212	0.4529
Op. Temp G2	1.7926	0.5634
Op. Temp G3	1.5733	0.5314
Op. Temp G4	1.6528	0.4482
Op. Temp G5	1.3735	0.5284
Type of insulation		
Insulation G1	0.1274	0.3775

The next step is to proceed with the result from Wald test analysis for each of the coefficient.

4.2.2 Wald Test

After generating the coefficient for each variable, it is necessary to test the significant and effect carries by the coefficient. Thus, here Wald test is chosen as one of the appropriate test for this purpose. Wald test is obtained by comparing the maximum likelihood estimate of the slope parameter, to an estimate of its standard error. The resulting ratio obtained from the equation below is provided in Table 3 with two tailed p-value P (|z|>W) with α=0.05, where W denotes a random variable following the standard normal distribution.

$$W = \frac{\hat{\beta}_1}{S\hat{E}(\hat{\beta}_1)} \tag{25}$$

For example, taken the ratio between the coefficients for age (year of service) with its standard error gives:

$$W = \frac{0.2365}{0.0410} = 5.7683 \tag{26}$$

The resulting ratio, P (|z|>5.7683) gives for two-tailed p-value as less than 0.0001. Therefore, by conventional criteria, this difference is considered to be extremely statistically significant. Refer Table 3 for the whole answer.

Table 4.3: Coefficients generated from Matlab

Parameter	**Coefficient**	**Wald test**	**p-value**
Intercept	-4.1067	-5.4822	0.0000
Age (year of service)	0.2365	5.7683	0.0000
Temperature group			
Op. Temp G1	1.9212	4.2420	0.0000
Op. Temp G2	1.7926	3.1816	0.0015
Op. Temp G3	1.5733	2.9607	0.0031
Op. Temp G4	1.6528	3.6876	0.0000
Op. Temp G5	1.3735	2.5994	0.0093
Type of insulation			
Insulation G1	0.1274	0.3375	0.7357

Here, it is observed that each of the coefficients give significant result for p-value as it is lower than 0.05 except for insulation type 1 (calcium silicate) give higher p-value. As p-value is less than 0.05, it is generally accepted in statistical modeling when by having p-because there is only a 5% chance that results we are seeing would

have come up in a random distribution. Thus, we can say with a 95% probability of being correct that the variable is having some effect, assuming the model is specified correctly. In this analysis, operating temperature show significant effect towards the model but insulation type seems not significant and may be removed from final model.
In this analysis, we need to rerun the data and repeat the same steps by excluding insulation type from the model. The new coefficients are given as in the table below:

Table 4.4: Final coefficients generated from MATLAB

Parameter	Coefficient	Standard error	Wald Test	p-value
Intercept	-3.9804	0.6461	-6.1610	0.0000
Age (year of service)	0.2366	0.0410	5.7701	0.0000
Temperature group				
Op. Temp G1	1.8954	0.4458	4.2515	0.0000
Op. Temp G2	1.6749	0.4404	3.8030	0.0001
Op. Temp G3	1.4695	0.4317	3.4038	0.0007
Op. Temp G4	1.6457	0.4473	3.6793	0.0002
Op. Temp G5	1.2761	0.4433	2.8785	0.0040

From the table above, all p-values has shown significant value as it is lower than 0.05. Thus, we can write final a general equation of *y(x)* as linear function of independent variables as

$$y(x) = -3.9804 + 0.2366\beta 1 + 1.8954\beta 2 + 1.67496\beta 3 + 1.4695\beta 4 + 1.6457\beta 5 + 1.2761\beta 6 \quad (27)$$

Where β_0 = Intercept; β_1= age (year of service); $\beta 2$- $\beta 6$=dummy variable for operating temperature groups.

4.2.3 Simplified Equation

Equation (1) above can be further simplified according to the condition of the piping itself. Both will give different value for intercept thus produce the new equations. For example, a 2- inch pipe with operating temperature group 1, the new equation will become:

$$y(x) = -3.9804 + 0.2366\beta 1 + 1.8954(1) + 1.67496(0) + 1.4695(0) + 1.6457(0) + 1.2761(0) \tag{28}$$

Thus, simplified it to be:

$$y(x) = -3.9804 + 0.2366\beta 1 + 1.8954(1) \text{ or } y(x) = -2.085 + 0.2366\beta 1 \tag{29}$$

The new equation generated is the linear function of the age (year of service) with particular condition provided. Below are the list of equation for each group of operating temperature and type of insulation used:

Table 4.5: Simplified equation for operating temperature group

No	Type of operating temperature	Equation
1.	Operating temperature group 1	$y(x) = -2.085 + 0.2366\beta 1$
2.	Operating temperature group 2	$y(x) = -2.3054 + 0.2366\beta 1$
3.	Operating temperature group 3	$y(x) = -2.5109 + 0.2366\beta 1$
4.	Operating temperature group 4	$y(x) = -2.3347 + 0.2366\beta 1$
5	Operating temperature group 5	$y(x) = -2.7043 + 0.2366\beta 1$
6.	Operating temperature group 6	$y(x) = -3.9804 + 0.2366\beta 1$

From the linear equation generated above, functional forms for this relationship need to be described. Let p_i be the probability for CUI failure in case *i* and simple linear logistic regression model is

$$logit\,(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta o + \beta_1\, x \tag{30}$$

Where x is the age (year of service) and β_0 and β_1 are interception and slope respectively. The logistic regression model presents the log odds of CUI failure as a linear function of age with respect to dummy operating temperature and types of insulation used.

To predict the estimated probability of the CUI failure at certain year of service, rearranged the equation 4 into

$$\hat{p} = \frac{e^{\beta o+\beta_1 x}}{1+e^{\beta o+\beta_1 x}} \tag{31}$$

For example, taking same sample from previous in equation (3) above, the estimated probability of the pipe at 10 years of service (age factor) is:

$$\hat{p} = \frac{e^{-2.085+0.2561(10)}}{1+e^{-2.085+0.2566(10)}} = 0.570 \tag{32}$$

4.2.4 Sensitivity Analysis of Model

The logistic model is developed using 80%, 90% of the sample data and the results generated from these two models are subsequently compared to the original logistic model (100% sample data). The coefficient generated by 3 sample data is given in Table # while the result from Kruskal- Wallis test is given in Table 4.6.

Table 4.6: Coefficients for 100%, 90% and 90% of sample data

	100% of sample data		**90% of sample data**		**80% of sample data**	
Variable	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept	-3.9804	0.0000	-4.0380	0.0000	-4.0624	0.0000
Age	0.2366	0.0000	0.2334	0.0000	0.2419	0.0000
Temp. 1	1.8954	0.0000	2.1649	0.0000	1.9004	0.0002
Temp. 2	1.6749	0.0001	1.7502	0.0001	1.7586	0.0002
Temp. 3	1.4695	0.0007	1.6522	0.0003	1.3076	0.0046
Temp. 4	1.6457	0.0002	1.7561	0.0001	1.8386	0.0004
Temp. 5	1.2761	0.0040	1.3631	0.0036	1.0156	0.0405

The sensitivity analysis has revealed that KW= 0.089 compared with $X^2_{0.05} = 5.991$. Therefore the null hypothesis may be accepted, indicating that there is no significant different between the three models. The proposed model seems to be a good representation of the observed data.

Table 4.7: Kruskal-Wallis Test for 100%, 90% and 90% of sample data

Probability of occurrence			**Rank Measure**		
100% data	90% data	80% data	100% data	90% data	80% sample
0.136	0.163	0.128	2	4	1
0.166	0.197	0.157	5	7	3
0.202	0.236	0.192	8	10	6
0.243	0.281	0.232	11	13	9
0.289	0.330	0.278	14	16	12
0.340	0.384	0.329	17	18	15
0.394	0.440	0.385	20	21	19
0.452	0.499	0.444	23	24	22
0.511	0.557	0.504	26	27	25
0.570	0.613	0.564	29	30	28
0.627	0.667	0.622	32	33	31
0.680	0.717	0.677	35	36	34
0.729	0.762	0.728	38	39	37
0.773	0.801	0.773	41	42	40
0.812	0.836	0.813	43	45	44
0.846	0.865	0.847	46	48	47
0.874	0.890	0.875	49	51	50
0.898	0.911	0.900	52	54	53
0.918	0.928	0.919	55	57	56
0.934	0.942	0.936	58	60	59
0.947	0.954	0.949	61	63	62
0.958	0.963	0.959	64	66	65
0.966	0.971	0.968	67	69	68
0.973	0.977	0.975	70	72	71
0.979	0.981	0.980	73	75	74
0.983	0.985	0.984	76	78	77
0.987	0.988	0.987	79	81	80
0.989	0.991	0.990	82	84	83
0.992	0.993	0.992	85	87	86
0.993	0.994	0.994	88	90	89
		SUM:	1349	1400	1346
		KW: 0.089			

4.3 Big bore

4.3.1 Backward Stepwise Elimination Method

Big bore piping system undergone same procedure as small bore whereby a backward stepwise elimination method has been employed for selecting specific variables for the model development. Table below provide initial coefficients from MATLAB software.

