Statistical Modeling
Of
Fracture Toughness Data

by

Guru Prakash

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil and Environmental Engineering

Waterloo, Ontario, Canada, 2007

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The fracture toughness of the zirconium alloy (Zr-2.5Nb) is an important parameter in determining the flaw tolerance for operation of pressure tubes in reactor. Fracture toughness data have been generated by performing rising pressure burst tests on sections of pressure tubes removed from operating reactors. The test data were used to generate a lower-bound fracture toughness curve, which is used in defining the operational limits of pressure tubes. The thesis presents a comprehensive statistical analysis of burst test data and develops a multivariate statistical model to relate toughness with material chemistry, mechanical properties, and operational history. The proposed model can be useful in predicting fracture toughness of specific in-service pressure tubes, thereby minimizing conservatism associated with a generic lower-bound approach.

Acknowledgements

I would like to thank my supervisor Professor Mahesh D. Pandey for his support, encouragement, guidance, and patience during my study at the University of Waterloo.

I would like to thank Professor Galena Morozova for providing me Teaching Assistantship for the CIVE153 course.

I would like to thank Mike Cocker for supporting me in fixing my computer.

I would also like to thank the University of Waterloo for providing financial support in the form of the International Master Student Award.

Finally, I would like offer my thanks to my friends for their financial, emotional, and social support.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Introduction

The pressure tubes used in CANDU® reactors are fabricated from cold-worked Zr-2.5Nb and are a length of 6.3 m, with an inside diameter of 103 mm, and a wall thickness of 4.2 mm. During service, irradiation and deuterium ingress from the pressurized heavy water coolant reduces the fracture toughness of the pressure tube material. Periodic assessments of surveillance tubes are removed from the reactors and inspection are conducted to ensure that the tubes remain "fit-for-service" [1]. Currently, 106 burst tests have been performed on sections irradiated Zr 2.5Nb pressure tubes removed from operating reactors using the standardized method [3]. The measured values of K_c from a portion of these tests were used to generate a lower-bound curve, thereby defining the operational limits of pressure tubes. Such a conservative approach was deemed necessary due in part to the significant tube-to-tube variability in measured fracture properties. Previous studies have identified specific material characteristics that influence pressure tube fracture toughness, and the variability in the burst test results. The role of chlorine in the formation of primary void nucleation sites for fracture, for exampled , highlighted the importance of controlling the chemical composition and fabrication routes of pressure tubes [3, 5].

1.2 Objectives

In the current study, a comprehensive multivariate statistical analysis of the burst test database is performed to correlate fracture toughness with relevant variables, such as material chemistry, mechanical properties and irradiation history. As a result, a significant portion of the burst test

® Canada Deuterium Uranium (CANDU) is a registered trademark of Atomic Energy of Canada Limited.

fracture toughness variability is addressed, and the statistical influences of the covariates are quantified. As part of the analysis, predictive models for pressure tube fracture toughness are developed from the most significant covariates.

The objective of this study is to develop an advanced multivariate statistical model to relate the fracture toughness with covariates such as material composition, mechanical properties and other parameters related to reactor in service conditions. To apply the regression procedure twelve independent variables have been taken.

1.3 Thesis Organization

The outline of the thesis is as follows. In Chapter two a description of the irradiated Zr-2.5Nb pressure tube material and burst test methodology is presented. In Chapter three a simple regression model is developed between lower bound for fracture toughness and test temperature. In Chapter four a description of the multivariate statistical techniques is presented followed by a multivariate analysis of data. The results from the multivariate analysis are then presented. In Chapter five stepwise regression analysis is performed. The result obtained from forward, backward and stepwise regression is compared and the best suitable model is considered. Finally, conclusion is presented in Chapter six.

Chapter 2
Preliminary Analysis

2.1 CSA Method to determine lower bound fracture toughness

CAN/CSA-N285.4-94 describes an empirical equation to determine the lower-bound fracture toughness.

For temperature less than or equal to 150^0C, the lower-bound fracture toughness is given by

$$K_c = 27 + 0.30T \quad MPa\sqrt{m} \tag{D.13-1}$$

and for temperature greater than 150^0 C, the lower-bound fracture toughness is given by

$$K_c = 72 \quad MPa\sqrt{m} \tag{D.13-2}$$

Where,

K_c = fracture toughness $MPa\sqrt{m}$, defined as the critical stress intensity factor at the onset of flaw instability,

T = temperature, (^{0}C)

2.1.1 Statistically based fracture toughness

For temperatures less than or equal to 150^0C, the relation for the statistical based fracture toughness is given by

$$K_{c1} = \exp(A_{kc1} + B_{kc2}T + \varepsilon_{kc1}) \quad MPa\sqrt{m} \tag{D.13-3}$$

Where,

K_{c1} = statistically based fracture toughness, $MPa\sqrt{m}$

$A_{kc\,1}$ = 3.762; B_{kc1} = 5.8 x 10^{-3}; T = Temperature, (^0C);

$$\varepsilon_{kc1} = t_{29} sd_{kc1}\left[1+\frac{1}{31}+\frac{(T-87.742)^2}{1.3653.9}\right]^{1/2}$$

t_{29} = Student's t-distribution with 29 degree of freedom

sd_{kc1} = 0.174

For temperature greater than 150^{0}C, the relation for the statistically based fracture toughness is given

$K_{c2} = exp\ (A_{kc2} + \varepsilon_{kc2})$

K_{c2} = statistically based fracture toughness, $MPa\sqrt{m}$

A_{kc2} = 4.6495

$$\varepsilon_{kc2} = t_{34} sd_{kc2}\left[1+\frac{1}{35}\right]^{1/2}$$

t_{34} = Student's t-distribution with 34 degree of freedom

sd_{kc2} = 0.1809

2.2 Data

2.2.1 Material

The majority of specimens initially tested were from sections of tubes removed after approximately 18 years of operation. These tubes were fabricated as standard cold-worked (~ 26%) Zr-2.5Nb pressure tubes prior to 1987, and it was specified that the ingots should be vacuum arc melted twice, as this process reduces some of the volatile impurity elements [6]. Some ingots, however, were produced using 100% recycled material, which is equivalent to the ingot being melted four times. The multiple melting of the material significantly reduced some of the volatile impurity elements (e.g. chlorine), which has a significant effect on the fracture toughness [5].

Cold-worked Zr-2.5Nb pressure tubes manufactured prior to 1987 were fabricated in accordance with the chemical specifications detailed in [6], which do not include any specific limits on the concentrations of impurity elements such as chlorine and phosphorus. Previous studies have demonstrated that these particular elements are among a few that have a significant effect on the deformation and fracture behaviour of pressure tube material [3, 5]. As a result, changes to the chemical composition specifications for Zr-2.5Nb pressure tubes were recommended [7] to improve the properties of newer tubes. Although the manufacturing route for pressure tubes has evolved with time, the overall changes to fabrication have not been substantial.

The chemical compositions of specimen used in the burst test program were taken from ingot analyses provided by the manufacturer, and Glow Discharge Mass Spectrography (GDMS) of offcuts (material removed from the ends of a pressure tube before installation in reactor). The measured values for the elements chlorine (Cl), carbon (C), oxygen (O), iron (Fe), and phosphorus (P) were used as part of the current study.

2.2.2 Burst Test Procedure and Analysis

The standardised procedure for conducting burst tests on irradiated Zr-2.5Nb pressure tube material was developed at AECL [3], and involves spark machining a through-wall axial crack of 55 mm length at the centre of a 0.5 m long section of tube. In addition, results from tests with non-standard crack lengths are included in this investigation (initial crack lengths $2a_0$ were in the range of 36.1 mm $\leq 2a_0 \leq$ 86.4 mm). The machined flaw is then sealed with a composite patch made of Teflon, stainless steel, and aluminium sheet that are secured to the pressure tube with silicone rubber. The test section is fitted with mechanical end caps before attachment to the pressurizing system, and enclosed in a protective bell-jar. The machined flaw in the specimen is extended approximately 5 mm axially in each direction by fatigue pressure cycling at room temperature using water and a maximum stress intensity of 15 MPa $m^{1/2}$. Stable crack growth is monitored using the direct current potential drop method. Once an experiment is to be conducted, the test section is heated to the desired test temperature using external heating coils and held for at least one hour. The test section is then pressurized with argon gas monotonically until failure. The Dugdale strip yield equation for an axial, through wall defect in a pressurized cylinder is used to calculate the Mode I stress intensity factor (K_c) as [3],

$$K_I = \left(\frac{8\sigma_f^2}{\pi} a \ln\left[\sec\left(\frac{\pi M \sigma_h}{2\sigma_f} \right) \right] \right)^{1/2} \tag{2.1}$$

where σ_f = flow stress (mean of the yield stress and ultimate tensile strength), $2a$ = total crack length, σ_h = hoop stress (pr_i / t), p = internal pressure, r_i = internal radius, t = wall thickness, and M = Folias bulging correction factor given approximately by [8]:

$$M = \{1 + 1.255[a^2 /(r_m t) - 0.0135[a^4 /(r_m t)^2]\}^{1/2} \tag{2.2}$$

for a given mean radius r_m. The resulting fracture toughness is expressed as the critical stress intensity factor K_c, corresponding to the stress intensity at the point of instability (rupture) calculated using the initial crack length $2a_0$ rather than the crack length at the point of instability ($2a_i$). As a result, K_c represents a conservative estimate of the fracture toughness. The CCL determined from K_c is also conservative provided that the pressure at rupture is less than the operating pressure.

2.3 Preliminary Analysis

The database for statistical analysis consists of fracture toughness *(K_c)* values obtained from 106 tests and values of 12 covariates for each test sample. A summary of variable affecting the fracture toughness is presented in Table 2-1. The average and standard deviation of K_c are estimated as 113.55 MPa√m and 32.83 MPa√m, respectively, and the coefficient of variation is 29 %.

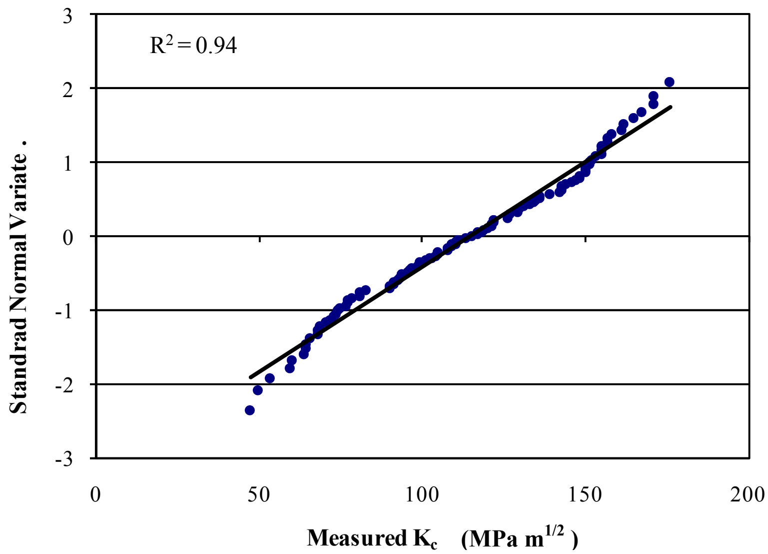

Figure 2-1 Normal probability plot of fracture toughness data

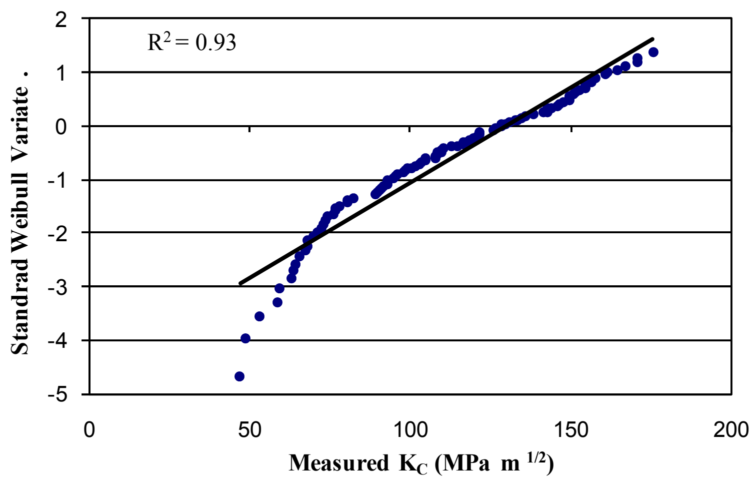

Figure 2-2 Weibull probability plot of fracture toughness data

The normal probability plot presented in Figure 2-1 shows that the normal distribution can model the test data reasonably well. From the fitted distribution, the 10% probability lower bound for K_c is estimated as 71.5 MPa√m. It should be remarked that the lower and upper tail regions of the empirical distribution (i.e., data) are not well represented by the normal distribution, and the use of other distributions for improving the goodness-of-fit will be explored in future work. Other distributions like Weibull and Gumbel are plotted in Figure 2-2 and Figure 2-3 respectively.

