CHAPTER 4

MODEL DEVELOPMENT

4.1 Research Framework

The research project focused on finding an appropriate quantitative model for the purpose of assessing the probability of failure for piping systems subject to CUI in RBI analysis. The study was driven by the type of data available for CUI where in this case the wall thickness data, which is typically used to assess corrosion failure, were found to be quite limited. Therefore, several mathematical models were explored to go well with the available data as shown in Figure 4.1.

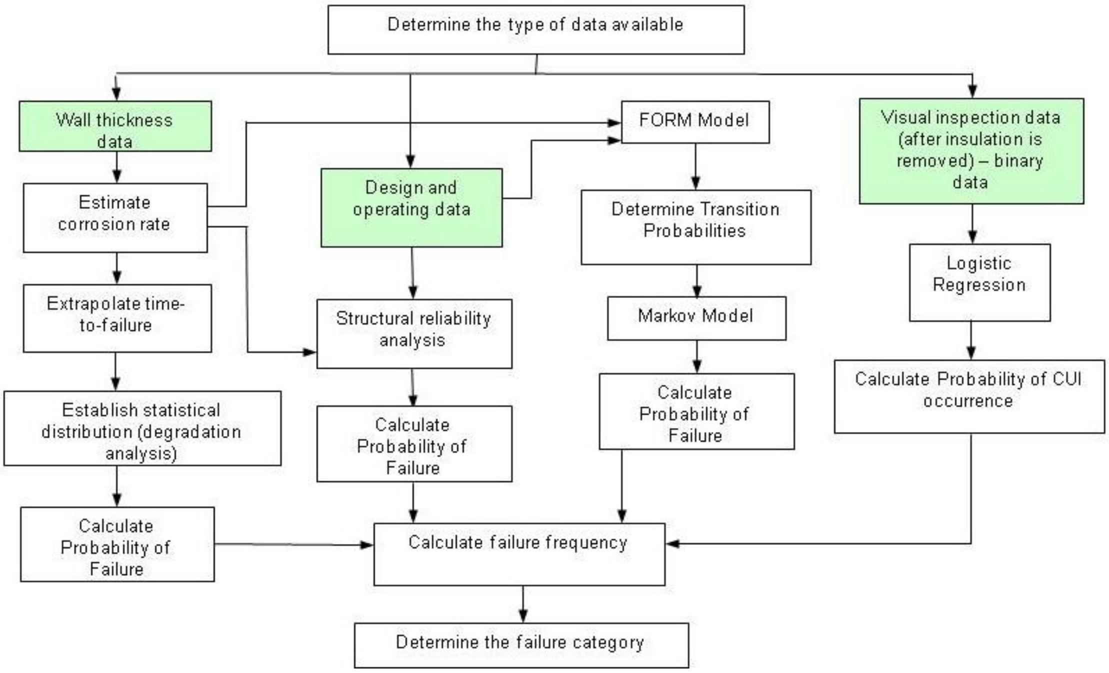

Figure 4.1: Research framework

From Figure 4.1, it can be seen that there are typically three types of CUI inspection data as follows:

1) *Visual inspection data*: This data were collected after insulation was removed, however, no wall thickness measurements were being taken. Data was recognized as either CUI was found or CUI was not found i.e. pass or fail data.
2) *Wall thickness data*: This data were collected after insulation was removed at the thickness measurement locations (TMLs).
3) *Design and operating data*: Examples of design and operating data collected are design pressure and temperature as well as operating pressure and temperature.

As discussed in Chapter 3, four models were proposed in this study which are as follows:

1) *Logistic regression model*: This model was applied based on the visual inspection data which can be treated as binary data (1 = CUI was found; 0 = CUI was not found).
2) *Degradation analysis*: This model was applied if the wall thickness data are available and the number of data is sufficient to carry out the degradation analysis. This model does not require the design and operating data.
3) *Structural reliability analysis*: This model was applied when the wall thickness data are limited and it took the advantage of having the design and operating data.
4) *Continuous time Markov model*: This model assumes the corrosion progresses through several states and in this study, the data required for this model was similar to the structural reliability model in determining the transition rates.