Table 4.8: Initial coefficients generated from MATLAB for big bore

Parameter	Coefficient	Standard error
Intercept	-2.2791	0.1590
Age (year of service)	0.2616	0.0132
Temperature group		
Op. Temp G1	0.1809	0.1376
Op. Temp G2	0.1001	0.2032
Op. Temp G3	0.0949	0.1706
Op. Temp G4	0.3100	0.2188
Op. Temp G5	-0.0773	0.1806
Type of insulation		
Insulation G1	-1.0596	0.1470

All of the coefficients produced will be tested using Wald test with the same α=0.05 and will be referred to the two tailed p-value for the significant effect of each.

4.3.2 Wald Test

Table below provides the result analysis from Wald Test

Table 4.9: Coefficients generated from MATLAB

Parameter	Coefficient	Wald test	p-value
Intercept	-2.2791	-14.33	0.0000
Age (year of service)	0.2616	1.9012	0.0000
Temperature group			
Op. Temp G1	0.1809	0.8903	0.1887
Op. Temp G2	0.1001	0.4926	0.6223
Op. Temp G3	0.0949	0.5563	0.5783
Op. Temp G4	0.3100	-0.4280	0.1567
Op. Temp G5	-0.0773	-0.4313	0.6688
Type of insulation			
Insulation G1	-1.0596	-7.208	0.0000

Table 4.10 shows the variables that do meet and do not meet the 0.05 significance level criterion. Here, we can see that the operating temperature do not much contributes to the model development for big bore piping system as it turn to be insignificant factor. Further analysis will be carried out in the next part to prove this condition. On the other hand, insulation type shows significant effect based on the resulting p-value.

In this analysis, we need to rerun the data and repeat the same steps by excluding operating temperature type from the model. The new coefficients are given as in the table below:

Table 4.10: Final coefficients generated from MATLAB

Parameter	Coefficient	Standard error	Wald Test	p-value
Intercept	-2.1913	0.1590	-13.7844	0.0000
Age (year of service)	0.2614	0.0132	19.8282	0.0000
Insulation type 1	-1.0558	0.0946	-11.1580	0.0000

From the table above, all p-values has shown significant value as it is lower than 0.05. Thus, we can write final a general equation of *y(x)* as linear function of independent variables as

$$y(x) = -2.1913 + 0.2614\beta 1 - 1.0558\beta 2 \qquad (33)$$

Where β_0 = Intercept; β_1= Age (year of service); $\beta 2$= Insulation Type 1.

4.3.3 Simplified Equation

The new equation generated is the linear function of the age (year of service) with the type of insulation used. Below are the lists of equation for each group of type of insulation used:

Table 4.11: Simplified equation for insulation type group

No	Type of operating temperature with Insulation type	Equation
1.	Insulation type 1(calcium silicate)	$y(x) = -3.2471 + 0.2614\beta 1$
2.	Insulation type 2 (cellular glass)	$y(x) = -2.1913 + 0.2614\beta 1$

From the linear equation generated above, functional forms for this relationship need to be described. Let p_i be the probability for CUI failure in case i and simple linear logistic regression model is

$$logit\,(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta o + \beta_1\, x \tag{34}$$

Where x is the age (year of service) and β_0 and β_1 are interception and slope respectively. The logistic regression model presents the log odds of CUI failure as a linear function of age with respect to dummy operating temperature and types of insulation used.

To predict the estimated probability of the CUI failure at certain year of service, rearranged the Equation (35) into

$$\hat{p} = \frac{e^{\beta o+\beta_1 x}}{1+e^{\beta o+\beta_1 x}} \tag{35}$$

For example, taking same sample from previous in Equation (35) above, the estimated probability of the pipe at 10 years of service (age factor) is:

$$\hat{p} = \frac{e^{-3.2471+0.2614(10)}}{1+e^{-3.2471+0.2614(10)}} = 0.3468 \tag{36}$$

4.3.4 Sensitivity Analysis of Model

The logistic model is developed using 80%, 90% of the sample data and the results generated from these two models are subsequently compared to the original logistic model (100% sample data). The coefficient generated by 3 sample data is given in Table 12 while the result from Kruskal- Wallis test is given in Table 13;

Table 4.12: Coefficients for 100%, 90% and 90% of sample data

	100% of sample data		**90% of sample data**		**80% of sample data**	
Variable	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept		0.0000		0.0000		0.0000
Age		0.0000		0.0000		0.0000
Temp. 1		0.0000		0.0000		0.0002

Table 4.13: Kruskal-Wallis Test for 100%, 90% and 90% of sample data

Probability of occurrence			**Rank Measure**		
100% data	90% data	80% data	100% data	90% data	80% data
0.0374	0.0363	0.0337	3	2	1
0.0481	0.0467	0.0435	6	5	4
0.0616	0.0599	0.0560	9	8	7
0.0785	0.0766	0.0717	12	11	10
0.0996	0.0974	0.0915	15	14	13
0.1256	0.1231	0.1160	18	17	16
0.1573	0.1544	0.1461	21	20	19
0.1951	0.1920	0.1823	24	23	22
0.2394	0.2362	0.2252	27	26	25
0.2902	0.2869	0.2747	30	29	28
0.3468	0.3436	0.3305	33	32	31
0.4081	0.4051	0.3915	36	35	34
0.4725	0.4698	0.4561	39	38	37
0.5377	0.5355	0.5222	42	41	40
0.6017	0.6000	0.5875	45	44	43
0.6624	0.6612	0.6499	48	47	46
0.7181	0.7174	0.7075	51	50	49
0.7679	0.7676	0.7592	54	53	52
0.8112	0.8113	0.8043	56	57	55
0.8481	0.8483	0.8427	59	60	58
0.8788	0.8792	0.8747	62	63	61
0.9040	0.9045	0.9010	65	66	64
0.9244	0.9249	0.9222	68	69	67
0.9408	0.9413	0.9392	71	72	70
0.9538	0.9542	0.9527	74	75	73
0.9640	0.9644	0.9633	77	78	76
0.9721	0.9724	0.9716	80	81	79
0.9783	0.9787	0.9781	83	84	82
0.9832	0.9835	0.9831	86	87	85
		SUM:	1426	1470	1475
		KW: 0.0668			

The sensitivity analysis has revealed that KW= 0.0668 compared with $X^2_{0.05} = 5.991$. Therefore the null hypothesis may be accepted, indicating that there is no significant different between the three models. The proposed model seems to be a good representation of the observed data.

4.4 General comparison for small bore and big bore

In the previous analysis, small bore and big bore has shown several differences in terms of factors that are significant towards the probability of CUI occurrence. For small bore, operating temperature seems significant while for big bore, p-value for operating temperature shows not significant. On the other hand, for small bore, insulation types are insignificant factor but turn to be significant for big bore.

In small bore, when a graph of probability for CUI failure is plotted using time of service as base and six groups of operating temperature as variables, the trend produce is replicated from API guidelines. That means for operating temperature group 1 (49C to 93C) shows higher probability of having CUI compare with other temperature group.