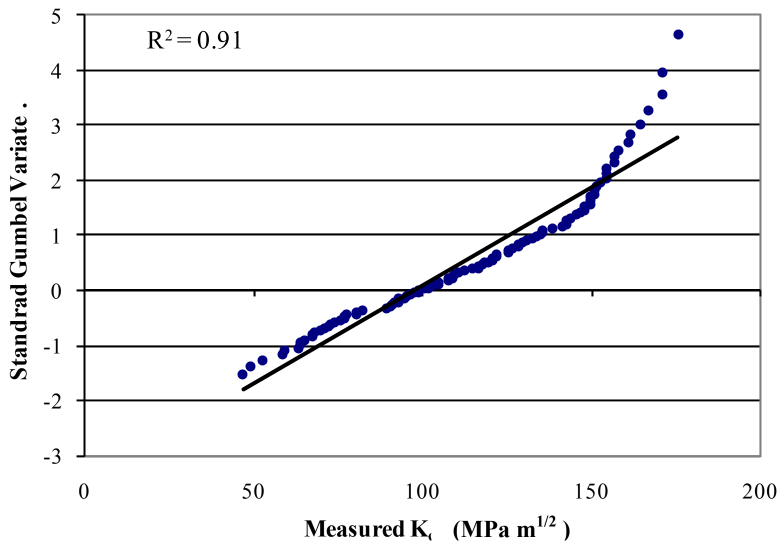

Figure 2-3 Gumbel probability plot of fracture toughness data

2.4 Concluding Remark

The normal, weibull and gumbel probability paper plots have R^2 values 0.94, 0.93 and 0.91 respectively. Since R^2 value is highest for normal distribution, so normal distribution is the best fit for the given data set. The traditional single-variate probabilistic analysis generally provides a conservative lower bound, especially when the data exhibit large variability. To improve the lower-bound estimate, a regression analysis is required as it can reduce the prediction variability by establishing a statistical relation with other random variables (covariates) that influence the fracture toughness.

Table 2-1: Variables affecting the fracture toughness

Variables	X_k	Name	Symbol	Mean (μ_k)	Standard Deviation (σ_k)	Units
Material Chemistry	X_1	Chlorine [C1]	Cl	4.31	3.77	Ppm
	X_2	Phosphorous[P]	P	25.03	17.78	Ppm
	X_3	Carbon [C]	C	159.95	25.97	Ppm
	X_4	Oxygen [O]	O	1133.51	95.49	Ppm
	X_5	Iron [Fe]	Fe	756.32	259.53	Ppm
Mechanical Property	X_6	Flow stress	FS	922.33	146.22	MPa
Operational Parameters	X_7	Irradiation Fluence	IRF	9.43	2.19	10^{25}n/m^2
	X_8	Irradiation Temperature	IRT	266.85	11.15	°C
	X_9	Test Temperature	TT	183.83	90.79	°C
Material Texture Parameters	X_{10}	Offcut Avg *Fr*[*]	OFR	0.32	0.024	
	X_{11}	Offcut Avg *Ft*[*]	OFT	0.63	0.027	
	X_{12}	Offcut Avg *Fl*[*]	OFL	0.049	0.0098	

[*] *Fr*, *Ft*, and *Fl* are measures of the fraction of grains with basal plane normal oriented in the radial, transverse, and longitudinal tube directions, respectively.

Chapter 3
Linear Regression Model

3.1 Introduction:

Modeling refers to the development of mathematical expressions that describe in some sense the behavior of a random variable of interest. This variable may be the fracture toughness of a pipe, thickness of a feeder, or the tensile strength of metal wire. In all cases, this variable is called the *dependent variable* and denoted with Y. Most commonly the modeling is aimed at describing how the mean of the dependent variable E[Y] changes with changing conditions; the variance of the dependent variable is assumed to be unaffected by the changing conditions.

Other variables which are thought to provide information on the behavior of the dependent variable are incorporated into the model as predictor or explanatory variables. These variables are called the *independent variables* and are denoted by X with subscripts as needed to identify different independent variables. In addition to the X's, all models involve unknown constants, called *parameters*, which control the behavior of the model.

The mathematical complexity of the model and the degree to which it is a realistic model depend on how much is known about the process being studied and on the purpose of the modeling exercise. In preliminary studies of a process or in cases where prediction is the primary objective, the models usually fall into the class of models that are *linear in the parameters*. That is, the parameters enter the model as simple coefficients on the independent variables or functions of the independent variables. Such models are referred as *linear models*. The more realistic models, on the other hand, are often *nonlinear in the parameters*. Most growth models, for example, are *nonlinear models*. Nonlinear models fall into two categories: the one, which can be linearized by an appropriate transformation on the dependent variable, and other are those that cannot be so transformed.

3.2 The Linear model and Assumptions

The simplest linear model involves only one independent variable and states that the true mean of the dependent variable changes at a constant rate as the value of the independent variable increases or decreases. Thus, the *functional relationship* between the true mean of Y_i, denoted by E $[Y_i]$ and X_i is the equation of a straight line:

$$\mathrm{E}[Y] = \beta_0 + \beta_1 X \tag{3.1}$$

β_0 is the intercept, the value of E $[Y_i]$ when X = 0, and β_1 is the slope of the line, the rate of change in $E[Y_i]$ per unit change in X. The observations on the dependent variable Y_i are assumed to be random observations from populations of random variables with the mean of each population given by E $[Y_i]$. The deviation of an observation Y_i from its population mean $E[Y_i]$ is taken into account by adding a random error ε_i to give the statistical model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{3.2}$$

The subscript *i* indicates the particular observational unit, $i = 1, 2, \ldots, n$. The X_i is the i^{th} observation on the independent variable and assumed to be measured without error. That is, the observed values of X are assumed to be a set of known values. The Y_i and X_i are paired observations; both are measured on every observational unit.

The random errors ε_i have zero mean and are assumed to have common variance σ^2 and to be pairwise independent. Since the only random element in the model is ε_i, these assumptions imply that the Y_i also have common variance σ^2 and are pairwise independent. For purposes of making tests of significance, the random errors are assumed to be normally distributed, which implies that the Y_i are also normally distributed. The random error assumptions are frequently stated as

$$\varepsilon_i \sim NID(0, \sigma^2) \tag{3.3}$$

where NID stands for “normally and independently distributed.” The quantities in parentheses denote the mean and the variance, respectively, of the normal distribution.

3.3 Least Square Estimation

The simple linear model has two parameters β_0 and β_1, which are to be estimated from the data. If there were no random error in Y_i, any two data points could be used to solve explicitly for the values of the parameters. The random variation in **Y**, however, causes each pair of observed data points to give different results. (All estimates would be identical only if the observed data fell exactly on the straight line.) A method is needed that will combine all the information to give one solution which is "best" by some criterion. The *least squares estimation procedure* uses the criterion that the least squares solution must give the smallest possible sum of squared deviations of the criterion observed Y_i from the estimates of their true means provided by the solution. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be numerical estimates of the parameters β_0 and β_1, respectively, and let

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \tag{3.4}$$

be the estimated mean of **Y** for each X_i, $i = 1,2, \ldots, n$. Note that $\hat{Y}_i$ is obtained by substituting the estimates for the parameters in the *functional form* of the model relating E (Y_i) to X_i, equation 3.1. The least squares principle chooses $\hat{\beta}_o$ *and* $\hat{\beta}_1$ be that minimize the sum of squares of the residuals, SS (Res):

$$\begin{aligned} SS(\mathrm{Re}\, s) &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n} e_i^2 \end{aligned} \tag{3.5}$$

Where $e_i = (Y_i - \hat{Y}_i)$ is the observed residual for the i^{th} observation. The summation indicated by Σ is over all observations in the data set as indicated by the index of summation, $i = 1$ to n.

The estimators for β_0 and β_1 are obtained by minimizing SS (Res). The derivatives of SS (Res) with respect to $\hat{\beta}$, $\hat{\beta}_1$ in turn are set equal to zero. This gives two equations in two unknowns called the normal equations:

$$n(\hat{\beta}_0) + (\sum X_i)\hat{\beta}_1 = \sum Y_i$$
$$(\sum X_i)\hat{\beta}_0 + (\sum X_i^2)\hat{\beta}_1 = \sum X_i Y_i \tag{3.6}$$

Solving the normal equations simultaneously for $\hat{\beta}_0$ and $\hat{\beta}_1$ gives the estimates of β_1 and β_0 as

$$\hat{\beta}_1 = \frac{\sum (X - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{x_i^2}$$
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{3.7}$$

Note that $x_i = (X_i - \overline{X})$ and $y_i = (Y_i - \overline{Y})$ denote observations expressed as deviations from their sample means X and Y respectively. These estimates of the parameters give the regression equation

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{3.8}$$

3.4 Accuracy of Estimates

3.4.1 Variance of $\hat{\beta}_1$

To determine the variance of $\hat{\beta}_1$ we express $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{3.9}$$

$$\hat{\beta}_1 = \left(\frac{x_1}{\sum x_i^2}\right) y_1 + \left(\frac{x_2}{\sum x_i^2}\right) y_2 + \left(\frac{x_n}{\sum x_i^2}\right) y_n \tag{3.10}$$

The coefficient on each Y_i is $\frac{x_i}{\sum x_i^2}$, which a constant in the regression model is. The Y_i are assumed to be independent and to have common variance σ^2. Thus, the variance of $\hat{\beta}_1$ is

$$Var(\hat{\beta}_1)=\left(\frac{x_1}{\sum x_i^2}\right)^2\sigma^2+\left(\frac{x_2}{\sum x_i^2}\right)^2\sigma^2+\left(\frac{x_n}{\sum x_i^2}\right)^2\sigma^2 \tag{3.11}$$

$$=\frac{\sum x_i^2}{\left(\sum x_i^2\right)}\sigma^2=\frac{\sigma^2}{\sum x_i^2} \tag{3.12}$$

3.4.2 Variance of $\hat{\beta}_0$

The variance of $\hat{\beta}_0$ can be derived using following formula

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{X}$$

The random variables in this linear function are $\overline{Y}$ and $\hat{\beta}_1$; the coefficients are 1 and (- $\overline{X}$). Equation 3.11 can be used to obtain variance of $\hat{\beta}_0$:

$$Var(\hat{\beta}_0)=Var(\overline{Y})+(-X)^2Var(\hat{\beta}_1)+2\left(-\overline{X}\right)Cov(\overline{Y},\hat{\beta}_1)$$

It can be easily shown that $Var(\overline{Y})=\frac{\sigma^2}{n}$ and $Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_i^2}$ and $Cov(\overline{Y},\hat{\beta}_1)=0$.

$$Var(\hat{\beta}_0)=Var(\overline{Y})+\left(\overline{X}\right)^2Var(\hat{\beta}_1)$$

$$=\frac{\sigma^2}{n}+\left(\overline{X}\right)^2\frac{\sigma^2}{\sum x_i^2}=\left(\frac{1}{n}+\frac{\left(\overline{X}\right)^2}{\sum x_i^2}\right)\sigma^2$$

3.5 Predicted Values and Residuals

Each quantity computed from the fitted regression line $\hat{Y}_i$ is used as both (1) the *estimate* of the population mean of **Y** for that particular value of Predictions **X** and (2) the *prediction* of the value of **Y** one might obtain on some future observation at that level of **X**. Hence, the $\hat{Y}_i$ are referred to both as *estimates* and as *predicted* values.