4.2 Logistic Regression Model

In developing a logistic regression model, several steps have been employed. Figure 4.2 shows the framework.

Step 1: Data Acquisition

All information related to CUI was collected.

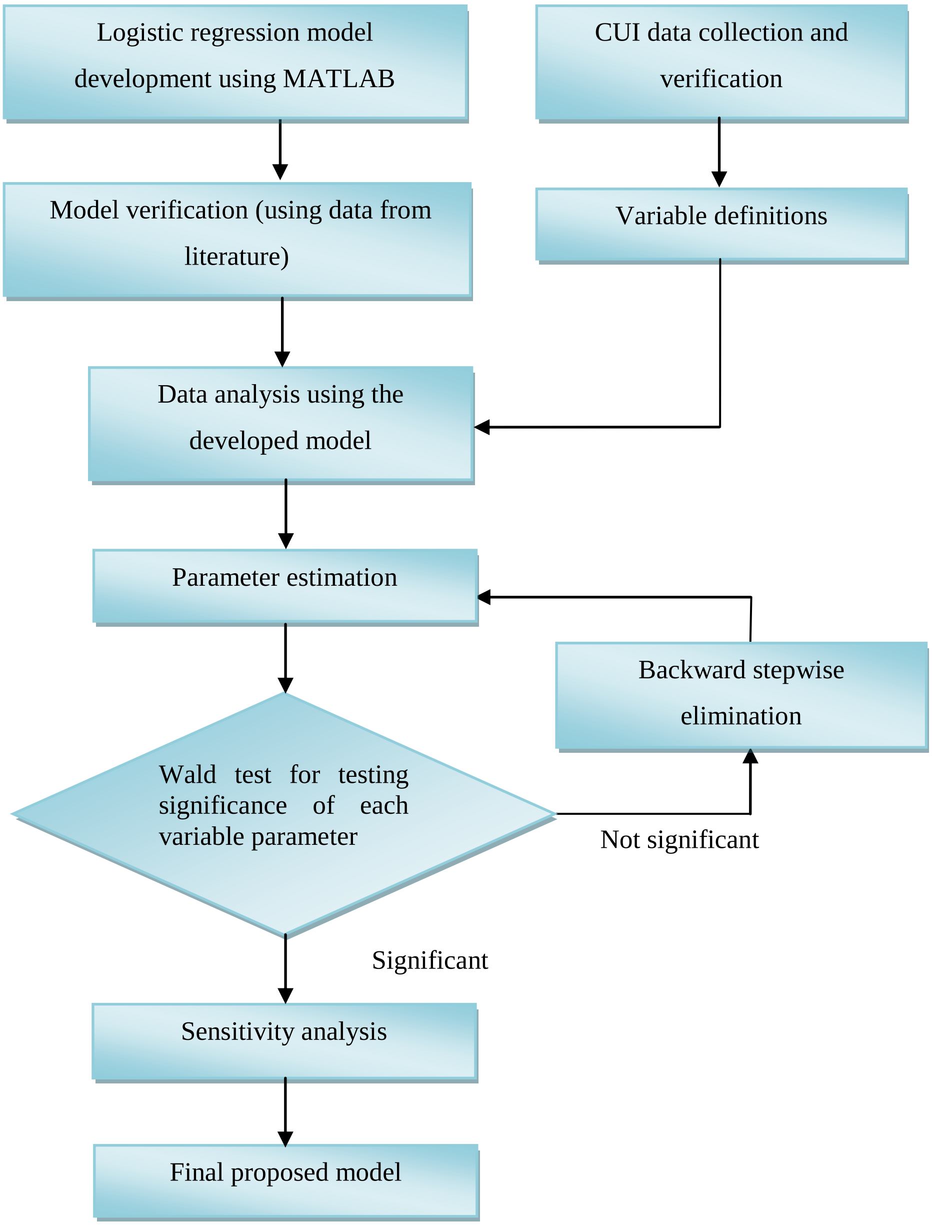

Figure 4.2: Logistic regression model flowchart for CUI

Step 2: Variable definitions

Based on the data collected, explanatory variables were identified. In this study, three explanatory variables, which were pipe age, operating temperature and insulation type, were employed for the logistic model development based on the availability of data in the inspection database.