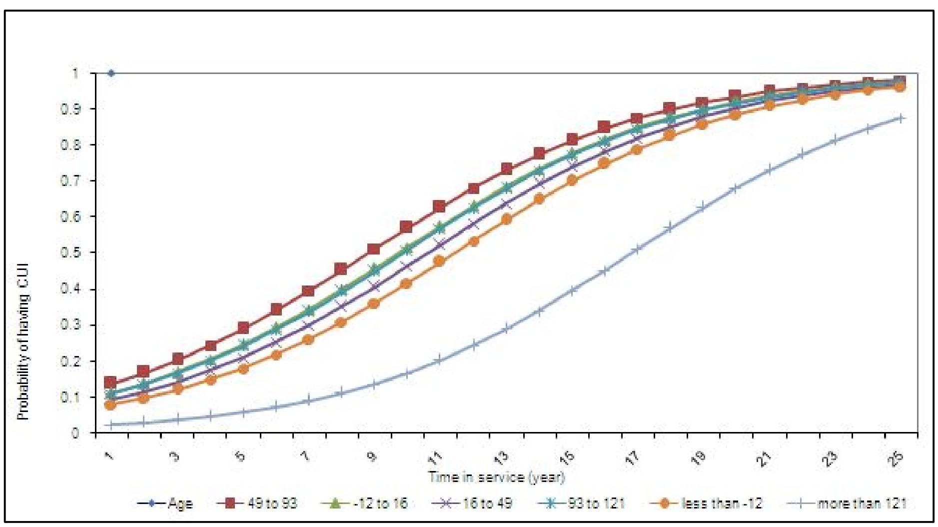


Figure 4.1: Graph probability of CUI occurrence for small bore

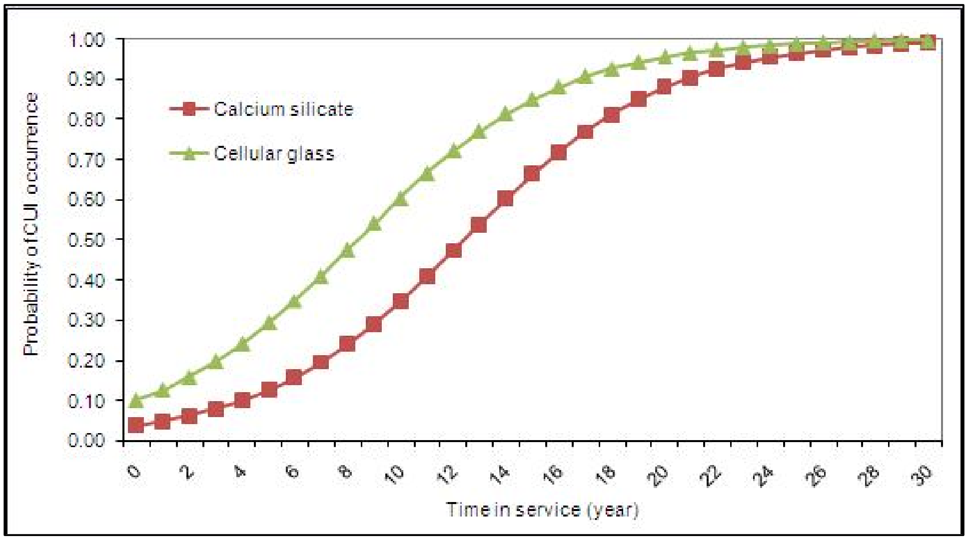
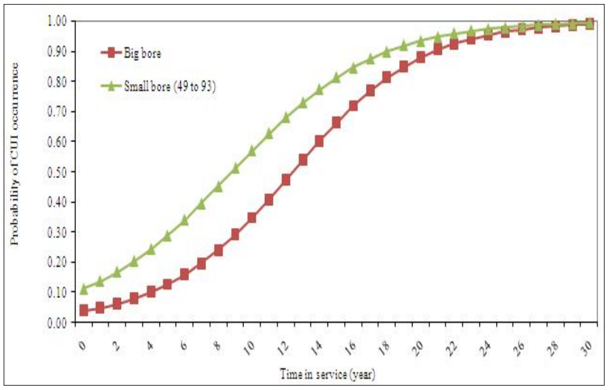


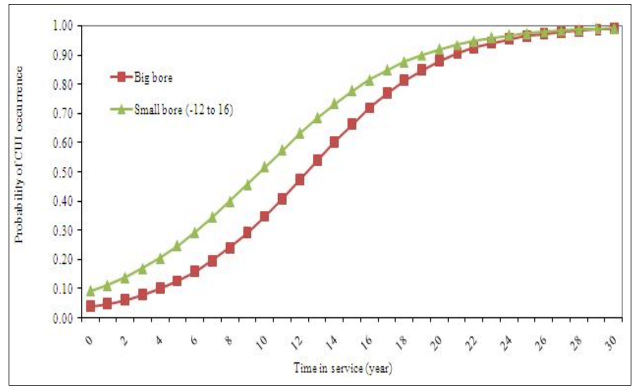
Figure 4.2: Graph probability of CUI occurrence for big bore piping

In this study, type of insulation becomes significant for big bore while operating temperature is insignificant. From the above graph, it is noted that cellular glass has higher probability of having CUI compare to calcium silicate. Further explanation regarding this situation will be discussed in the next chapter.

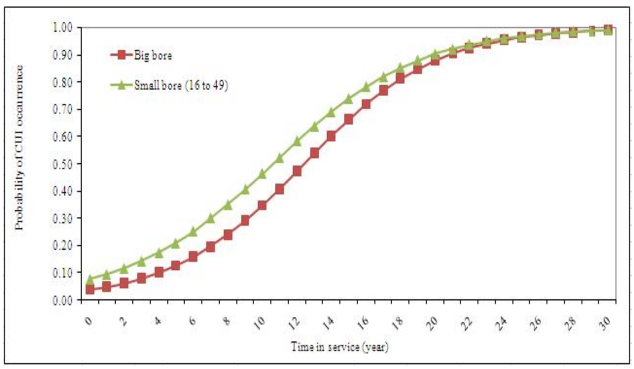
Above all, generally in normal condition, small bore will exhibit higher probability to have CUI compare to big bore. One of the factors is due to the significant effect from operating temperature towards the pipe itself. With regards to this scenario, a general comparison is done between small bore and big bore using six groups of temperature range as variables.



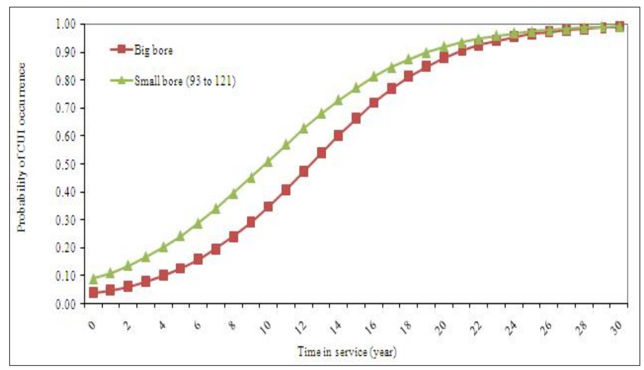
(a) Group 1: 49°C to 93°C



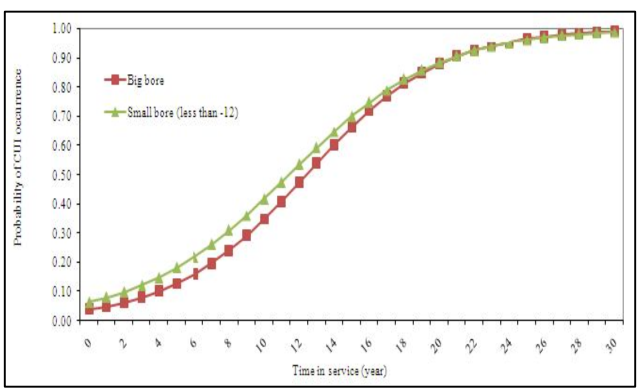
(b) Group 2: -12°C to 16°C



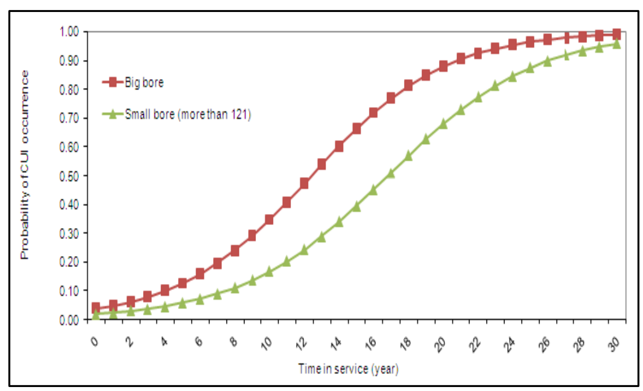
(c) Group 3: 16°C to 49°C



(d) Group 4: 93°C to 121°C



(e) Group 5: Less than -12°C



(f) Group 6: More than 121°C

Figure 4.3: Graph probability of CUI occurrence for small bore piping and big bore system

CHAPTER 5

DISCUSSION

5.2 Case study

From the analysis which has been conducted previously, the hypothesis of this project is proved as true when small bore and big bore group shown different end result for the final model (parameters involved and probability of CUI occurrence). However, it is insufficient to conclude in such a way without including analysis from API 581 for further understanding.