If the observed values Y_i in the data set are compared with their corresponding values $\hat{Y}_i$ computed from the regression equation, a measure of the degree of agreement between the model and the data is obtained. The residuals

$$e_i = Y_i - \hat{Y}_i \tag{3.13}$$

measure the discrepancy between the data and the fitted model.

3.6 Analysis of Variation in the Dependent Variable

The residuals are defined in equation 3.13 as the deviations of the observed values from the estimated values provided by the regression equation. Alternatively, each observed value of the dependent variable Y_i can be written as the sum of the estimated population mean of **Y** for the given value of **X** and the corresponding residual:

$$Y_i = \hat{Y}_i + e_i \tag{3.14}$$

$\hat{Y}$ is the part of the observation $\hat{Y}_i$ "accounted for" by the model, whereas e_i reflects the "unaccounted for" part. The **total uncorrected sum of squares** of Y_i, SS(Total $_{\text{uncorr}}$) = $\sum Y_i^{\ 2}$, can be similarly partitioned. Substitute $\hat{Y}_i + e_i$ for each Y_i and expand the square. Thus

$$\sum Y_i^2 = \sum (\hat{Y}_i + e_i)^2$$

$$
\begin{aligned}
&= \sum \hat{Y}_i^2 + \sum e_i^2 \\
&= ss(Model) + ss(\operatorname{Re} s)
\end{aligned}
\tag{3.15}
$$

(The cross-product term $\sum \hat{Y}_i e_i$ is zero. The term SS (Model) is the sum of squares "accounted for" by the model; SS (Res) is the "unaccounted for" part of the sum of squares. The forms SS (Model) = $\sum \hat{Y}_i^2$ and SS (Res) = $\sum e_i^{\,2}$ show the origins of these sums of squares. The more convenient computational forms are

$$
\begin{aligned}
&ss(Model) = n\bar{Y}^2 + \hat{\beta}_1^2 \sum (X_i - \bar{X})^2 \\
&ss(\operatorname{Re} s) = ss(Total_{uncorr}) - ss(Model)
\end{aligned}
\tag{3.16}
$$

The partitioning of the total uncorrected sum of squares can be re expressed in terms of the **corrected sum of squares** by subtracting the sum of squares due to correction for the mean, the correction factor $n\overline{Y}^2$, from each side of equation (3.15):

$$
ss(Total_{uncorr}) - n\bar{Y}^2 = [ss(Model) - n\bar{Y}^2] + ss(\operatorname{Re} s) \tag{3.17}
$$

or, using equation (3.16):

$$
\begin{aligned}
\sum y_i^2 &= \hat{\beta}_1 \sum (X_i - \bar{X})^2 + \sum e_i^2 \\
&= ss(\operatorname{Re} gr) + ss(\operatorname{Re} s)
\end{aligned}
\tag{3.18}
$$

The lower case y is the deviation of **Y** from $\bar{Y}$ so that $\sum y_i^2$ is the corrected total sum of squares. Henceforth, $ss(Total)$ is used to denote the corrected sum of squares of the dependent variable. $ss(Model)$ denotes the sum of squares attributable to the entire model, where as $ss(\operatorname{Re} s)$ denotes only that part of $ss(Model)$ that exceeds the correction factor. The correction factor is the sum of squares for a model that contains only the constant term β_0. Such a model postulates that the mean of

Y is a constant, or is unaffected by changes in **X**. Thus, $ss(\mathrm{Re}\,s)$ measures the additional information provided by the independent variable.

The degrees of freedom associated with each sum of squares is determined by the sample size *n* and the number of parameters *p'* in the model. [*p'* to denote the number of parameters in the model and *p* (without the prime) to denote the number of independent variables; $p' = p+1$ when the model includes an intercept as in equation 3.2.] The degrees of freedom associated with SS(Model) is $p' = 2$; the degrees of freedom associated with *SS(Regr)* is always 1 less to account for subtraction of the correction factor, which has 1 degree of freedom. *SS(Res)* will contain the $(n - p')$ degrees of freedom not accounted for by *SS(Model)*. The mean squares are found by dividing each sum of squares by its degrees of freedom. One measure of the contribution of the independent variable(s) in the model is the coefficient of determination, denoted by R^2.

$$R^2 = \frac{ss(\mathrm{Re}\,gr)}{\sum y_i^2} \tag{3.19}$$

$$\sum y_i^2 = \text{ the corrected total sum of squares.}$$

This is the proportion of the (corrected) sum of squares of **Y** attributable to the information obtained from the independent variable(s). The coefficient of determination ranges from zero to one and is the square of the product moment correlation between Y_i and $\hat{Y}_i$ if there is only one independent variable, it is also the square of the correlation coefficient between Y_i and X_i.

3.7 Tests of Significance and Confidence Intervals

The most common hypothesis of interest in simple linear regression is the hypothesis that the true value of the linear regression coefficient, the slope, is zero. This says that the dependent variable **Y** shows neither a linear increase nor decrease as the independent variable changes. In some cases, the nature of the problem will suggest other values for the null hypothesis. The computed regression coefficients, being random variables, will never exactly equal the hypothesized value. The role of the test of significance is to protect against being misled by the random variation in the estimates. Is the difference between the observed value of the parameter $\hat{\beta}_1$ and the hypothesized value of the parameter greater can be reasonably attributed to random variation? If so, the null hypothesis is rejected. To accommodate the more general case, the null hypothesis is written as $H_0 : \beta_1 = m$, where m is any constant of interest and of course can be equal to zero. The alternative hypothesis is

$H_a : \hat{\beta}_1 = m$, $H_a : \beta_1 > m$, or $H_a : \beta_1 < m$ depending on the expected behavior of β_1 if the null hypothesis is not true. In the first case, $H_a : \hat{\beta}_1 = m$ is referred to as the two-tailed alternative hypothesis (interest is in detecting departures of β_1 from m in either direction) and leads to a two-tailed test of significance. The latter two alternative hypotheses, $H_a : \beta_1 > m$ and $H_a : \beta_1 < m$, are one-tailed alternatives and lead to one-tailed tests of significance. If the random errors in the model, the ε_i, are normally distributed, the Y and any linear function of the Y will be normally distributed. Thus, $\hat{\beta}_1$ is normally distributed with mean β_1 ($\hat{\beta}_1$ is unbiased) and variance Var($\hat{\beta}_1$). If the null hypothesis that $\beta_1 = m$ is true, then $\hat{\beta} - m_1$ is normally distributed with mean zero. Thus,

$$t = \frac{\hat{\beta}_1 - \mathrm{m}}{s(\hat{\beta}_1)}$$

is distributed as Student's t with degrees of freedom determined by the degrees of freedom in the estimate of σ^2 in the denominator. The computed t-value is compared to the appropriate critical value of Student's t, determined by the Type I error α and whether the alternative hypothesis is one-tailed or two-tailed. The critical value of Student's t for the two-tailed alternative hypothesis places probability *α/2* in each tail of the distribution. The critical values for the one-tailed alternative hypotheses place probability α in only the upper or lower tail of the distribution, depending on whether the alternative is $\beta_1 > m$ or $\beta_1 < m$, respectively.

3.8 Effect of Texture Parameter

To address the effect of texture parameter a sample of 55 data points at 250^{0}C from low chlorine sample has been taken. Figure 3-1 and Figure 3-2 shows the dependency of fracture toughness with texture parameter.

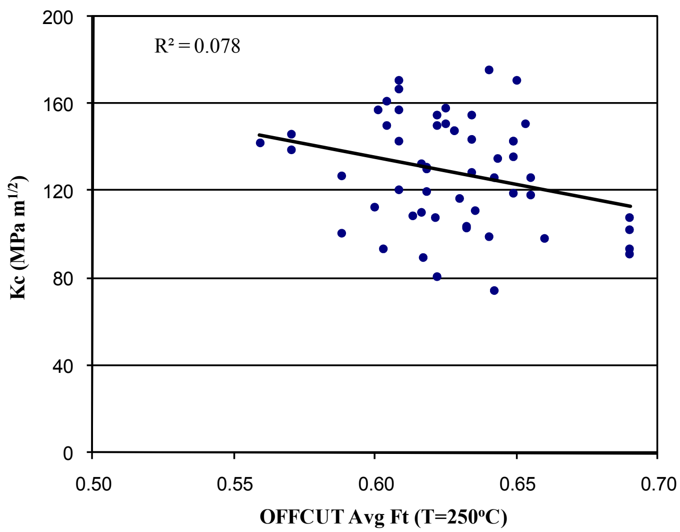

Figure 3-1 Effect of Texture parameter Ft on K_c

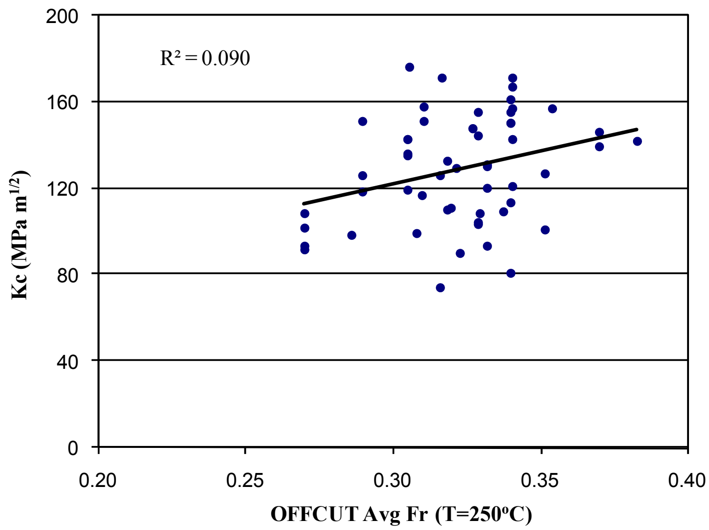

Figure 3-2 Effect of Texture parameter Fr on K_c

3.9 Case study: Fracture Toughness Vs Temperature

3.9.1 Temperature less than 150^{0}C

To find the equation for lower bound fracture toughness a sample of 43 data points (test temperature less than 1500^{0}C) has been taken from the database. The result of the simple regression between fracture toughness and Test temperature is given in Table 3-1. The R^2 for this model is 0.44 meaning 44% of the variation in the fracture toughness (dependent variable) is "explained' by its linear relationship with the Test Temperature (independent variable).

Table 3-1: Simple regression result Test Temperature taken as a variable

Covariate	Unstandardized Coefficients		Std Coff.	t value	sig	95% confidence interval for B	
	B	Std error				Lower Bound	Upper Bound
(Constant)	62	6.6		9.49	0.00	49	75
Test Temperature	0.36	0.06	0.66	5.64	0.00	0.23	0.49

Test of Significance: Using the two tailed alternative hypothesis and α = 0.05 gives a critical *t*-value of $t_{(0.05,2)} = 4.303$. Since $|t| > 4.303$, the conclusion is that data provide convincing evidence that $\hat{\beta}_1$ is different from zero.

For each data point 90% lower bound fracture toughness has been predicted with above model. A linear curve fit for these predicted value results in the following equation

$$K_c = 20 + 0.36T \tag{3.20}$$

The lower bound fracture toughness obtained in the equation (3.20) can be compared with the clause 13.2.2 CAN/CSA – N285.4-94 of

$$K_c = 27 + 0.30T \tag{D.13-1}$$

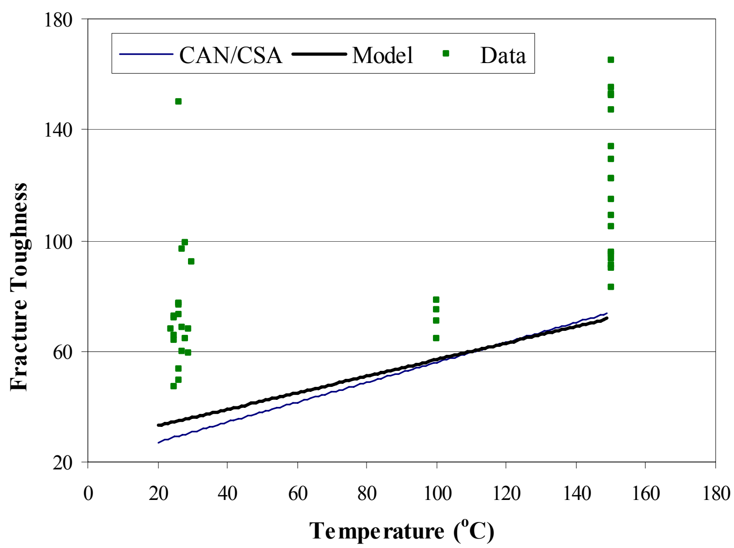

Figure 3-3 Comparison CAN/CSA Vs Model (T ≤ 150)

3.9.2 Temperature greater than 150^{0}C

For temperature greater than 150^0 C, a sample of 63 data points have taken from the database and a simple regression is performed. The result of regression is shown in Table 3-2 below. The R^2 for this model is 0.031.