Pipe age was classified as a continuous variable where data from three age groups were collected (i.e. pipe age 6, 10 and 15). Operating temperature was clustered based on API 581 operating temperatures and as such were defined as categorical variables. Categorical variables are the same as dummy variables which are artificial explanatory variables in a regression model. In this case, the dummy variables represent the categories of the operating temperature. Each dummy variable assumes one of two values, 0 or 1, indicating whether an observation falls in a particular group.

Operating temperature more than 121°C was named as Group 6 and was referred as the reference group. Hence, five additional dummy variables were defined for operating temperature with respect to the reference as follow:

- Group 1: 1 when operating temperature is in the range 49°C to 93°C, 0 otherwise
- Group 2: 1 when operating temperature is in the range -12°C to 16°C, 0 otherwise
- Group 3: 1 when operating temperature is in the range 16°C to 49°C, 0 otherwise
- Group 4: 1 when operating temperature is in the range 93°C to 121°C, 0 otherwise
- Group 5: 1 when operating temperature is less than -12°C, 0 otherwise

Insulation type can be classified into two groups and therefore, was also defined as categorical variables in the logistic model. Table 4.1 shows the different requirements of the insulation that determines the different insulation material. There are two categories of insulation material: calcium silicate (type 1) and cellular glass (type 2). 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

Table 4.1: Classification for types of piping insulation class (Engineering Specification for Thermal Insulation Design, 1994)

Insulation requirement	Temperature range (°C)	Insulation material	Insulation type
Heat Input Control	0 to 650	Calcium silicate	1
Heat Conservation	10 to 650	Calcium silicate	1
Personnel Protection	65 to 650	Calcium silicate	1
Cold conservation	+10 to -180	Cellular glass	2
Prevention of Surface Condensation	10 to ambient	Cellular glass	2

The response variable is classified as binary response and may be classified as either CUI is found ($Y = 1$) or CUI is not found ($Y = 0$).

Step 3: Model development using MATLAB

The logistic model was developed in MATLAB R2009a using two main functions; *glmfit* and *glmval* functions. The *glmfit* function is used to estimate the coefficients of the parameters in the logistic regression model. The syntax is as follows:

b = glmfit (*data*, *Y*, '*distr*')

where , *b* is a vector of coefficients, *data* is a vector of the predictors, *Y* is a vector of observed responses and '*distr*' is the distribution that can be any of the following strings: 'binomial', 'gamma', 'inverse gaussian', 'normal' (by default), and 'poisson'.

The *glmval* function is used to compute the predicted values for the generalized linear model with link function '*link*' and predictors *X*. The syntax is as follows:

yfit = glmval (*b*, data, '*link*')

where *b* is a vector of coefficients estimated as returned by the *glmfit* function, *data* is a vector of the predictors and '*link*' is the link function that can be any of the strings used in the *glmfit* function. Refer to Appendix B for further explanation.

Step 4: Model verification

To verify the logistic regression model developed using MATLAB, data taken from Evans Country case study (Kleinbaum et al., 1982) was used. In the case study, the data was fitted to the logistic regression model using Java Script developed by Sullivan and Pezzullo (2007). The results generated by MATLAB were compared to the results generated via Java Script.

Step 5: Parameter Estimation

Once a logistic regression is specified with its parameter and data have been collected, one is in a position to evaluate its goodness of fit, that is, how well it fits the observed data. Goodness of fit is assessed by finding the parameter value of a model and the procedure is known as parameter estimation (Agresti, 1990). Estimates of the parameters can be found by using a mathematical technique called Maximum Likelihood Estimation (MLE).