Here, the analysis using API 581 guidelines is an additional tool in order to verify the hypothesis made whereby small bore and big bore will produce different probability of failure given the same operating parameters and conditions. Besides, the idea behind this case study is to expose the semi- quantitative method applied in RBI analysis as it has tendency to be affected by the different interpretation from different people. Two types of temperature range are chosen for this case study:

- 16°C to 49°C (Operating temperature group 3 with 0.0508 mmpyr)
- 49°C to 93°C (Operating temperature Group 1 with highest corrosion rate, 0.254mmpyr)

The case study is calculated using technical module subfactor (TMSF) and semi-quantitative which included universal subfactor, mechanical subfactor, process safety management factor. For further reference, template for likelihood analysis is provided in the **Appendix 3-1.**

For this case study, two pipes are chosen from Gas Processing Plant B, GPPB, PETRONAS Gas Berhad, Kerteh. All of the pipes which come from two different range of operating temperature can be divided into small bore and big bore group, having same pipe schedule and same operating parameters for each lines. The lines are given as below:

1. Main line : **8"-P-5-06-203-31020-50HCS**

   Inspection point:

   - 18”-1-IG-116-L-416 (big bore)
   - 1.5”-1-IG-116-G-410 (small bore)

2. Main line : **36"-P-5-03-016-61010-34CCG**

   Inspection point:

   - 36"-1-IS -110-L 033 (big bore)
   - 3/4"-1-IS -110-V 034(small bore)

Further pipe specification for these lines can be referred in the **Appendix 3-7 and Appendix 3-8.** Below are the result for the case study using logistic regression model at instant time (current phase) and semi- quantitative likelihood analysis (without time base).

Table 5.1: Probability of failure for Operating Temperature G3

	Probability of failure for Operating Temperature G3 (49°C to 93°C)	
	Statistical approach	API 581 Risk-Based Inspection
Line type	Logistic regression model (at instant age= 0 year)	Semi-Quantitative Likelihood Analysis(without time based)
Small bore	0.076	0.0025
Big bore	0.0374	0.000075

Table 5.2: Probability of failure for Operating Temperature G1

	Probability of failure for Operating Temperature G1 (16°C than 49°C)	
	Statistical approach	API 581 Risk-Based Inspection
Line type	Logistic regression model (at instant, age= 0 year)	Semi-Quantitative Likelihood Analysis (without time based)
Small bore	0.110	0.0025
Big bore	0.100	0.000015

Based on the probability generated from the both tables, we can observed that in logistic regression model, small bore produces higher probability compare to the big bore. In Table 5.1 where the operating temperature lies under group3 (49°C to 93°C), the probability of having CUI in small bore pipe at instant condition is 0.076 while for big bore is 0.0374. On the other hand, for Table 5.2, the operating temperature is in group 1 (16°C to 49°C), again, small bore shows higher probability compare with big bore.

As for RBI analysis using semi-quantitative likelihood, it is observed the value for small bore is constant even in two different ranges of temperature. This is because; refer **Appendix 3-1**, in semi-quantitative analysis, operating temperature for that particular process is not considered directly as data input or factor contributes for CUI deterioration whereby it uses general temperature range for carbon steel material which is considered to be as greater than 288°C (-550°F). This range is too broad and it is not suitable for the CUI prediction. Intuitively, by having group 1 as the operating temperature, big bore in Table 16 supposed to give higher probability of failure for CUI compare to big bore in Table 15 but the result shows is contradict with the real situation.

In reality, small bore pipe will give higher tendency to experience CUI compare to big bore. As for this project, operating temperature is considered as one of the factor which can contribute directly to the CUI deterioration. At this point, temperature shows significant effect towards the small bore but it is relatively insignificant to the big bore. Consequently, we can see from the result whereby as the temperature increase, the probability of having CUI for small bore is higher compare to big bore. One of the reasons is due to the thickness of the pipe itself. Small bore will have lower thickness compare to big bore which in this case, if there is any external corrosion occurs on the pipe, the thickness can be affected easily. Thus, it will cause small bore to be more severed compare to big bore.

Another factor to be considered in this analysis is the insulation used for wrapping the pipe. From the analysis, insulation seems significant to big bore but it is insignificant for small bore. This is due to the size of the pipe itself. With bigger pipe size, it requires thick insulation compares with small bore. Then, due to its big size, it will cause side effect whereby the tendency for people to step onto it is higher. This phenomenon usually occurs in refinery plant as there are certain areas which are difficult to access. People may be randomly step onto big pipe in order to enter that area if it is the possible solution. Thus, by continuously stepping onto the pipe, it will affect the condition of the insulation itself. Hence, when the insulation is damaged or the sealant is loosed, water will easily ingress into it.

Then, among the type of insulation itself, cellular glass shows higher tendency for having CUI instead of calcium silicate. This is due to the properties of cellular glass as it impermeable to liquid, does not absorb moisture and it is hydro-barrier whereby at the same time it strengths the integrity of the barrier for the insulation function [34.] However, calcium silicate acts in different way as it has high physical water absorption function and good porosity. With these characteristics, both serves as advantages for insulation purposes as it can avoid water from being accumulated inside the insulation. Nonetheless, these advantages can counter-back its advantages when the condition of insulation is bad, damaged or broken. In that case, if the insulation material used is made from cellular glass, the water will accumulate onto the pipe surface as this material is not good in absorbing water, and thus it will leave the surface continuously wet. On the other hand, if the type of insulation used is calcium silicate that has high physical water absorption and avoiding heat losses for high temperature, it will reduce the amount of water by absorbing certain amount of it. As a result, the pipe surface will not be as wet as under cellular glass.

Despite of these properties of the material, this analysis has shown the significant of using logistic regression instead of semi-quantitative method in RBI. From the above tables, it is noted semi-quantitative analysis in RBI is generated without considering time as the base function. The likelihood is then predicted at the instant time based on general condition of the piping system which still subject to qualitative interpretation of personnel involved. Therefore, this analysis is unable to provide enough information regarding CUI deterioration for future action. Conversely, logistic regression serves in different modes whereby time-base function is considered as one of the significant factor in the model development. For both tables, each probability obtained is based on certain year of service (etc. at 0 or after pipe installation and so on), resulting in more systematic way of prediction. Using time as based function will offer advantage for inspection monitoring system since people can forecast CUI deterioration at any periods and provide more confidence to management for the integrity of piping system.

CHAPTER 6

CONCLUSIONS

Logistic regression model is successfully developed based on MATLAB software to determine probability of failure for CUI. POF for small bore pipes is found to be influenced by operating temperature while POF for big bore pipes is significantly affected by the types of insulation used. The general equation for logistic regression model in small bore is given as:

$$log\left(\frac{P_i}{1-P_i}\right) = \log(odds) = -3.9804 + 0.2366\beta1 + 1.8954\beta2 + 1.67496\beta3 + 1.4695\beta4 + 1.6457\beta5 + 1.2761\beta6 \quad (37)$$

Where β_0 = Intercept; β_1= age (year of service); $\beta2$- $\beta6$=dummy variable for operating temperature groups. While for big bore pipe, the general equation is given as:

$$log\left(\frac{P_i}{1-P_i}\right) = \log(odds) = -2.1913 + 0.2614\beta1 - 1.0558\beta2 \quad (38)$$

Where β_0 = Intercept; β_1= Age (year of service); $\beta2$= Insulation Type 1.

Analyzing the case study via logistic regression model reveals that the prediction of POF for CUI can be done in more accurate way with time as a based function. In other words, logistic regression is not only providing POF for CUI but also the prediction for future inspection monitoring plan. However, in API 581 the POF is given instantaneously without considering year of service as one of the significant parameter. That means, API 581 is unable to provide enough information for future prediction.

CHAPTER 7

RECOMMENDATION

Even though logistic regression model is able to give better prediction compare with RBI, API 581 but this model still need to be improved in order to improve its reliability and accuracy for CUI.

7.1 Add more information in inspection report

In this study, the parameters used are limited to the age (year of service), operating temperature and type of insulation. It is advisable to have more parameters which can lead to the more accurate result and prediction. Therefore, it is recommended that in the next inspection procedure for CUI, CUI inspection report need to be more details instead of having typically binary data whether CUI is observed or not observed.

7.2 Data distribution should be mentioned clearly.

For future work, the CUI data distribution should be determined whether it is tabulated normally or binomial distribution. This is important step which can help to increase the accuracy of the project and avoid error in statistical modeling.

REFERENCES

[1] *Risk base inspection resource document.* (2000). American Petroleum Institute standard API 581.

[2] N. Dowling, M. Lorenz & F. Ropital. (2004). *Corrosion in the refinery industry.* San Donato Milanese Italy: Eni Technologie Research Centre.

[3] A.Jovanovic. (2003). Risk-based inspection and maintenance in power and process plants in Europe. *Science Direct* , 166-181.

[4] R. J. Davis. (2000). *Corrosion understanding the basic.* United State of America: ASM International.

[5] W. I. Pollock, J. M. Barnhart. (1985). *Corrosion of metal under thermal insulation.* United States: American Society for Testing and Material.

[6] PETRONAS Gas Berhad's financial year statement (FY08/09). Inspection Department. Kerteh, Terengganu

[7] F.D.Vogelaere. (2008). Corrosion under insulation. *Wiley Science* , *28* (1), 6.

[8] Philip A. Schweitzer, P.E. (1989). *Corrosion protection handbook.* New York: Marcel Dekker, Inc.

[9] *Damage mechanisms affecting fixed Equipment in the refining industry.* (2003). American Petroleum Institute (API) 571.