Table 3-2 Simple regression result Test Temperature taken as a variable

Covariate	Unstandardized Coefficients		Std Coff.	*t* value	Sig	95% confidence interval for B	
	B	Std error				Lower Bound	Upper Bound
(Constant)	41	61		0.67	0.50	-81	164
Test Temperature	0.34	0.24	0.18	1.4	0.17	-0.15	0.83

Test of Significance: Using the two tailed alternative hypothesis and α = 0.05 gives a critical *t*-value of $t_{(0.05,2)} =$ 4.303. Since $|t| <$ 4.303, the conclusion is that the data do not provide convincing evidence that $\hat{\beta}_1$ is different from zero. For each data point 90% lower bound fracture toughness has been predicted with above model. A linear curve fit for these predicted value results in the following equation

$$K_c = 0.36T - 6 \tag{3.21}$$

The lower bound fracture toughness obtained in the equation (3.21) can be compared with the clause 13.2.2 CAN/CSA – N285.4-94.

$$K_c = 72 \tag{D.13-2}$$

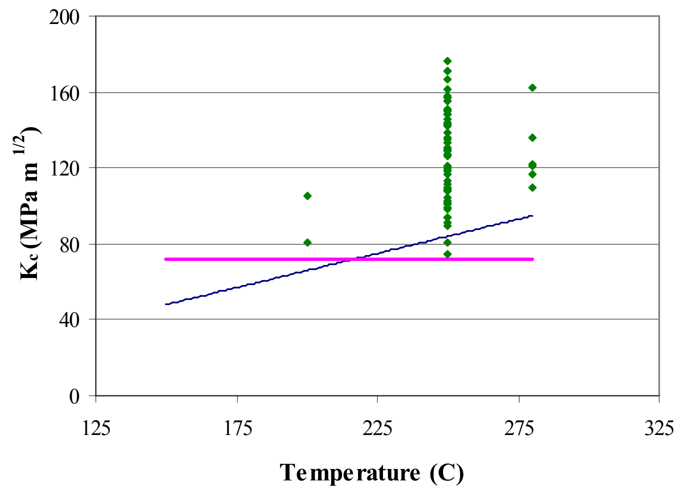

Figure 3-4 Comparison CAN/CSA Vs Model (T>150)

3.10 Concluding Remark

For temperature less than equal to 150^0 C, linear regression model shows a close resemblance with clause 13.2.2 CAN/CSA – N285.8-94.

For temperature greater than 150^0 C, linear regression model does not show resemblance with clause 13.2.2 CAN/CSA – N285.8-94. An increase in data point can improve the model.

Chapter 4
Linear Regression Model

4.1 The model

The linear additive model for relating a dependent variable to p independent variables is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_p X_{ip} + \varepsilon_i. \tag{4.1}$$

The subscript i denote the observational unit from which the observations on Y and the p independent variables were taken. The second subscript designates the independent variable. The sample size is denoted with n, $i = 1, \ldots, n$, and p denotes the number of independent variables. There are $(p + 1)$ parameters β_j $(j = 0, \ldots, p)$ to be estimated when the linear model includes the intercept β_0. For convenience, another parameter $p' = (p+1)$ is used. Four matrices are needed to express the linear model in matrix notation:

Y: The $n \times 1$ column vector of observations on the dependent variable Y_i;

X: The $n \times p'$ matrix consisting of a column of ones, which is labeled 1;

followed by the p column vectors of the observations on the independent variables;

$\boldsymbol{\beta}$**:** The $p' \times 1$ vector of parameters to be estimated; and

$\boldsymbol{\varepsilon}$**:** The $n \times 1$ vector of random errors.

With these definitions, the linear model can be written as

$$Y = X\beta + \varepsilon, \tag{4.2}$$

or,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \ldots & X_{1p} \\ 1 & X_{21} & X_{22} & X_{23} & \ldots & X_{2p} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{n1} & X_{n2} & X_{n3} & X_{n4} & Xnp \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix} \tag{4.3}$$

Each column of **X** contains the values for a particular independent variable. The elements of a particular row of **X**, say row *r*, are the coefficients on the corresponding parameters in $\boldsymbol{\beta}$ that give $E(Y_r)$. β_0 has the constant multiplier **1** for all observations; hence, the column vector **1** is the first column of **X**. Multiplying the first row of **X** by $\boldsymbol{\beta}$, and adding the first element of $\boldsymbol{\varepsilon}$ confirms that the model for the first observation is

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \ldots\ldots + \beta_p X_{1p} + \varepsilon_1 \tag{4.4}$$

The vectors **Y** and $\boldsymbol{\varepsilon}$ are random vectors; the elements of these vectors are random variables. The matrix **X** is considered to be a matrix of known constants.

The vector $\boldsymbol{\beta}$ is a vector of unknown constants to be estimated from the data. Each element β_j is a partial regression coefficient reflecting the change in the dependent variable per unit change in the j^{th} independent variable, *assuming all other independent variables are held constant*. The definition of each partial regression coefficient is dependent on the set of independent variables in the model. Whenever clarity demands, the subscript notation on β_j is expanded to identify explicitly both the independent variable to which the coefficient applies and the other independent variables in the model. For example, $\beta_{2.13}$ would designate the partial regression coefficient for X_2 in a model that contains X_1, X_2, and X_3.

The usual assumption about ε_i are now expressed in terms of the random vector $\boldsymbol{\varepsilon}$. $\boldsymbol{\varepsilon}$ is said to have multivariate normal distribution with mean vector **0** (of order *n* x 1). The variance of an individual element ε_i is replaced with the variance covariance matrix for any random vector of *n* elements is defined as *n* x *n* symmetric matrix with diagonal elements equal to the variance of the random variables (in order) and the $(i,j)^{th}$ off-diagonal element equal to the covariance between ε_i and ε_j. For example, if **Z** is a 3 x 1 vector of random variables z_1, z_2, z_3, the variance-covariance matrix of **Z** is the 3 x 3 matrix

$$Var(Z) = \begin{bmatrix} \sigma^2(z_1) & Cov(z_1, z_2) & Cov(z_1, z_3) \\ Cov(z_2, z_1) & \sigma^2(z_2) & Cov(z_2, z_3) \\ Cov(z_3, z_1) & Cov(z_3, z_2) & \sigma^2(z_3) \end{bmatrix} \tag{4.5}$$

The variance co-variance matrix for ε is $\boldsymbol{I\sigma^2}$ where **I** is the n x n identity matrix and σ^2 is the common variance of all ε_i. The distribution of ε can be written as

$$\varepsilon \sim \mathrm{N}(0, \; \boldsymbol{I\sigma^2}) \tag{4.6}$$

The statement that the variance-covariance matrix of ε, $\boldsymbol{Var(\varepsilon)}$, is $\boldsymbol{I\sigma^2}$ included two usual assumption that

1. The ε_i have common variance σ^2 ; and
2. They are statistically independent. (Independence is reflected in zero covariance)

Since the element of **X** and **β** are constants, the $\boldsymbol{X\beta}$ term in the model is a set of constant being added to the vector of random errors, $\boldsymbol{\varepsilon}$. Thus, **Y** is a random vector with mean vector $\boldsymbol{X\beta}$ and variance-covariance matrix $\boldsymbol{I\sigma^2}$:

$$E\,(\boldsymbol{Y}) = E(X\beta + \varepsilon) = E(X\beta) + E(\varepsilon) = X\beta \tag{4.7}$$

$$Var\,(\mathrm{Y}) = Var(X\beta + \varepsilon) = Var(\varepsilon) = I\sigma^2 \tag{4.8}$$

$Var\,(\mathrm{Y})$ is the same as $Var(\varepsilon)$ since during a constant to a random variable does not change the variance. When ε is normally distributed, **Y** is also multivariate normally distributed. Thus,

$$Y \sim N(X\beta, I\sigma^2) \tag{4.9}$$

This result is based on the assumption that the linear model being used in the correct model. If important independent variables have been omitted or if the functional form of the model is not correct, $\boldsymbol{X\beta}$ will not be the expectation of **Y**.

4.2 Normal Equations and Solutions

In matrix notation, the normal equations are written as

$$X'X\hat{\beta} = X'Y \tag{4.10}$$

The normal equations are always consistent and hence will always have a solution of the form

$$\hat{\beta} = (X'X)^{-1}(X'Y) \tag{4.11}$$

The multiplication $X'X$ generates a p' x p' matrix where the diagonal elements are the sums of squares of each of the independent variables and the off-diagonal elements are the sums of products between independent variables. The general form of $\mathbf{X'X}$ is

$$X'X = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ip} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{11}X_{ip} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum X_{ip} & \sum X_{i1}X_{ip} & \sum X_{i2}X_{ip} & \cdots & \sum X_{ip}^2 \end{bmatrix} \tag{4.12}$$

Summation in all cases is over $i = 1$ to n, the n observations in the data. When only one independent variable is involved, $\mathbf{X'X}$ consists of only the upper-left 2 × 2 matrix. Inspection of the normal equations in Chapter 3, equation (3.6), reveals that the elements in this 2 × 2 matrix are the coefficients on $\hat{\beta}_0$ *and* $\hat{\beta}_1$. The elements of the matrix product $\mathbf{X'Y}$ are the sums of products between each independent variable in turn and the dependent variable:

$$X'Y = \begin{pmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \\ \cdots \\ \sum X_{ip}Y_i \end{pmatrix} \tag{4.13}$$

The first element ΣY_i , is the sum of products between the vector of ones (the first column of **X**) and **Y**. Again, if only one independent variable is involved, $\mathbf{X'Y}$ consists of only the first two elements.

The unique solution to the normal equations exists only if the inverse of the first element ΣY_i is the sum of products between the vector of ones (the first column of X) and **Y** . Again, if only one independent variable is involved, $X'Y$ consists of only the first two elements. The unique solution to the normal equations exists only if the inverse of $X'X$ exists. This, in turn, requires that the matrix **X** be of full column rank; that is, there can be no linear dependencies among the independent variables. The practical implication is that there can be no redundancies in the information contained in **X**.

It is always possible to rewrite the model such that the redundancies among the independent variables are eliminated and the corresponding **X** matrix is of full rank. In this chapter, **X** is assumed to be of full column rank.

4.3 Correlation Analysis

The correlation of *Kc* with all 12 covariates is calculated using SPSS and the same is summarized in Figure 4-1.