Step 6: Testing the significance of each parameter

After generating the parameters for each variable, it is necessary to test the statistical significance of each parameter in the model. In the formulation of a logistic model, Wald test was performed on each variable or model parameter to investigate its significance (Yang et al., 2005). Wald tests are based on the 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.

Step 7: Backward Stepwise Elimination

If the parameters obtained from result analysis are not significant, then a backward stepwise elimination method will be conducted to eliminate the parameter which is insignificant. Backward stepwise elimination method 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 biggest p-value is deleted (if the p-value is bigger than the chosen cutoff level). Refer back to Chapter 3.

Step 8: Sensitivity analysis

The objective of the sensitivity analysis is to validate the proposed model and to test the reliability of the model by evaluating its sensitivity to minor changes in the data set. In order to conduct sensitivity analysis, new logistic models are developed using 80% and 90% of the set of data based on the proposed method by Ariaratnam et al. (2001). Then, these new models will be compared to the 100% data set (i.e. original data).

To show that there is no difference among these three models from statistical point of view, Kruskal-Wallis test was performed. Kruskal-Wallis test is used to evaluate the differences between three or more treatment conditions (or populations) and this test was performed to test the following hypotheses:

- H_0: The models are equal (no significant difference between models)
- H_A: The models are different.

Step 9: Estimate the probability of CUI occurrence to be used in RBI analysis

Once an appropriate model was validated, then the probability of CUI occurrence was estimated for inspection planning purpose.

4.3 Degradation Analysis Framework

The steps shown in Figure 4.3 were followed in order to perform degradation analysis. The steps were based on the methodology outlined in Meeker and Escobar (1998).

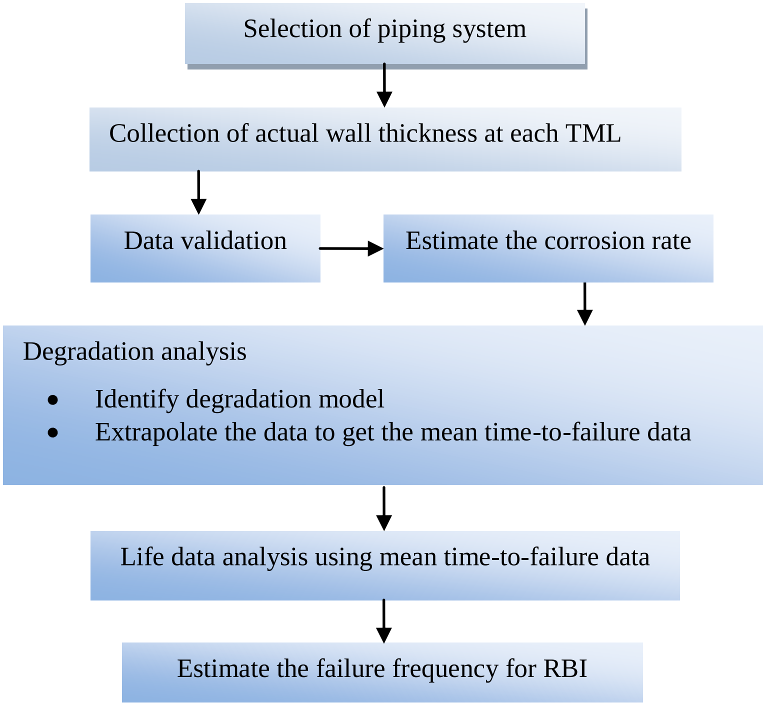

Figure 4.3: Degradation analysis framework to estimate reliability of piping system subject to CUI

Step 1: Data collection

The data collection phase includes identifying the piping systems which are susceptible to CUI based on API 581 guideline and gathering all the necessary information that need to be used in degradation analysis model. The required data are as follows:

- the nominal wall thickness t_{nominal} (from design data) or the initial wall thickness t_{initial} (wall thickness measured at the commencement of the piping system)
- the actual wall thickness measures t_{actual} after being in service for T number of years
- the minimum required wall thickness t_{min} (from design data)