[10] K. Fujiyama, S. Nagai, Y. Akikuni, T. Fujiwara, K. Furuya, S. Matsumoto, K. Takagi & T. Kawabata. (2004). *Risk-basd inspection and maintenance systems for steam turbines.* Yokohama, Japan: Science Direct.

[11] Vinod, G., Bidhar, S. K., Kushwaha, H. S., Verma, A. K. & Srividya, A. (2003). A comprehensive framework for evaluation of piping reliability due to erosion-corrosion for risk-informed in-service inspection, *Reliability Engineering and System Safety*, 82, 187-193.

[12] Fleming, K. N. (2004). Markov models for evaluating risk-informed in-service inspection strategies for nuclear power plant piping systems, *Reliability Engineering and System Safety*, 83, 27-45.

[13] S.Madanat & W. Hashim. (1995). Poisson regression model of infratsurcture transition probabilities. *Journal of Transportation Engineering* , 268-272.

[14] Ariaratnam, S.T.,El-Assly,A. and Yang,Y. (2001). Assessment of infrastructure inspection needs using logistic models. *Journal of Infrastructure Systems , 7* (4), 160-165.

[15] K.E Spezzaferro. (1996). Applying logistic regression to maintenance data to extablish inspection interval. *Preoceeding of the American Society of Mechanical Engineers Symposium on Structural Intgerity in Aging Air-carft* , 296.

[16] Wikipedia.com

[17] Agresti, A. (1990). Categorical data analysis. New York.

[18] I.J Myung. (2002). Tutorial on maximum likelihood estimation. *Science Direct* , doi: 10.1016/S0022-2496(02)00028-7.

[19] P. Meagher. (2004). Implement bayesian inference using PHP: Part 2: Solving parameter estimation problems. *Developer Works* .

[20] J. Fox. (2004). Maximum likelihood estimation of the logistic regression model. *UCLA/CCPR* .

[21] .J.Azman, C. &. (2008). *CUI Inpsection Report*. Terengganu: Petronas Gas Berhad.

[22] Hanapizan. (2009). *Piping CUI near A2-401*. Kerteh, Terengganu: Inspection Deparment, GPPA

[23] *The Trustees of Princeton University*. (2007). Retrieved February Wednesday, 2010, from DSS-Interpreting Regression Output.

[24] *API 570 Piping Inspection Code, Inspection, Repair, Alteration, and Rerating* (Vol. 2). (1998). Washington DC: American Petroleum Institute.

[25] Ming-Kuen Chang, R.-R. C.-M.-N. (2005). Application of risk based inspection in refinery and processing piping. *Science Direct* , 1.

[26] (1994). *Engineering Specification for Thermal Insulation Design, Dew Point Control Unit, PETRONAS Gas Sdn. Bhd.* Consortiom of Toyo Engineering Corporation.

[27] Jidong Yang, M. G. (2005). Use of Recurrent Markov Chains for Modelling the Crack Performance of Flexible Pavements. *Journal of Transportation Engineering ASCE* , 861-872.

[28] *The MathWorks*. (1994-2010 The MathWorks, Inc) Retrieved 2010, from Statistics Toolbox 7.2 Fitting Data with Generalized Linear Models: http://www.mathworks.com/products/statistics/demos.html?file=/products/demos/shipping/stats/glmdemo.html

[29] P.Roxy, O. C. (2009). *Introduction to statistic and data analysis.* Canada: Brooks/Cole Cengage Learning.

[30] Stockburger, D. W. (n.d.). *Multiple Regression with Categorical Variables.* Retrieved 2006, from The Psychology Department at Missouri State University: http://www.psychstat.missouristate.edu/

[31] Ming-Kuen Changa, R.-R. C.-M.-N. (2005). Application of risk based inspection in refinery and processing piping. *Science Direct* , 397-402.

[32] Melissa A. Hardy. (1993). *Regression with dummy variables.* California: Sage Publication Inc.

[33] Bruin, J. 2006. newtest: command to compute new test. UCLA: Academic Technology Services, Statistical Consulting Group. http://www.ats.ucla.edu/stat/stata/ado/analysis/.

[34] Silowash, B. (2010). *Piping System Manual.* United States of America: McGraw Hill.

[35] Tchantchane, A. (2009). Categorical data modelling: Logistic regression software. *World Academy of Science, Engineering and Technology* , 230- 235.

[36] Sullivan K & Pezzullo, J.C (2007) Logistic Regression version; version 05.07.02. Retreived April 12,2010, from http://statpages.org/logistic.html

[37] Kleinbaum, Kupper, and Morgenstern (1982), *Epidemiologic Research: Principles and Quantitative Methods*, New York: Van Nostrand Reinhold.

APPENDIX 1-1

Gantt chart FYP 2

Table 6: Gantt chart for final year project 2

	1	2	3	4	5	6	7		8	9	10	11	12	13	14
Data Analysis PART 2								M							
Develop model								I							
Find equation coefficient								D							
Validate data and model testing															
Preparation for submission of Progress Report 1								S							
Submission of Progress Report 1				☻				E							
Preparation for submission of Progress Report 2								M							
Submission of Progress Report 2								E	☻						
Preparation for Seminar								S							
Seminar 2								T	☻						
Preparation for poster								E							
Poster Exhibition								R				☻			
Preparation for Dissertation Final Draft															
Preparation for Oral															
Submission of Dissertation (Hard bound)															☻

LEGEND:

Completed

In progress

Preparation

Milestone

APPENDIX 2-1

Specification for operating temperature and insulation types

Table A.2.1: Classification for operating temperature

Temperature range (°C)	Corrosion rate(mmpy)	Temperature group
49°C to 93°C	0.254	1
-12°C to 16°C	0.127	2
16°C to 49°C	0.0508	3
93°C to 121°C	0.0508	4
Less than -12°C	0	5
More than 121°C	0	6

This classification made based from corrosion rate which is adapted in API 581, Risk Based-Inspection guideline. Since data collection is made at marine location, thus the corrosion rate taken is mainly referring for marine while eliminating temperate and arid location.

Table A.2.2: Classification for types of piping

Insulation class	Insulation requirement	Temperature range (°C)	Insulation material	Insulation group
N10A	Heat Input Control	0 to 650	Calcium silicate	1
N20A	Heat Conservation	10 to 650	Calcium silicate	1
N23A	Personnel Protection	65 to 650	Calcium silicate	1
HCS	Heat Conservation	10 to 650	Calcium silicate	1
N31A	Cold conservation	+10 to -180	Cellular glass	2
N34A	Prevention of Surface Condensation	10 to ambient	Cellular glass	2
CCG	Cold conservation	+10 to -180	Cellular glass	2

APPENDIX 3-1

Quantitative likelihood analysis

Objective: This worksheet is used to calculate the likelihood of failure [1]

	Enter Data
	Auto Computed

General Info	
Piping type	Piping, 0.75 in. diameter, per ft
Leak size	1/4 in.
Leak frequency	1.0E-05
Pipe length (in m)	10

Technical module subfactor (TMSF): Ratio of the frequency of failure due to damage to the generic failure frequency times the likelihood that the damage level is present.

Key-in TMSF

Technical Module Subfactor				2
Damage State	Probability of failure	"Generic" probability of failure	Ratio to "Generic"	Confidence in predicted damage rate (before inspection)
1		1.00E-05	0.0E+00	Moderate reliability data
2		1.00E-05	0.0E+00	Low reliability data
3		1.00E-05	0.0E+00	Low reliability data
Total technical module subfactor (TMSF)				

Universal subfactor

Plant condition		About equal to industry standards	0
Cold weather		Above 5°C (40°F)	0
Seismic activity		0 or 1	0
Mechanical subfactor			
Piping complexity	Number of connections	1	10
	Number of injections points	1	20
	Number of branches	1	3
	Number of valves	1	5
	Complexity factor	38	
	Complexity factor per foot	1.16	0
Construction code		The code for this type of equipment has been significantly modifies since the time of fabrication	1

Life cycle	Years in service	15	
	Design life	25	
	% of design life elapsed	0.6	
	Life cycle value	0 to 7	2
Safety factors	Operating pressure, $P_{operating}$	70	
	Design pressure, P_{design}	81.7	

	$P_{operating}/P_{design}$	0.86	
	Operating pressure value	0.7 to 0.89	0
	Operating temperature, $T_{operating}$	For carbon steels: > 288°C (-550°F)	2
Vibration monitoring	For pumps	No vibration monitoring program	0.5
	For compressors	No vibration monitoring program	1

Process subfactor				
Continuity	Planned Shutdowns (per year)	0 to 1	-1	
	Unplanned Shutdowns (per year)	0 to 1	-1.5	
Stability		Process has about average stability	0	
Relief valves	Maintenance program	Less than 5% of RVs overdue	-1	
	Fouling service	No significant amount of fouling	0	
	Corrosive service	Yes	3	
	Very clean service	Yes	-1	
Equipment modification factor				7
Process safety management modification factor				50
Adjusted failure frequency				0.0035