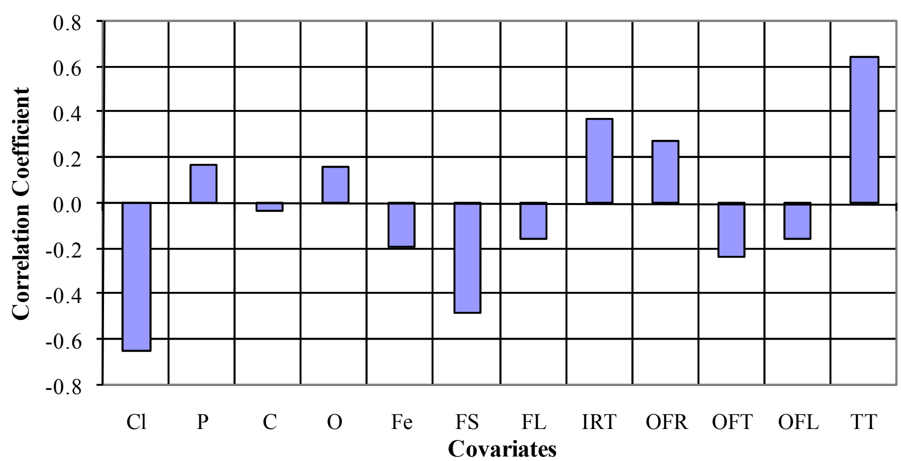

Figure 4-1 Correlation of Measured K_c with all 12 covariates

Note that the correlation coefficient between two random variables X_i and X_k is defined as $\rho_{ik} = E[(X_i - \mu_i)(X_k - \mu_k)]/\sigma_i\sigma_k$, where the operator E[.] denotes the mathematical expectation. Figure 4-1 shows that Kc has high negative correlation with chlorine (ρ = -0.65) and flow stress (ρ = -0.49) and high positive correlation with the test temperature (ρ = 0.65). These observations are consistent with physical reasoning. For example, material with high flow stress, which is an average of the yield and ultimate tensile stresses, is expected to have low fracture toughness. Similarly, the positive correlation of test temperature on *Kc* is well understood. *Kc* exhibits modest correlation ($0.2 \leq |\rho| \leq 0.5$) with irradiation temperature, iron and grain size parameters, Fr and Ft.

The correlation plot does not convey a complete picture since the covariates could be significantly correlated among themselves. In that case, more refined measures than the correlation coefficient are required to relate fracture toughness with them. This issue is investigated in the next section.

4.4 Correlation among Covariates

4.4.1 Pair-wise Correlation Analysis

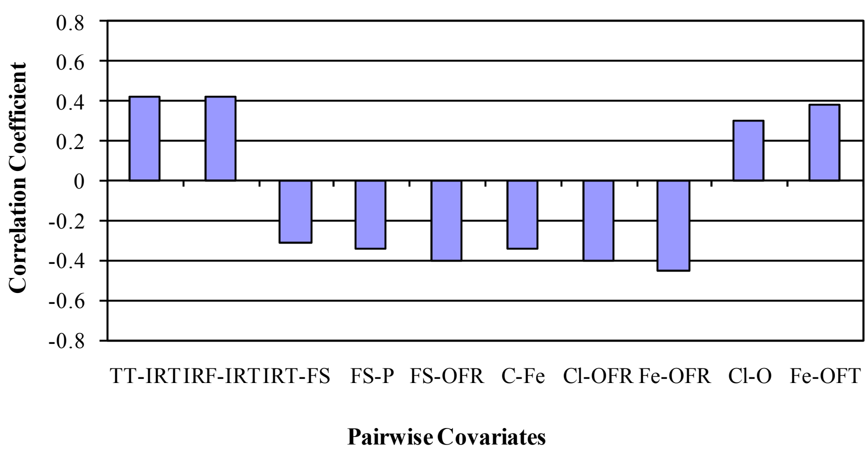

Figure 4-2 Covariates with modest correlation ($0.3 \leq |\rho| \leq 0.5$)

A 12×12 (Appendix A) correlation matrix for the covariates was calculated from the test data, and only significant results are summarized here. High negative correlation is found between two pairs of covariates: (1) flow stress and test temperature (ρ_{69} = -0.74), and (2) grain size parameters *Fr* and *Ft* ($\rho_{10,11}$ = -0.95). In Figure 4-2 ten pairs of covariates with modest correlation ($0.3 \leq |\rho| \leq 0.5$) are presented. The test temperature is correlated with several other covariates, and grain size parameters are correlated with chemical impurities.

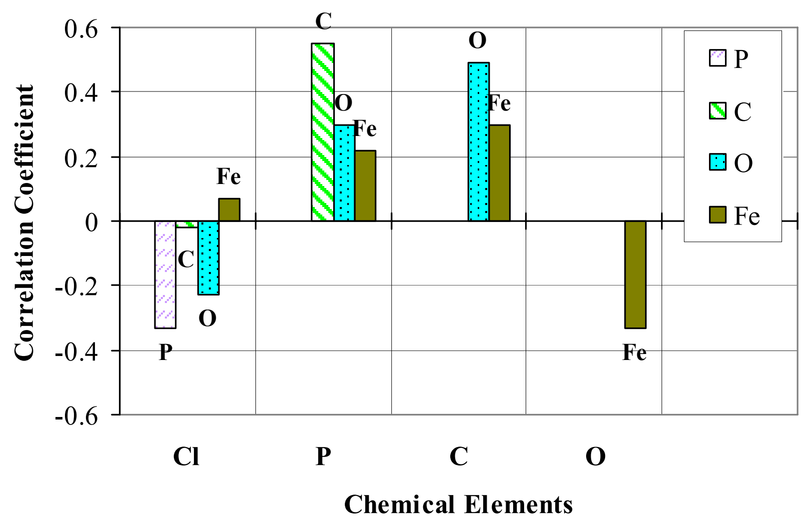

Figure 4-3 Correlation among chemical impurity elements

The correlation among chemical impurities is presented in Figure 4-3. Chlorine concentration is negatively correlated with phosphorus (ρ_{12} = -0.33) and oxygen (ρ_{14} = -0.23). Carbon has high correlation with phosphorus (ρ_{23} = 0.55) and oxygen (ρ_{34} = 0.49). Iron has modest correlation with phosphorus (ρ_{25} = 0.22), carbon (ρ_{35} = 0.3) and oxygen (ρ_{45} = -0.33). Because of this correlation structure, all chemical impurity elements may not be required in modeling the effect of material chemistry on the fracture toughness.

4.4.2 Stepwise Estimation: Selecting the first Variable

Table 6-1 (Appendix A), shows all the correlations among the twelve independent variables and their correlations with the dependent variable **(Y)**. Examination of the correlation matrix indicates that chlorine concentration [Cl] (X_1), and Test temperature (X_9) is reasonably high correlated with Fracture toughness. Fracture toughness is also moderately correlated with Flow stress (X_6), irradiation temperature (X_7) and OFFCUT Avg Fr (X_{10}). So, there are five variables (namely X_1, X_6, X_7, X_9, X_{10}) have high potential to predict the fracture toughness. Chlorine concentration [Cl] (X_1), and Test temperature (X_9) will account for 42% ($r^2 = 0.65^2$) each, of the variation in Fracture toughness, if used separately as the only independent variable in the regression. Also from Table 6-1 (Appendix A), it is clear that there is no correlation between X_1 and X_9, so both predictors can be included in the regression model. Since, the correlation is highest for chlorine concentration [Cl] and Test temperature, so in forward regression Test temperature **(X_9)** has been considered as a first entering variable. The result of regression using Test Temperature as only predicting variable is shown in Table 4-1.

Table 4-1 Regression result with variable Entered: (X_9) Test temperature

Multiple R	0.65
Multiple R^2	0.42
Adjusted R^2	0.42
Standard Error of estimate	25.0

Multiple R: Multiple R is the correlation coefficient for the simple regression of $\mathbf{X_9}$ and the dependent variable. It has no plus or minus sign because in multiple regression the sign of the individual variable may vary, so this coefficient reflects only the degree of association.

R square: R square (R^2) is the correlation coefficient squared, also referred to as the coefficient of determination. This value indicates that percentage of total variation of **Y** explained by $\mathbf{X_9}$. The total sum of squares (48111+65096=113208) is the squared error that would occur if we used only the mean of **Y** to predict the dependent variable. Using the value of $\mathbf{X_9}$ reduces this error by (48111/113208 = 42.1%).

Standard Error of the Estimate: The standard error of the estimate is another measure of the accuracy of our predictions. It is the square root of the sum of the squared errors divided by the degree of freedom. It represents an estimate of the standard deviation of the actual dependent values around the regression line; *i.e,* a measure of variation around the regression line. The standard error of the estimate can also be viewed as the standard deviation of the prediction errors and thus become measure to asses the absolute size of the confidence interval for the predictions

Variables in equation: A single predictor variable X_9 is used to calculate the regression equation for predicting the dependent variable. For each variable in the equation, several measures need to be defined: the regression coefficient, the standard error of the coefficient and the *'t'* value of variables in the equation.

- *Regression coefficient:* The value 0.24 in the regression coefficient (b_1) of the predictor variable (X_9). Thus the predicted value for each value of X_9 is the intercept plus the regression coefficient times the value of the predictor variable ($70.2 - 0.24*X_9$). The standardized regression coefficient, or beta value, of 0.65 is the value calculated from the standardized data. With only one independent variable, the squared beta coefficient equals the coefficient of determination. The beta values allow us to compare the effect of X_9 on Y to the effect on Y of other predictor variables at each stage, because this value reduces the regression coefficient to a comparable unit, the number of standard deviations. (At this stage we don't have any variable to compare).
- *Standard error of coefficient:* the standard error of the coefficient is the standard error of the estimate of b_9. The value of b_9 divided by the standard error ($-0.24/0.027 = -8.8$) is the calculated *t* value for a *t* test of the hypothesis $b_9 = 0$. A smaller standard error implies more reliable prediction. Thus it is good to have small standard errors and therefore smaller confidence interval. This coefficient is also referred to as the standard error of the regression coefficient; It is an estimate of how much the regression coefficient will vary between samples of the same size from the same population and use them to calculate the regression equation, this would be an estimate of how much the regression coefficient would vary frm sample to sample.

- *t value of variables in the equation:* the *t* value of variables in the equation measures the significance of the partial correlation of the variable reflected in the regression coefficient. It is useful to determine whether a variable should be dropped from the equation once a variable has been added. Also given in the table is the level of significance, which is compared to the threshold level for dropping the variable. In our case, we have set a 0.1 for dropping variables from the equation. The critical value for a significance level of 0.1 (two tailed) with 104 degree of freedom is 1.66. Therefore, X_9 meets our requirement for inclusion in the regression equation. F values are often given at this stage rather than *t* values. (*t* values is the square root of the F value)

Variables not in the equation: Although X_9 has been included in the regression equation, four other potential independent variables remain for inclusion to improve the prediction of the criterion variable. For those values, two measures are available to assess their potential contribution: partial correlations and *t* values.

- *Partial correlation.* The partial correlation is a measure of the variation in Y not accounted for by the variables in the equation (only X_9 in step 1) that can be accounted for by each of these additional variables. For example, in Table 4-2 the value 0.69 represents the partial correlation of X_1 given that X_9 is in the equation. Remember, the partial correlation can be misinterpreted. It does not mean that we explain 69.0 percent of the previously explained variance. It means that 47 percent (0.69^2 =47%, the partial coefficient of determination) of the unexplained (not the total) variance can now be accounted for by X_9. Because 42 percent was already explained by X_1, (1-0.42)*0.47=0.27 or 27 percent of the total variance could be explained by adding variable X_1.
- *t values of the variable not in the equation*. The column of *t* values measures the significance of the partial correlation for variables not in the equation. These are calculated as a ratio of

the additional sum of squares explained by including a particular variable and the sum of the squares explained by including a particular variable and the sum of the squares left after adding that same variable. If this *t* values does not exceed a specified significance level, the variable will not be allowed to enter the equation. The tabled *t* value for a significance level of 0.1 with 103 degree of freedom is 1.66. Looking at the column of *t* values, note that five variables (X_1, X_2, X_4, $X_{10,}$ and X_{11}) exceed this value and are candidates for inclusion. Simple correlation of X_1 with the dependent variable is 0.65, and the partial correlation is also largest among rest of the variable to be included. Therefore, X_1 would be included in the model next.