Step 2: Estimation of corrosion rate

CUI corrosion rates were then determined. The long-term corrosion rate was then estimated using Eq. (3.7). The reason why the short-term corrosion rate was not used is because of, for insulated piping systems, the wall thickness was not measured at each TML during each inspection period. Therefore, the information in hand was only the actual wall thickness after being in service for T number of years in service. Moreover, since the initial thickness t_{initial} is not always available, the nominal thickness t_{nominal} will replace t_{initial} for Eq. (3.7). In the case where t_{initial} is available, t_{initial} will replace t_{nominal}. Thus, the corrosion rate formula is stated as:

$$CR = \frac{t_{\text{nominal}} - t_{\text{actual}}}{T} \tag{4.1}$$

where CR = corrosion rate (in mm/yr), t_{nominal} = the nominal pipe thickness, t_{actual} = the actual pipe thickness and T = number of years in service.

Step 3: Perform degradation analysis

The degradation model for corrosion was identified. In this study, the linear model was assumed (i.e. using Eq. (4.1)) and the extrapolation of the time-to-failure for each TML is based on this linear model. Failure is defined as when the wall thickness reaches the minimum wall thickness specified.

Step 4: Perform life data analysis

The mean time-to-failure data generated in Step 3 was fitted to an appropriate distribution using Weibull++ software in order to estimate the distribution parameters. Then, the distribution fitting was checked by comparing the log-likelihood values generated for each distribution. The largest log-likelihood value indicates the best fit.

Step 5: Estimate the probability of failure for RBI analysis

Once an appropriate distribution was validated, then the failure probability and the failure frequency were estimated to be used in RBI analysis. The cumulative density function for failure probability is estimated using Eq. (3.21) for Weibull distribution, Eq. (3.23) for exponential distribution or Eq. (3.26) for lognormal distribution. The failure frequency/rate is estimated using Eq. (3.22).

4.4 Structural Reliability Analysis Framework

The following steps illustrate the process used to develop structural reliability analysis methodology.

Step 1: Identify the limit state function

A limit state function needs to be defined where this function expresses the criterion for failure of the pipe. Often it is common for the limit state function to be the difference between resistance and the load. For corrosion failure in pipe, the thinning model proposed by Khan et al. (2006) was used as the limit state function in this study. Khan et al. (2006) stated that the thinning model can be used for both internal and external corrosions; however, the corrosion rate may differ from one case to the other. The limit state function can be defined as

$$g(x) = R - St = S\left(1 - \frac{CR \times T}{t}\right) - \frac{PD}{2t} \qquad (4.2)$$

where R is material resistance (in MPa), St is the applied stress (in MPa), S is material yield strength (in MPa), CR is corrosion rate (in mm/yr), P is operating pressure (in MPa), D is the diameter of the component (in mm), t is the material thickness (in mm) and T is number of years in service (in year).

Step2: Identify the basic random variables

Based on the thinning model, the following were the basic random variables: S is material yield strength, $\boldsymbol{CR}$ is corrosion rate, $\boldsymbol{P}$ is operating pressure, $\boldsymbol{D}$ is the diameter of the component and $\boldsymbol{t}$ is the material thickness.

Step 3: Data collection

The data collection phase includes gathering all the necessary information that need to be used in structural reliability analysis including the corrosion rate for CUI. Corrosion rate is estimated using Eq. (4.1).

Step 4: Estimation of probability distribution and its parameter

In this step, the variability of the random variables S, CR, P, D and t was reviewed and estimation of their probability distributions and their parameters are made. Further explanation on how to estimate the distributions and their parameters can be found in Appendix D.

Step 5: Development of FORM model in using spreadsheet

FORM model was built using Microsoft Excel and Visual Basic for Application. A practical procedure presented by Zhao and Ono (1999), Low and Tang (2004) and Low and Tang (2007) was used to develop FORM model in Microsoft Excel. The reliability index was obtained by calling Excel's built-in optimization program, Solver, with the objective function to minimize the reliability index subject to the constraint that the limit-state function, $g(x) = 0$. The probability of failure is then calculated. Refer to Appendix D for the FORM model framework built in Microsoft Excel and the code for Visual Basic Application.