APPENDIX 3-2

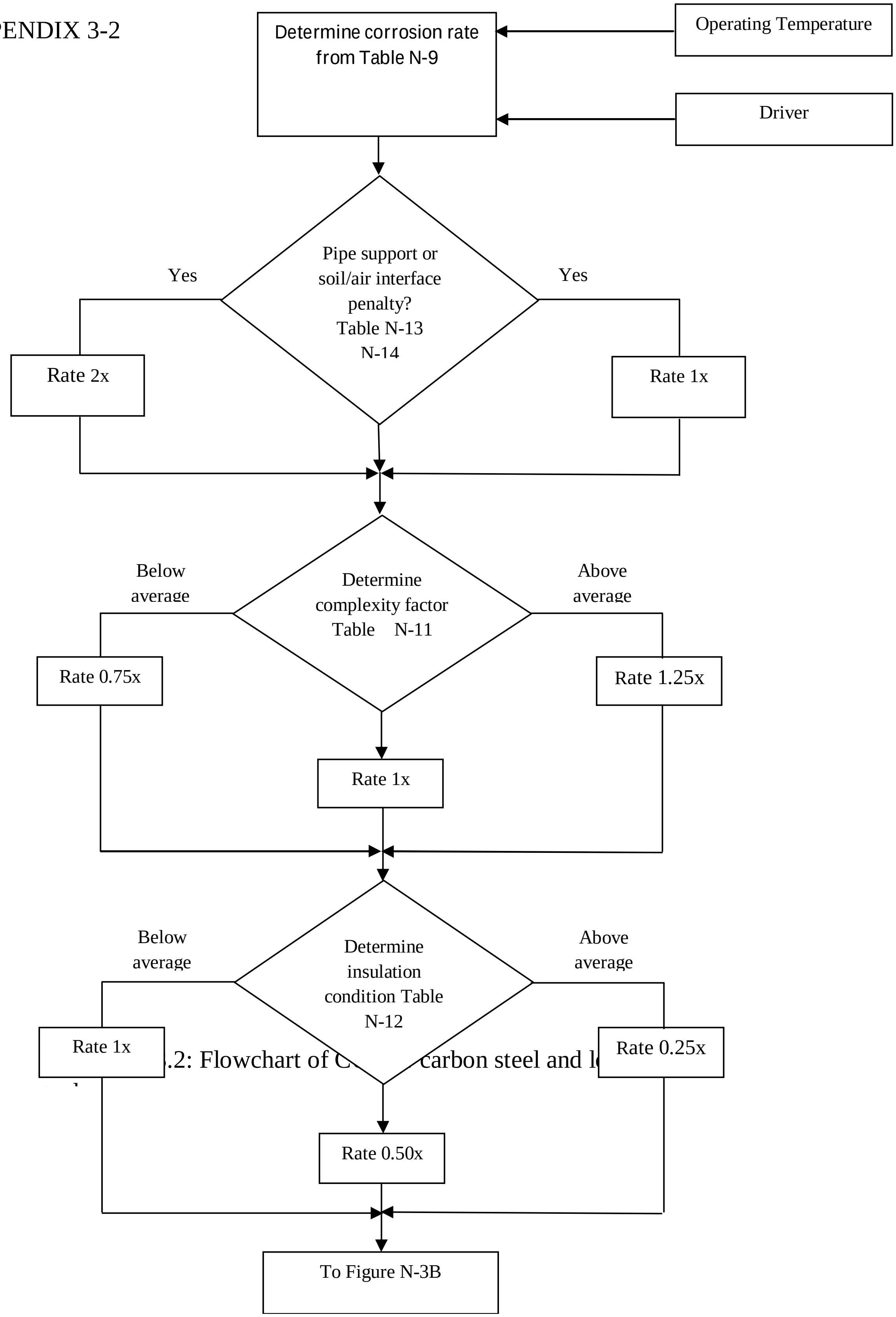
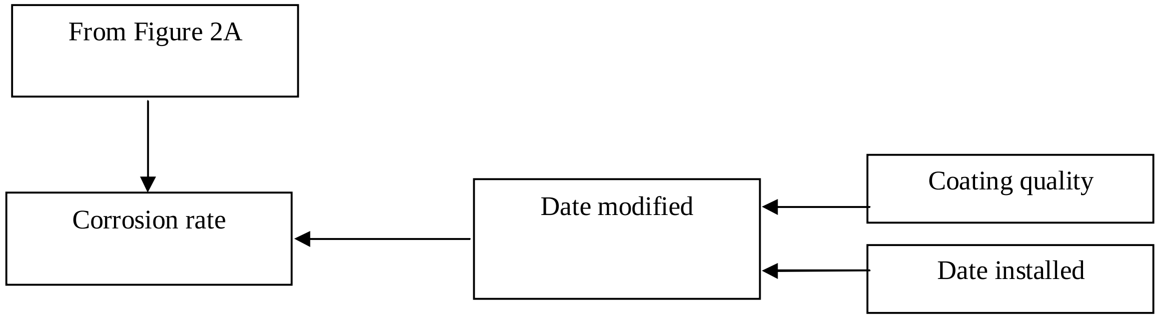


Figure A.3.2: Flow chart of CUI for Carbon Steel [1]


From Figure 2A
Corrosion rate
Date modified
Coating quality
Date installed

APPENDIX 3-3

Table A3.3: Basic data required for CUI for Carbon Steels [2]

Parameter	**Explanations**
Driver	The drivers for external corrosion under insulation. This can be weather at a location.
Rate, in mmpy	Corrosion rate for external corrosion. Based on temperature and driver.
Date	Determine the time (in years). Default to date installed. Can change based on date of coating, time since last inspection.
Inspection Effectiveness	The effectiveness of the CUI inspection program.
Inspection Number	The number of CUI inspection.
Coating Quality	Related to the type of coating applied under insulation: • None- no coating or primer only. • Medium- single coat epoxy. • High- multi coat epoxy or filled epoxy.
Complexity	The number of branches: • Below average • Average • Above average
Insulation Condition	Determine whether insulation condition is good based on external visual inspection. Good insulation will show no sign of damage (i.e. punctured, torn or missing water proofing) or standing water (i.e. brown, green or black stains).
Pipe Support Penalty (Y/N)	If piping is supported directly on beams or other such configuration that does not allow for coating maintenance, external corrosion can be more severe.
Interface Penalty (Y/N)	If piping has interface where it enters either soil or water, this area is subject to increased corrosion.

APPENDIX 3-4

1. Adjustment for Coatings

Coating Quality		
None	Medium	High
Date = Installed	Date = Coating date + 5	Date = Coating date + 15

2. Adjustment for Complexity

Below Average	Average	Above Average
Rate= Rate x 0.75	Rate= Rate x 1	Rate= Rate x 1.25

3. Adjustment for Insulation Conditions

Below Average	Average	Above Average
Rate= Rate x 1	Rate= Rate x 0.5	Rate= Rate x 0.25

4. Adjustment for Pipe Support Penalty

Penalty Apply	Penalty Does Not Apply
Rate= Rate x 2.0	Rate= Rate x 1.0

5. Adjustment for Interface Penalty

Penalty Apply	Penalty Does Not Apply
Rate= Rate x 2.0	Rate= Rate x 1.0

APPENDIX 3-5

Data specification for case study 1 for TMSF

Table A.3.5: Data collection and specification for **36"-P-5-03-016-61010-34CCG**

Parameter	Small bore 3/4"-1-IS -110-V 034	Rating	Big bore 36"-1-IS -110-L 033	Rating
Operating temperature (Marine)	27°C	0.0508	27°C	0.0508
Pipe support or penalty interface	Yes	2	Yes	2
Complexity factor	Default	1	Default	1
Equipment Insulation condition	Below average	1	Average	0.5
Coating quality	None	1.25	Medium	1
Calculated corrosion rate	0.0508x2x1x1 x1.25=0.127mmpyr ≈ 0.005in./yr	Below than state 1	0.0508x2x1x0.5 x1.25=0.0635mmpyr ≈ 0.0025in/yr	Below than state 1

Table A.3.5.1: Calculate the Composite Technical Module Subfactor for all Damage Mechanisms [1]

Table 8-8—Calculated Frequency of Failure for Different Damage States

Damage State	Corrosion Rate	Wall Loss	Remaining Wall	Frequency of Failure
1	0.010 in./yr	0.06	0.315	8×10^{-6}
2	0.020 in./yr	0.12	0.255	2×10^{-5}
3	0.040 in./yr	0.24	0.135	5×10^{-3}