Table 4-2 Inclusion of first variable Test Temperature

Variables in the equation							Variables not in the equation	
Variables		Unstandardized Coefficients	Std. Error of coefficient	Standardized Coefficients (β)	t value	Sig.	Partial correlation	t value
(Constant)		70.2	5.5		12. 7	0.000		
Test Temp	X9	0.24	0.027	0.65	8.7	0.000		
[Cl]	X1						-0.69	-9.68
[C]	X3						0.078	0.79
[O]	X4						0.287	3.04
[Fe]	X5						-0.142	-1.45
Flow Stress	X6						-0.008	-0.08
Irrad Flu	X7						-0.130	-1.4
Irradiation Temp	X8						0.140	1.45
[P]	X2						0.234	2.44
OFFCUT Avg Fr	X10						0.251	2.63
OFFCUT Avg Ft	X11						-0.218	-2.26
OFFCUT Avg Fl	X12						-0.148	-1.51

4.4.3 Selecting the second Variable

The multiple R (from 0.65 to 0.83) and the R squared (from 0.42 to 0.70) values have both increased with additional X_1 inclusion. The increase in R^2 of 28 percent is derived by multiplying the 58 percent of variation that was not explained after step 1 by the partial correlation squared: $58*(0.69)^2 = 27.6$; that is, of the 50.9 percent unexplained with X_9, $(0.69)^2$ of this variance was explained by adding X_1, yielding a total variance explained of 0.70 that is, $0.42+0.58*(0.69)^2 = 0.70$

The value of unstandardized coefficient β has changed very little (0.24 to 0.19). This is a further clue that variables X_9 and X_1 are relatively independent (the simple correlation between two variables is 0.07). If the effect of X_1 on Y were totally independent of the effect of X_9, the β coefficient would not change at all.

The partial *t* values indicate that both X_9 and X_1 are statistically significant predictors of Y. The *t* value for the X_9 is now 9.3, where it was 8.7 in the previous step. The *t* value for X_1 examines the contribution of this variable given that X_9 is already in the equation.

Because predictors X_9 and X_1 both make significant contributions to the explanation of variation in the dependent variable, we can ask, are other predictors available? Looking at the partial correlation for variables not in the equation in Table 4-3, we see that X_8 (Irradiation temperature) has highest partial correlation (0.187).

4.4.4 Selecting the third Variable

With X_8 entered into the regression equation, the results are shown in the Table 4-4. As we predicted, the value of R^2 increases by 1.0 percent (0.70 to 0.71).

Table 4-3 Introduction of second variable Chlorine

Variables in the equation								Variables not in the equation	
Variables		Unstandardized Coefficients	Std. Error of coefficient	Standardized Coefficients (β)	*t* value	Sig.		Partial correlation	*t* Value
(Constant)		97	4.9		19.9	0.000			
Test Temperature	X_9	0.19	0.02	0.53	9.3	0.000			
[Cl]	X_1	-4.6	0.48	-0.53	-9.6	0.000			
[C]	X_3							0.052	0.531
[O]	X_4							0.155	1.589
[Fe]	X_5							-0.151	-1.54
Flow Stress	X_6							0.102	1.035
Irradiation Fluence	X_7							-0.176	-1.802
Irradiation Temperature	X_8							0.187	1.927
[P]	X_2							-0.008	-0.077
OFFCUT Avg Fr	X_{10}							-0.022	-0.22
OFFCUT Avg Ft	X_{11}							0.031	0.031
OFFCUT Avg Fl	X_{12}							-0.165	-0.165

In addition, examination of the partial correlation for rest of the variable not in the equation (most potential X_7 and X_4) indicates that no additional value will be gained by adding them to the predictive equation. These partial correlations are very small and have partial *t* values (for X_2, X_3, X_5, X_6, X_{10}, X_{11}, X_{12}) associated with them that would not be statistically significant at the level (0.1) chosen for this model.

Table 4-4 Selecting the third variable Irradiation Temperature

Variables in the equation								Variables not in the equation	
Variables		Unstandardized Coefficients	Std. Error of coefficient	Standardized Coefficients (β)	t value	Sig.		Partial correlation	t Value
(Constant)		11.9	44.8		0.26	0.79			
Test Temperature	X_9	0.18	0.022	0.49	8.2	0.000			
[Cl]	X_1	-4.6	0.47	-0.53	-9.7	0.000			
Irradiation Temperature	X_8	0.33	0.17	0.113	1.9	0.057			
[O]	X_4							0.186	1.9
[Fe]	X_5							-0.122	-1.23
Flow Stress	X_6							0.106	1.06
Irradiation Fluence	X_7							-0.259	-2.69
[C]	X_3							0.091	0.92
[P]	X_2							0.056	0.56
OFFCUT Avg Fr	X_{10}							0.032	-0.32
OFFCUT Avg Ft	X_{11}							0.031	0.313
OFFCUT Avg Fl	X_{12}							-0.153	-1.55

In developing the multivariate equation following **assumptions** have been considered.

Linearity: The first assumption linearity will be assessed through the analysis of residuals and partial regression plots. Figure 4-4 does not exhibit any nonlinear pattern to residual, thus ensuring that the overall equation is linear. But we must also be certain, when using more than one predictor variable, that each predictor variable's relationship is linear as well as to ensure its best representation in the equation. To do so, we use the partial regression plot for each predictor in the equation. Figure 4-5, Figure 4-6, Figure 4-7 shows the relationship for X_9 (Test Temperature), and X_1 (Chlorine Concentration) are quite well defined; thus they have strong and significant effect in the regression equation.

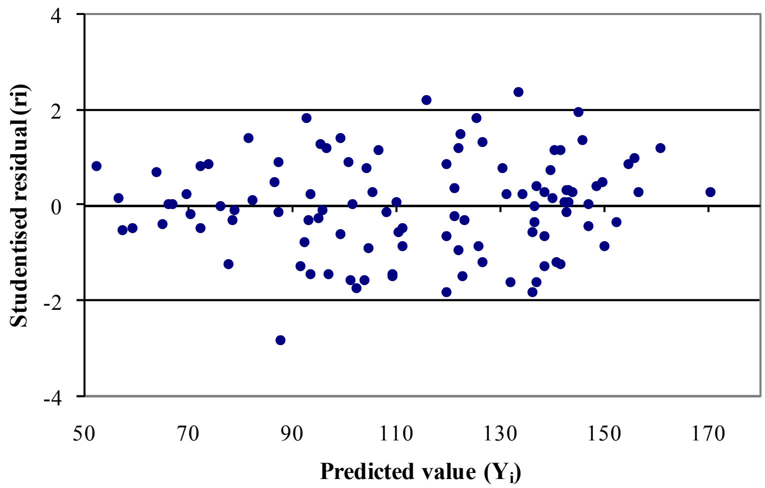

Figure 4-4 Analysis of studentized residuals

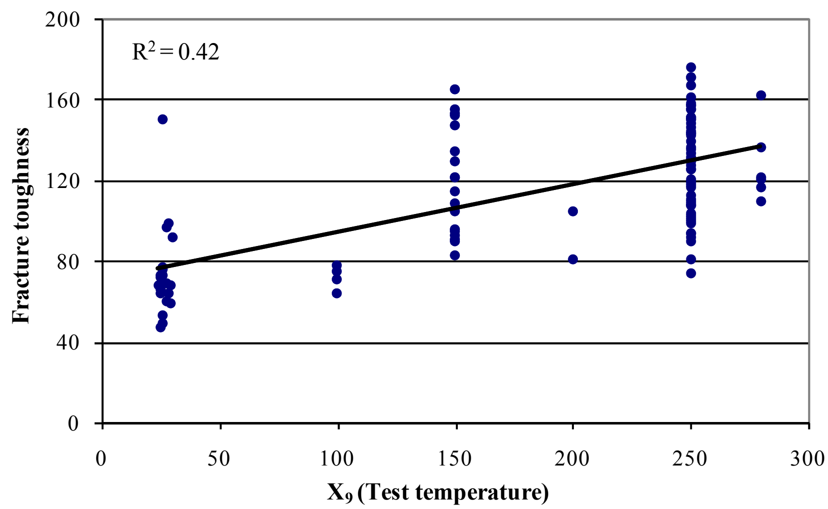

Figure 4-5 Partial regression plot, Test Temperature Vs Fracture Toughness

The variable X_8 (Irradiation Temperature) is less well defined, both in slope and scatter of the points, thus explaining its lesser effect in the equation (evidenced by the smaller coefficient (0.33), beta value (0.11), and significance level (0.06)). For all three variables, almost no non linear pattern is shown, thus meeting the assumption of linearity for each predictor variable.

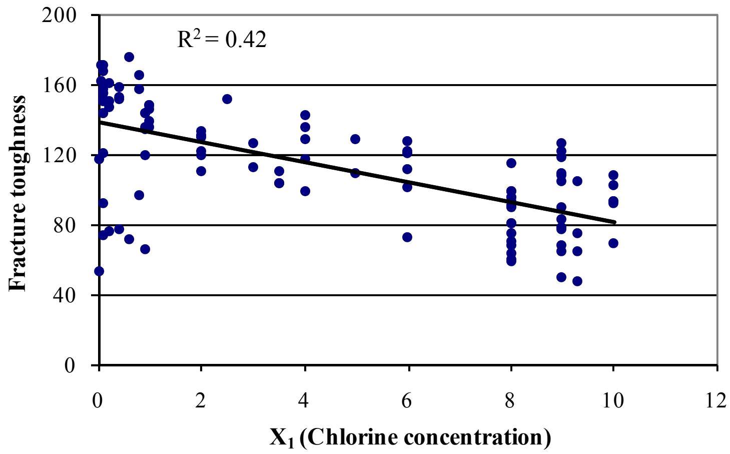

Figure 4-6 Partial regression plot, Chlorine Concentration Vs Fracture Toughness

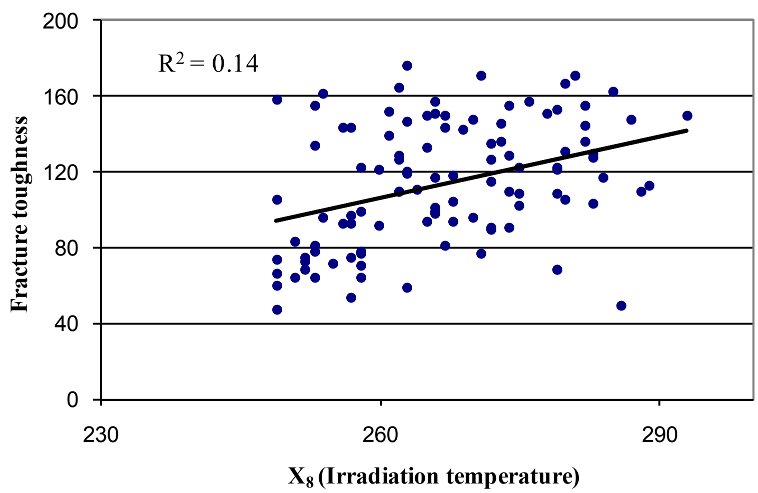

Figure 4-7 Partial regression plot, Irradiation Temperature Vs Fracture Toughness

Homoscedasticity: The next assumption deals with the constancy of the residuals across values of the predictor variables. Examination of the residuals Figure 4-4 shows no pattern of increasing or decreasing residuals. This finding indicates homoscedasticity in the multivariate case.

Identifying Outliers for assumption Violations: For our final analysis, we attempt to identify any observations, that are influential (having a disproportionate impact on the regression results) and determine whether they should be excluded from the analysis. The residual has been used for identifying outliers. Figure 4-9 shows the studentized residuals for each observation. Because the values correspond to *t* values, upper and lower limits can be set once the desired confidence interval has been established.