Step 6: Validation of developed FORM model

The FORM algorithm developed was validated using two case studies published in the literature (Cardoso et al., 2008; Teixeira et al., 2008). Both case studies were not about CUI; however, they were on assessing reliability of pipelines with corrosion defect.

Step 7: Running FORM model using CUI data

Once the FORM model was validated, data on CUI samples were run using the model.

Step 8: Results verification

The results were verified using Monte Carlo simulation approach. The results were also compared to another limit state function that typically used for corroded piping and pipelines which is the difference between the pipe failure pressure P_f and the pipe operating pressure P_{op} as stated (Caleyo et al., 2002; Vinod et al. 2003; Teixeira et al., 2008):

$$g(x) = P_f - P_{op} \tag{4.3}$$

A failure pressure model P_f that is typically used to determine the failure probability is the ASME failure model, the modified B31G (Vinod et al., 2003; Escoe, 2006). The modified B31G failure pressure model is defined as

$$P_f = \frac{2(S+68.95)t}{D}\left(\frac{1-0.85\frac{d(T)}{t}}{1-0.85\frac{d(T)}{t}M^{-1}}\right) \tag{4.4}$$

where

$$M = \begin{cases} \sqrt{1 + 0.6275\frac{l^2}{Dt} - 0.003375\frac{l^2}{D^2t^2}} & for\ \frac{l^2}{Dt} \leq 50 \\ 0.32\frac{l^2}{Dt} + 3.3 & for\ \frac{l^2}{Dt} > 50 \end{cases} \tag{4.5}$$

where $d(T)$ is depth of corrosion (in mm), S is the material yield strength (in MPa), t is the thickness of the pipe (in mm), D is the outer diameter of the pipe (in mm), CR is the corrosion rate (in mm/yr), P_{op} is the operating pressure (in MPa), T is the time of inspection (in year) and l is the axial length of corrosion defect (in mm). It is assumed that the depth of corrosion, $d(T)$ is

$$d(T) = 0.20t + CR \times T \tag{4.6}$$

4.5 Continuous-Time Markov Model Framework

The following steps described the process employed to develop a continuous Markov model.

Step 1: Identify the Markov model to be used

This study proposed a three-state Markov model for CUI damage mechanism. The proposed model was based on a three-state Markov model for general wall thinning model proposed by Fleming (2004).

Step 2: Identify the definition for each state

Fleming model consists of three states reflecting the progressive stage of pipe failure mechanism: the state with no damage, development of detectable damage and the occurrence of leaks. In this model, the occurrence of leaks is defined as a failure which does not mean there are actual leaks but rather the thickness of wall has reached the minimum wall thickness specified.

The definitions given for each state in Fleming model were very general. Therefore, this study proposed the definition shown in Table 3.7 in Chapter 3. The definition of states can be based on the depth of corrosion defects.

Step 3: Develop the differential equations

The Markov model was solved by setting up the differential equations (refer to Chapter 3) and the solution of the differential equations was obtained using Laplace transforms as can be seen in Appendix C.

Step 4: Determine the state probabilities

Once the solutions for abovementioned differential equations were obtained, the calculations were performed via spreadsheet.

Step 5: Estimate the transition rates using FORM model

The transition rates are determined using FORM model based on the formulation of the limit state functions discussed in Chapter 3. Also, the repair rate for corrosion defects was estimated based on the characteristics of inspection and mean time to repair corrosion defects upon detection.

Step 6: Model verification and validation

The developed Markov model using spreadsheet was validated by comparing the state probabilities generated with the results generated using MATLAB (i.e. MATLAB was used to solve numerically the differential equations).

Step 7: Running Markov model using CUI data

Once the Markov model was validated, data on CUI samples were run using the model.

4.6 Concluding Remarks

This chapter outlines the research methodology for the four proposed models in assessing quantitatively the probability of failure due to CUI which are logistic regression, degradation analysis, structural reliability analysis and continuous-time Markov model. Each step is discussed in details.