Table 8-9—Calculated Technical Module Subfactor

Damage State	Probability of Failure	"Generic" Probability of Failure	Ratio to "Generic"	Likelihood of Damage (before inspection)	Partial Damage Factor (no insp.)	Likelihood of Damage (after inspection)	Partial Damage Factor (1 insp.)
1	8×10^{-6}	1×10^{-4}	0.08	0.5	0.04	0.81	0.06
2	2×10^{-5}	1×10^{-4}	0.2	0.3	0.06	0.14	0.03
3	5×10^{-3}	1×10^{-4}	50	0.2	10	0.05	2
Total Technical Module Sub-Factor					10		2

APPENDIX 3-6

Data specification for case study 2 for TMSF

Table A.3.6: data collection and specification for **8"-P-5-06-203-31020-50HCS**

Parameter	Small bore 3/4"-1-IS-116-V-195	Rating	Big bore 8"-1-IG-116-L-172	Rating
Operating temperature (Marine)	75°C	0.0508	75°C	0.0508
Pipe support or penalty interface	Yes	2	Yes	2
Complexity factor	Default	1	Default	1
Equipment Insulation condition	Below average	1	Average	0.5
Coating quality	None	1.25	Medium	1
Calculated corrosion rate	0.0508x2x1x1x1.25= 0.127mmpyr ≈ 0.005in./yr	Below than state 1	0.0508x2x1x10.5 x1.25=0.0.0635mmpyr ≈ 0.0025in/yr	Below than state 1

Table A.3.6.1: Calculate the Composite Technical Module Subfactor for all Damage Mechanisms [1]

Damage State	Corrosion Rate	Wall Loss	Remaining Wall	Frequency of Failure
1	0.010 in./yr	0.06	0.315	8×10^{-6}
2	0.020 in./yr	0.12	0.255	2×10^{-5}
3	0.040 in./yr	0.24	0.135	5×10^{-3}

Table 8-9—Calculated Technical Module Subfactor

Damage State	Probability of Failure	"Generic" Probability of Failure	Ratio to "Generic"	Likelihood of Damage (before inspection)	Partial Damage Factor (no insp.)	Likelihood of Damage (after inspection)	Partial Damage Factor (1 insp.)
1	8×10^{-6}	1×10^{-4}	0.08	0.5	0.04	0.81	0.06
2	2×10^{-5}	1×10^{-4}	0.2	0.3	0.06	0.14	0.03
3	5×10^{-3}	1×10^{-4}	50	0.2	10	0.05	2
Total Technical Module Sub-Factor					10		2

APPENDIX 3-7

Table A3.7: data collection and specification for **36"-P-5-03-016-61010-34CC**

Parameter	Small bore 3/4"-1-IS -110-V 034	Big bore 36"-1-IS -110-L 033
Operating parameters • Op.temperature • Op.pressure	 27°C 64.7 BAG	 27°C 64.7 BAG
Design parameters Design temperature Design pressure	 80°C 81.7 BAG	 80°C 81.7 BAG
Pipe size	0.75 inch	8 inch
Pipe Schedule	XS	XS
Universal subfactor • Plant condition • Cold weather • Seismic Activity	Set by default	Set by default
Mechanical Subfactor	Set by default	Set by default
Process Subfactor • Continuity • Stability • Relief valves	Set by default	Set by default

APPENDIX 3-8

Table A3.8: data collection and specification for **8"-P-5-06-203-31020-50HCS**

Parameter	Small bore 3/4"-1-IS-116-V-195	Big bore 8"-1-IG-116-L-172
Operating parameters • Op.temperature • Op.pressure	 75°C 14.3BAG	 75°C 14.3BAG
Design parameters • Design temperature • Design pressure	 163°C 29.3BAG	 163°C 29.3BAG
Pipe size	1.5 inch	18 inch
Pipe Schedule	Std	Std
Universal subfactor • Plant condition • Cold weather • Seismic Activity	Set by default	Set by default
Mechanical Subfactor	Set by default	Set by default
Process Subfactor • Continuity • Stability • Relief valves	Set by default	Set by default

APPENDIX 4-1

Sample of MATLAB coding for logistic regression (small bore)

```
>> x=x;
>> y=y;
>> n=[1];
>> b=glmfit(x,[y ones(318,1)],'binomial','link','logit');
>> fitted = glmval(b,x,'logit');
>> loglikelihood = sum(log(binopdf(y,n,fitted)));
>>                                [b,dev,stats]=glmfit(x,[y
ones(318,1)],'binomial','link','logit')

b =

   -3.9804
    0.2366
    1.8954
    1.6749
    1.4695
    1.6457
    1.2761


dev =

  389.4893


stats =

         beta: [7x1 double]
          dfe: 311
         sfit: 1.0160
            s: 1
      estdisp: 0
         covb: [7x7 double]
           se: [7x1 double]
    coeffcorr: [7x7 double]
            t: [7x1 double]
            p: [7x1 double]
        resid: [318x1 double]
       residp: [318x1 double]
       residd: [318x1 double]
       resida: [318x1 double]
```

APPENDIX 4-2

Sample of MATLAB coding for logistic regression (1parameter)

```
function [beta,Iter] = NR_logistic(data,beta_start)
x=data(:,1); % x is first column of input data
y=data(:,2); % y is second column of response data
n=length(x)
diff = 1;
beta = beta_start; % initial values
while diff>0.0001 % set the convergence criterion
beta_old = beta;
p = exp(beta(1)+beta(2)*x)./(1+exp(beta(1)+beta(2)*x));
Loglikelihood = sum(y.*log(p)+(1-y).*log(1-p))
s = [sum(y-p); % scoring function for Newton Rahpson
sum((y-p).*x)];
Iter = [sum(p.*(1-p)) sum(p.*(1-p).*x); % information
matrix
sum(p.*(1-p).*x) sum(p.*(1-p).*x.*x)]
beta = beta_old + Iter\s % new value of beta
diff = sum(abs(beta-beta_old)); % sum of absolute
differences
end

*******************************************

xy=[0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 ; 0 0 0 0 1 1 1
1 1 1 0 0 1 1 1 1 1 1 1 1];

mat=transpose(xy) % transpose the value in vertical matrix

bnot=[0;0];    % start initial guess (0,0)

NR_logistic(mat,bnot)% call NR_logistic function
```

APPENDIX 5-1

Example of calculation for probability of CUI occurrence for operating temperature G1 within 30 years (small bore).

Age	**y**	**exp(y)**	**1+exp(y)**	**exp(y)/[1+exp(y)]**
1	-1.848	0.157	1.157	0.136
2	-1.612	0.200	1.200	0.166
3	-1.375	0.253	1.253	0.202
4	-1.139	0.320	1.320	0.243
5	-0.902	0.406	1.406	0.289
6	-0.665	0.514	1.514	0.340
7	-0.429	0.651	1.651	0.394
8	-0.192	0.825	1.825	0.452
9	0.044	1.045	2.045	0.511
10	0.281	1.324	2.324	0.570
11	0.518	1.678	2.678	0.627
12	0.754	2.126	3.126	0.680
13	0.991	2.693	3.693	0.729
14	1.227	3.412	4.412	0.773
15	1.464	4.323	5.323	0.812
16	1.701	5.477	6.477	0.846
17	1.937	6.939	7.939	0.874
18	2.174	8.792	9.792	0.898
19	2.410	11.138	12.138	0.918
20	2.647	14.112	15.112	0.934
21	2.884	17.879	18.879	0.947
22	3.120	22.651	23.651	0.958
23	3.357	28.697	29.697	0.966
24	3.593	36.357	37.357	0.973
25	3.830	46.063	47.063	0.979
26	4.067	58.358	59.358	0.983
27	4.303	73.936	74.936	0.987
28	4.540	93.672	94.672	0.989
29	4.776	118.676	119.676	0.992
30	5.013	150.355	151.355	0.993

APPENDIX 5-2

The probability of CUI occurrence for all operating temperature groups within 30 years (small bore).