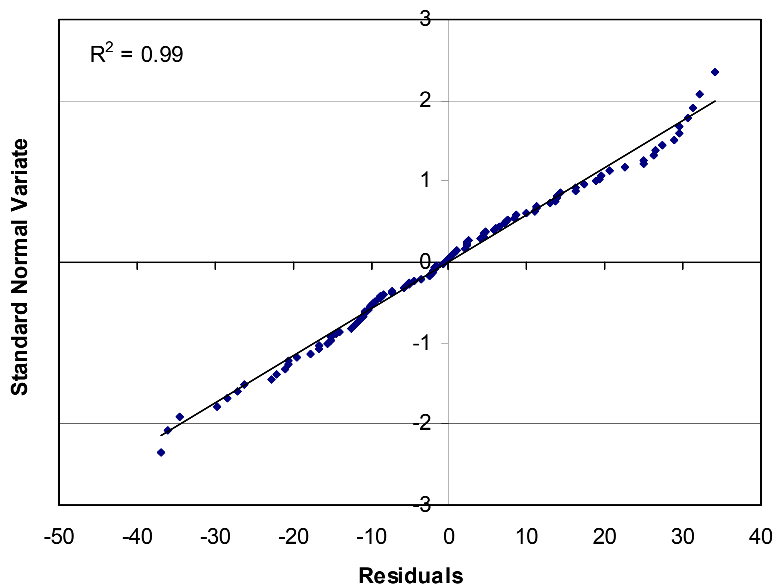

Figure 4-8 Normal probability paper plot of residuals for the model validation

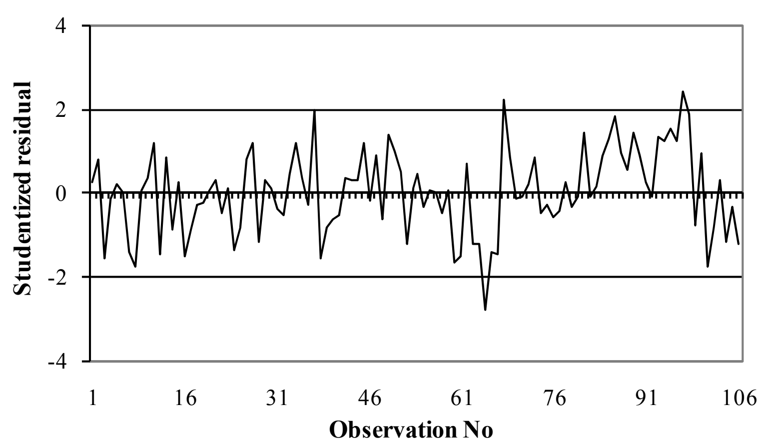

Figure 4-9 Plot of studentized residuals

For 95 percent confidence interval, the corresponding t value is 1.66, thus identifying statistically significant residuals as those with residuals greater than this value. Seven observation (5,37,60,65,97,98,101) have significant residuals and can be classified as outliers. Outliers are important because they are observations not represented by the regression equation for one or more reasons-one of which may be an influential effect on the equation that requires a remedy.

4.4.5 Predictive Model for samples with Low Chlorine Content

Current Zr-2.5Nb pressure tubes have a [Cl] < 0.2 ppm. As a result, a statistical model for predicting fracture toughness of pressure tubes with low chlorine content is required. From the database, a smaller sample of size = 28 data points having [Cl] $\leq$ 1 ppm was extracted for analysis. The number of samples used, from the database at a given test temperature, is outlined in Table 4-5.

Table 4-5 Test sample of low chlorine data

Test Temperature (°C)	Number of Samples
150	5
250	21
280	2
Total	28

From this sample, the average and standard deviation of Kc are estimated as 152.46 MPa√m and 10.64 MPa√m, respectively. The coefficient of variance of this sample is 6.97%. A regression model has developed using 7 variables including only ingot chemistry and operational conditions Table 4-7, resulting in the following model:

$$y = 31.878 - .367X_1 - 0.071X_2 + 0.056X_3 + 0.036X_4 - 1.1526X_5 + 0.141X_6 + 0.029X_7 \quad (4.14)$$

The regression results using these methods are summarized in Table 4-6. This model can explain 57% of the variability associated with the data. The standard error of the model with 20 degrees of freedom is 8.05 MPa√m, which is smaller than the standard deviation of the test data (10.64 MPa√m).

Table 4-6 Regression result for low chlorine data

Covariates	Regression Coefficients		t value	Significance (p value)	95% confidence intervals for b_k	
	b_k	Standard Error			Lower Bound	Upper Bound
(Constant)	31.878	60.201	.530	.602	-93.699	157.456
P (X_1)	-.367	.148	-2.473	.022	-.059	.118
C (X_2)	.071	.100	.711	.485	-4.141	1.090
O (X_3)	.056	.020	2.774	.012	-.274	.556
Fe (X_4)	.036	.013	2.730	.013	-.676	-.057
Irradiation Fluence (X_5)	-1.526	1.254	-1.217	.238	-.138	.280
Irradiation Temperature (X_6)	.141	.199	.709	.487	.014	.098
Test Temperature (X_7)	.029	.042	.696	.494	.008	.063

Table 4-7 Variable for low chlorine data

Variables		Name	Mean	Standard Deviation	Units
Ingot Chemistry	X_1	Phosphorous[P]	35.96	22.61	Ppm
	X_2	Carbon [C]	160.68	34.79	Ppm
	X_3	Oxygen [O]	1160.14	103.30	Ppm
	X_4	Iron [Fe]	743.04	214.39	Ppm
Operational Parameters	X_5	Irradiation Fluence	9.17	1.75	10^{25}n/m^2
	X_6	Irradiation Temperature	269.11	10.87	°C
	X_7	Test Temperature	234.29	40.77	°C

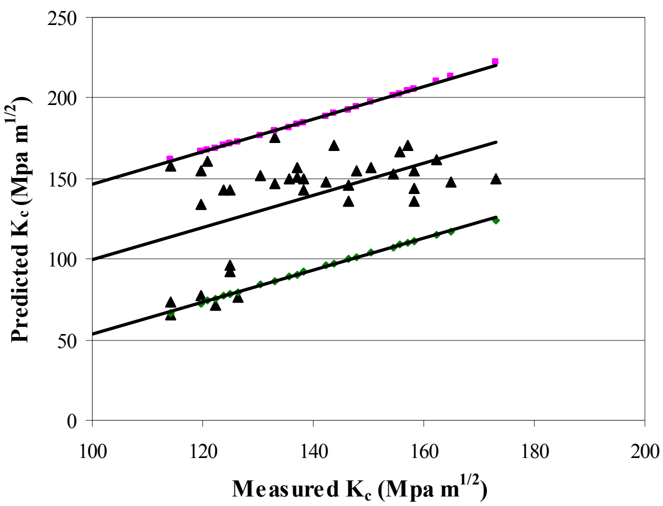

Figure 4-10 Regression model for low chlorine content data

A comparison of the observed values and predicted values using this model is shown in Figure 4-10 with 90% confidence interval, and the order of importance of the covariates is listed in the Figure 4-11. Here, it is observed that Phosphorous has the highest influence, followed by Irradiation Fluence and Iron.

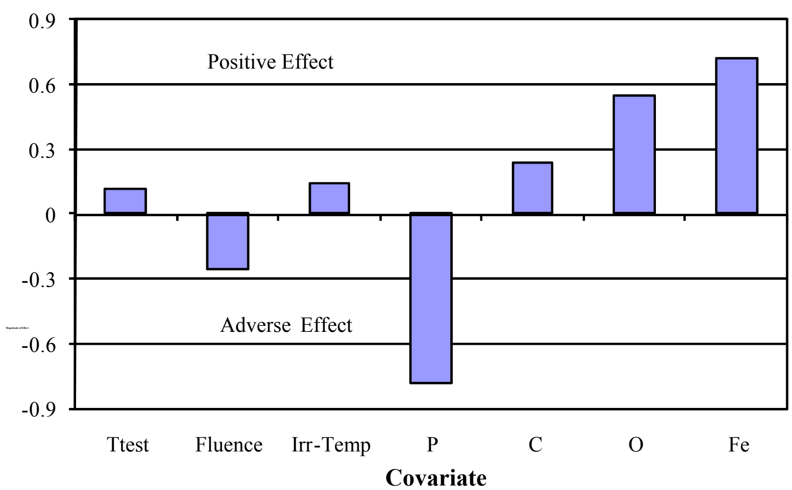

Figure 4-11 Effect of Covariates

4.5 Concluding Remark

Figure 4-1 shows that Kc has high negative correlation with chlorine ($\rho = -0.65$) and flow stress ($\rho = -0.49$) and high positive correlation with the test temperature ($\rho = 0.65$). Kc exhibits modest correlation ($0.2 \leq |\rho| \leq 0.5$) with irradiation temperature, iron and grain size parameters, Fr and Ft.

High negative correlation is found between two pairs of covariates: (1) flow stress and test temperature ($\rho_{6,\ 9} = -0.74$), and (2) grain size parameters Fr and Ft ($\rho_{10,11} = -0.95$). Chlorine concentration is negatively correlated with phosphorus ($\rho_{1,2} = -0.33$) and oxygen ($\rho_{1,4} = -0.23$). Carbon has high correlation with phosphorus ($\rho_{2,3} = 0.55$) and oxygen ($\rho_{3,4} = 0.49$). Iron has modest correlation with phosphorus ($\rho_{2,5} = 0.22$), carbon ($\rho_{3,5} = 0.3$) and oxygen ($\rho_{4,5} = -0.33$).

Test temperature (X_9) has been considered as a first entering variable. The result of regression using Test Temperature as only predicting variable is shown in Table 4-1. The standardized coefficient for Test temperature is 0.65 and R^2 value for model is 0.42.

For selecting second variable *t* value and largest correlation coefficient has been considered as criteria. Simple correlation of X_1 with the dependent variable Y is 0.65, and the partial correlation is also largest among rest of the variable to be included. Therefore, X_1 would be included in the model next. The multiple R (from 0.65 to 0.83) and the R squared (from 0.42 to 0.70) values have both increased with additional X_1 inclusion.

Looking at the partial correlation for variables not in the equation in Table 4-3, we see that X_8 (Irradiation temperature) has highest partial correlation coefficient (0.187) among rest of the variable to be included. With X_8 entered into the regression equation, the results are shown in the Table 4-4. The value of R^2 increases by 1.0 percent (0.70 to 0.71).

Current Zr-2.5Nb pressure tubes have a [Cl] < 0.2 ppm. So, a statistical model for predicting fracture toughness of pressure tubes with low chlorine content is required. From the database, a smaller sample of size = 28 data points having [Cl] $\leq$ 1 ppm was extracted for analysis. A regression model has developed using 7 variables including only ingot chemistry and operational conditions has been developed. The order of importance of the covariates is listed in the Figure 4-11. It is observed that Phosphorous has the highest influence, followed by Irradiation Fluence and Iron.

Chapter 5
Stepwise Regression

5.1 Stepwise Regression Methods

A stepwise regression method is a procedure by which the best model is developed in stages. A list of several potential explanatory variables is available and this list is repeatedly searched for variables which should be included in the model. The best explanatory variable is used first, then the second best, and so on. This procedure is known as stepwise regression. There are three procedures available to perform stepwise regression analysis [15].

5.1.1 Forward Selection

Forward stepwise selection of variables chooses the subset models by adding one variable at a time to the previously chosen subset. Forward selection starts by choosing as the one-variable subset the independent variable that accounts for the largest amount of variation in the dependent Variable. This will be the variable having the highest simple correlation with **Y**. At each successive step, the variable in the subset of variables not already in the model that causes the largest decrease in the residual sum of squares is added to the subset. Without a termination rule, forward selection continues until all variables are in the model.

5.1.2 Backward Elimination

Backward elimination of variables chooses the subset models by starting with the full model and then eliminating at each step the one variable whose deletion will cause the residual sum of squares to increase the least. This will be the variable in the current subset model that has the smallest partial sum of squares. Without a termination rule, backward elimination continues until the subset model contains only one variable.

5.1.3 Stepwise Selection

Neither forward selection nor backward elimination takes into account stepwise the effect that the addition or deletion of a variable can have on the contributions of other variables to the model. A variable added early to the model in forward selection can become unimportant after other variables are added, or variables previously dropped in backward elimination can become important after other variables are dropped from the model. The variable selection method commonly labeled is a forward selection process that rechecks at each step the importance of all previously included variables. If the partial sums of squares for any previously included variables do not meet a minimum criterion to stay in the model, the selection procedure changes to backward elimination and variables are dropped one at a time until all remaining variables meet the minimum criterion. Then, forward selection resumes. Stepwise selection of variables requires more computing than forward or backward selection but has an advantage in terms of the number of potential subset models checked before the model for each subset size is decided. It is reasonable to expect stepwise selection to have a greater chance of choosing the best subsets in the sample data, but selection of the best subset for each subset size is not guaranteed.

5.2 Stopping Rules

The computer programs for the stepwise selection methods generally include criteria for terminating the selection process. In forward selection, the common criterion is the ratio of the reduction in residual sum of squares caused by the next candidate variable to be considered to the residual mean square from the model including that variable. This criterion can be expressed in terms of a critical "F-to-enter" or in terms of a critical "significance level to enter" (SLE), where F is the "F-test" of the partial sum of squares of the variable being considered. The forward selection terminates when no variable outside the model meets the criterion to enter. This "F-test," and the ones to follow, should be viewed only as stopping rules rather than as classical tests of significance. The use of the data to select the most favorable variables creates biases that invalidate these ratios as tests of significance (Berk, 1978).