	Operating Temperature (degree Celsius)					
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Age	49 to 93	-12 to 16	16 to 49	93 to 121	less than -12	more than 121
1	0.136	0.112	0.093	0.109	0.078	0.023
2	0.166	0.138	0.115	0.135	0.097	0.029
3	0.202	0.169	0.142	0.165	0.120	0.037
4	0.243	0.204	0.173	0.200	0.147	0.046
5	0.289	0.246	0.210	0.240	0.179	0.057
6	0.340	0.292	0.251	0.286	0.217	0.072
7	0.394	0.343	0.298	0.337	0.260	0.089
8	0.452	0.398	0.350	0.391	0.308	0.110
9	0.511	0.456	0.406	0.449	0.360	0.136
10	0.570	0.515	0.464	0.508	0.416	0.166
11	0.627	0.574	0.523	0.567	0.475	0.201
12	0.680	0.630	0.581	0.624	0.534	0.242
13	0.729	0.684	0.638	0.677	0.592	0.288
14	0.773	0.732	0.690	0.727	0.648	0.339
15	0.812	0.776	0.738	0.771	0.699	0.394
16	0.846	0.815	0.782	0.810	0.747	0.451
17	0.874	0.848	0.819	0.844	0.789	0.510
18	0.898	0.876	0.852	0.873	0.826	0.569
19	0.918	0.899	0.879	0.897	0.857	0.626
20	0.934	0.919	0.902	0.917	0.884	0.680
21	0.947	0.935	0.921	0.933	0.906	0.729
22	0.958	0.948	0.937	0.946	0.924	0.773
23	0.966	0.958	0.949	0.957	0.939	0.812
24	0.973	0.967	0.960	0.966	0.951	0.845
25	0.979	0.974	0.968	0.973	0.961	0.874
26	0.983	0.979	0.974	0.978	0.969	0.898
27	0.987	0.983	0.980	0.983	0.975	0.917
28	0.989	0.987	0.984	0.986	0.981	0.934
29	0.992	0.990	0.987	0.989	0.985	0.947
30	0.993	0.992	0.990	0.992	0.988	0.958

APPENDIX 6-1

Example of calculation for probability of CUI occurrence for calcium silicate within 30 years (big bore)

Age	**y**	**exp(y)**	**1+exp(y)**	**exp(y)/[1+exp(y)]**
0	-3.247	0.039	1.039	0.037
1	-2.986	0.051	1.051	0.048
2	-2.724	0.066	1.066	0.062
3	-2.463	0.085	1.085	0.079
4	-2.202	0.111	1.111	0.100
5	-1.940	0.144	1.144	0.126
6	-1.679	0.187	1.187	0.157
7	-1.417	0.242	1.242	0.195
8	-1.156	0.315	1.315	0.239
9	-0.895	0.409	1.409	0.290
10	-0.633	0.531	1.531	0.347
11	-0.372	0.690	1.690	0.408
12	-0.110	0.896	1.896	0.472
13	0.151	1.163	2.163	0.538
14	0.413	1.511	2.511	0.602
15	0.674	1.962	2.962	0.662
16	0.935	2.548	3.548	0.718
17	1.197	3.309	4.309	0.768
18	1.458	4.298	5.298	0.811
19	1.720	5.582	6.582	0.848
20	1.981	7.249	8.249	0.879
21	2.242	9.415	10.415	0.904
22	2.504	12.228	13.228	0.924
23	2.765	15.881	16.881	0.941
24	3.027	20.625	21.625	0.954
25	3.288	26.787	27.787	0.964
26	3.549	34.789	35.789	0.972
27	3.811	45.182	46.182	0.978
28	4.072	58.680	59.680	0.983
29	4.334	76.211	77.211	0.987
30	4.595	98.978	99.978	0.990

APPENDIX 6-2

The probability of CUI occurrence for all insulation type within 30 years (big bore).

Age	Calcium silicate	Cellular glass
0	0.037	0.101
1	0.048	0.127
2	0.062	0.159
3	0.079	0.197
4	0.100	0.241
5	0.126	0.292
6	0.157	0.349
7	0.195	0.411
8	0.239	0.475
9	0.290	0.540
10	0.347	0.604
11	0.408	0.665
12	0.472	0.720
13	0.538	0.770
14	0.602	0.813
15	0.662	0.849
16	0.718	0.880
17	0.768	0.905
18	0.811	0.925
19	0.848	0.941
20	0.879	0.954
21	0.904	0.964
22	0.924	0.972
23	0.941	0.979
24	0.954	0.983
25	0.964	0.987
26	0.972	0.990
27	0.978	0.992
28	0.983	0.994
29	0.987	0.995
30	0.990	0.996

APPENDIX 7-1
Sample data for small bore piping system

LINE DESIGNATION	TEMPERATURE GROUPS	INSULATION GROUP	RESPONSE
P-2-1083-D6103-D(N34A)	3	2	1
P-1-1066-D6103-D(N34A)	3	2	1
P-1039-E6123-H(N20A)	3	1	1
P-1038-E6123-H(N20A)	3	1	1
P-1082-D6123-D(N34A)	3	2	1
P-1-1086-D6103-D(N34A)	3	2	1
P-2-1084-D6103-D(N34A)	3	2	1
P-2-1043-D6103-D(N34A)	3	2	0
P-3/4-1084-D6103-D(N34A)	3	2	1
P-2-3009-D6308-D(N34A)	3	2	1
P-3/4-1040-D6103-H(N34A)	3	2	1
P-3/4-1073-D6308-P(N23A)	3	1	1
P-1-1074-D6308-P(N23A)	3	1	1
P-3/4-1074-D6508-P(N23A)	3	1	1
P-1-1044-D6103-H(N34A)	3	2	0
HF-1-1014-D1101-H(N20A)	1	1	1
P-0.75-2004-D6308-H(N20A)	1	1	1
PL-1-2045-D1306-H(N10A)	4	1	1
PL-3/4-2014-D1306-H(N20A)	4	1	1
PL-1/2-2016-D1306-H(N10A)	4	1	1
PL-3/4-2016-D1306-H(N10A)	4	1	1
PL-2-2017-D1306-H(N10A)	4	1	1
PL-11/2-2017-D1306-H(N10A)	4	1	1
PL-0.75-2001-D6038-H(N10A)	4	1	0
PL-1-2001-D6038-H(N10A)	4	1	0
PL-1-2001-D6038-H(N10A)	4	1	1
PL-0.75-2010-D1306-H(N20A)	4	1	0
PL-11/2-2016-D1306-H(N10A)	4	1	1
PL-3/4-2011-D1306-H(N20A)	4	1	1
LLS-1-G4-AGRU-LP-036-D1101-H(N20A)	2	1	1
LLS-3/4-G4-AGRU-LP-037-D1101-H(N20A)	2	1	1
PL-3/4-2007-D1306-H(N10A)	4	1	1
PL-1-2007-D1306-H(N10A)	4	1	1
PL-3/4-2007-D1306-H(N10A)	4	1	1

PL-3/4-2068-D1306-H(N10A)	1	1	1
PL-3/4-2067-D1306-H(N10A)	1	1	1
PL-3/4-2007-D1306-H(N20A)	4	1	1
PL-0.75-2021-D6308-H(N20A)	4	1	1
PL-1-2021-D6308-H(N20A)	4	1	1
PL-0.75-2019-D1306-H(N10A)	4	1	1
PL-0.5-2019-D1306-H(N10A)	4	1	1
PL-1-2019-D1306-H(N10A)	4	1	1
PL-3/4-2037-D1306-H(N20A)	1	1	1
PL-1/2,3/4-2036-D1306-H(N20A)	1	1	1
PL-3/4-2037-D1306-H(N20A)	1	1	1
PL-3/4-2036-D1306-H(N20A)	1	1	1
PL-0.75-2022-D6308-H(N10A)	4	1	1
PL-1-2022-D6308-H(N10A)	4	1	1
P-11/2-4509-D1101-H(N20A)	3	1	0
PR-1-7023-D1101-C(N31A)	3	2	1
P-2-3015-D6103-D(N34A)	3	2	1
P-1-4058-C6120-C(N31A)	2	2	1
P-1/2-3519-D3102-C(N31A)	3	2	1
P-1/2-4058-C6120-C(N31A)	2	2	1
P-3/4-4058-C6120-C(N31A)	2	2	1
P-3/4-3519-D3102-C(N31A)	3	2	1
PR-3/4-7010-D1101-C(N31A)	3	2	1
PR-3/4-7017-D1101-C(N31A)	3	2	1
LD-1-4001-C6192-C(N31A)	5	2	0
P-0.75-6001-C3110-C(N31B)	2	2	1
P-1-4001-C6120-C(N31A)	5	2	1
P-0.75-4001-C6120-C(N31A)	5	2	1
P-0.5-4028-C6021-C(N31B)	5	2	0
P-0.75-4013-D3102-C(N31B)	5	2	1
P-2-4013-D3102-C(N34A)	5	2	1
P-0.75-6027-C3110-C(N31A)	2	2	0
P-0.5-6029-C3110-C(N31A)	2	2	1
P-2-4011-C3110-C(N31A)	5	2	1
P-1.5-6001-C1109-C(N31A)	2	2	0
P-1-4006-C3110-C(N31A)	5	2	1
P-0.75-4006-C3110-C(N31A)	5	2	0
P-2-4010-C3110-C(N31A)	5	2	1
LD-0.75-6001-D3102-H(N20A)	5	1	1
PR-1.5-7034-D1101-C(N31A)	2	2	1
CF-1-6007-C3110-C(N31A)	2	2	1