The stopping rule for backward elimination is the “F-test” of the smallest partial sum of squares of the variables remaining in the model. Again, this criterion can be stated in terms of an “F-to-stay” or as a “significance level to stay” (SLS). Backward elimination terminates when all variables remaining in the model meet the criterion to stay. The stopping rule for stepwise selection of variables uses both the forward and backward elimination criteria. The variable selection process terminates when all variables in the model meet the criterion to stay and no variables outside the model meet the criterion to enter (except, perhaps, for the variable that was just eliminated). The criterion for a variable to enter the model need not be the same as the criterion for the variable to stay. There is some advantage in using a more relaxed criterion for entry to force the selection process to consider a larger number of subsets of variables.

5.3 Model development fracture toughness data

The forward selection, backward elimination and stepwise selection methods of variable selection in SPSS are illustrated with the fracture toughness data. In this program, the termination rules are expressed in terms of significance level to enter, and significance level to stay. For this example, the criteria were set at SLE = 0.05 and SLS = 0.10 for all the three methods. These values are user defined. The regression results using these methods are summarized in Table 5-1,Table 5-2,Table 5-3 respectively.

Table 5-1 Backward Regression

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	Sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
1	(Constant)	-56.69	73.43		-0.77	0.44	-202.49	89.11
	Cl (X_1)	-4.55	0.59	-0.52	-7.69	0.00	-5.73	-3.38

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	Sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
	P (X_2)	-0.03	0.13	-0.02	-0.24	0.81	-0.30	0.24
	C (X_3)	0.10	0.11	0.08	0.91	0.36	-0.12	0.32
	O (X_4)	0.03	0.03	0.09	1.17	0.25	-0.02	0.09
	Fe (X_5)	-0.01	0.01	-0.06	-0.75	0.46	-0.03	0.01
	Flow Stress (X_6)	0.01	0.02	0.07	0.91	0.36	-0.02	0.05
	Fluence (X_7)	-2.65	0.89	-0.18	-2.96	0.00	-4.42	-0.87
	Irr-Temp (X_8)	0.54	0.19	0.18	2.80	0.01	0.16	0.93
	Test Temp (X9)	0.21	0.03	0.58	7.21	0.00	0.15	0.27
	OFFCUT Avg Fr (X_{10})	-51.59	90.43	-0.04	-0.57	0.57	-231.14	127.95
	OFFCUT Avg Ft (X_{11})	-253.03	197.37	-0.08	-1.28	0.20	-644.92	138.85
2	(Constant)	-92.20	52.48		-1.76	0.08	-196.32	11.91
	Cl (X_{11})	-4.29	0.46	-0.49	-9.22	0.00	-5.21	-3.36
	O (X_4)	0.05	0.02	0.15	2.69	0.01	0.01	0.09
	Fluence (X_7)	-2.87	0.87	-0.19	-3.31	0.00	-4.59	-1.15
	Irr-Temp (X_8)	0.60	0.18	0.20	3.36	0.00	0.24	0.95
	Test Temp (X9)	0.19	0.02	0.54	9.33	0.00	0.15	0.24

Table 5-2 Forward regression

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
1	(Constant)	70.22	5.51		12.75	0.00	59.29	81.14
	Test Temp (X_9)	0.24	0.03	0.65	8.77	0.00	0.18	0.29
2	(Constant)	97.82	4.91		19.90	0.00	88.07	107.57
	Test Temp (X_9)	0.19	0.02	0.54	9.75	0.00	0.16	0.23
	Cl (X_1)	-4.66	0.48	-0.54	-9.69	0.00	-5.61	-3.71

Table 5-3 Stepwise Regression

Model	Covariate	Unstandardized Coefficients		Std Coff.	t value	sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
1	(Constant)	70.22	5.51		12.75	0.00	59.29	81.14
	Test Temp (X_9)	0.24	0.03	0.65	8.77	0.00	0.18	0.29

Model	Covariate	Unstandardized Coefficients		Std Coff.	t value	sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
2	(Constant)	97.82	4.91		19.90	0.00	88.07	107.57
	Test Temp (X_9)	0.19	0.02	0.54	9.75	0.00	0.16	0.23
	Cl (X_1)	-4.66	0.48	-0.54	-9.69	0.00	-5.61	-3.71

5.4 Results

The backward elimination method results in best regression model with only 5 covariates, namely, chlorine (X_1), oxygen (X_4), irradiation fluence (X_7), irradiation temperature (X_8) and test temperature (X_9). The model equation is estimated as,

$$y = -92.203 - 4.287X_1 + 0.05X_4 - 2.869X_7 + 0.595X_8 + 0.194X_9 \tag{5.1}$$

The model statistics is summarized in Table ***5-4*** show that all the regression coefficients are significant at 5% level, except the intercept (b_0) for which the p value is 0.08.

Table 5-4 Regression results for the predictive model

Covariates	Regression Coefficients		t value	Significance (p value)	95% confidence intervals for b_k	
	b_k	Standard Error			Lower Bound	Upper Bound
(Constant)	-92.20	52.48	-1.76	0.08	-196.32	11.91
Cl (X_1)	-4.29	0.46	-9.22	0.00	-5.21	-3.36
O (X_4)	0.05	0.02	2.69	0.01	0.01	0.09
Fluence (X_7)	-2.87	0.87	-3.31	0.00	-4.59	-1.15
Irradiation Temperature (X_8)	0.60	0.18	3.36	0.00	0.24	0.95
Test Temperature (X_9)	0.19	0.02	9.33	0.00	0.15	0.24

For this model R^2 = 0.747, which means it explains 75% of the variability in the data, i.e., sum of squares of deviation from the mean of K_c data. The standard error of the model with 100 degrees of freedom is 16.90 MPa√m, which is approximately half of the standard deviation of the K_c (32.83 MPa√m). A comparison of the observed and predicted values presented in Figure 5-1 shows that 90%

prediction intervals enclose the sample data quite well. The adjusted R^2 of this model (=0.734) is slightly higher than that consisting 11 covariates (0.728), which indicates that elimination of variables has not resulted in loss of model predictability. Another way to check the suitability of any *m* covariate model in comparison to that with 11 covariates is to compute the Mallow statistic (C_m), given as [10]

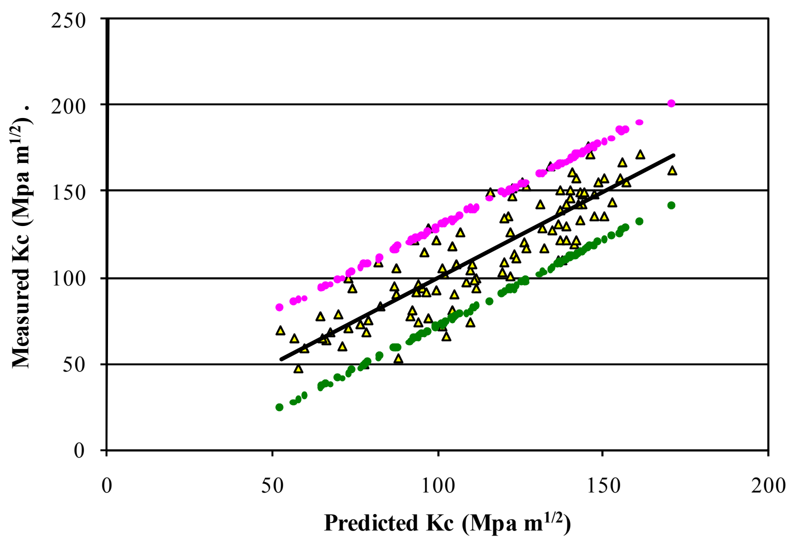

Figure 5-1 Regression model for K_c with 90% prediction interval

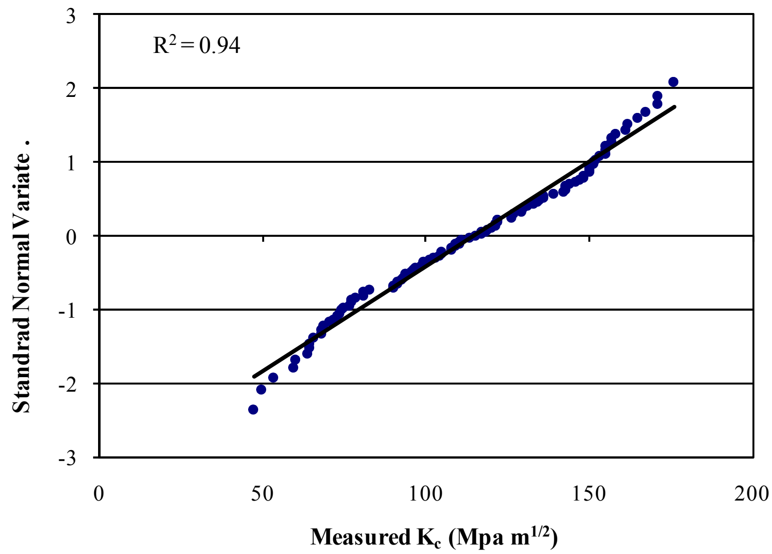

Figure 5-2 Residuals of the regression plotted on the normal probability paper

$$C_m = (n - m - 1)\frac{(s^2)_m}{(s^2)_{11}} + 2(m+1) - n \tag{5.2}$$

Cm < (m+1) signifies that adequacy of the reduced model. In case of the proposed model with m = 5, n = 106, $(s_2)_5$ = 16.90 and $(s_2)_{11}$ = 17.10, C_m = 4.8 (< 6), which also confirms its adequacy. The residuals, i.e., the differences between measured and predicted values, are plotted on a normal probability plot in Figure 5-2. It shows that the residuals can be fitted quite closely by a normal distribution, which validates the modelling assumption underlying the regression analysis.

5.5 Discussion

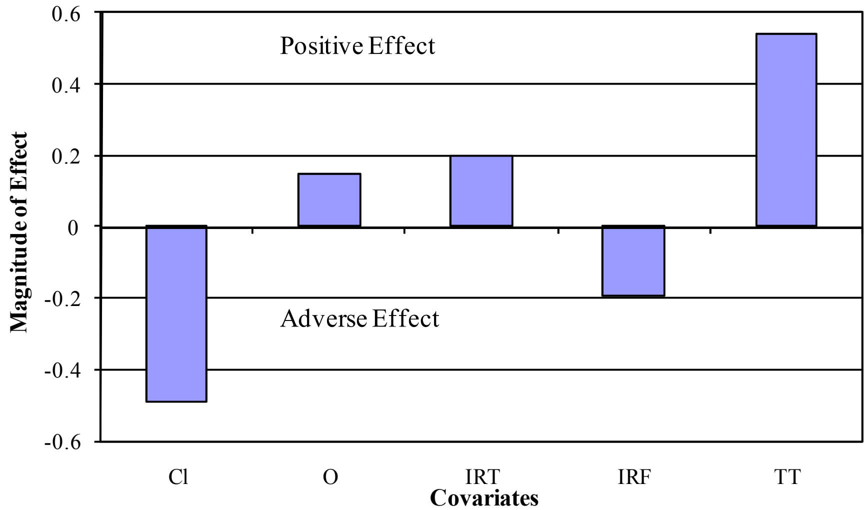

Figure 5-3: Effect of covariates on the fracture toughness

The importance and nature of the effect of a covariate on *Kc* can be investigated by examining the sign and magnitude of standardized regression coefficients, which are obtained by carrying out multivariate regression on standardized the covariates obtained as $(X_k - \mu_k)/\sigma_k$.The magnitudes of standardized regression coefficients are indicative of their relative importance in explaining the variability associated with *Kc* [10]. These standardized regression coefficients should be bounded within ±1, otherwise they are considered inconsistent due to large inter-correlation among covariates. The negative sign indicates that increasing the value of that covariate would decrease the toughness, and the reverse is implied by the positive sign.

Figure 5-3 shows that the test temperature has the highest influence (positive), followed by [Cl] which has an adverse effect on *Kc.* The effect of the irradiation temperature (positive) and irradiation fluence (adverse) is of similar order, and oxygen has a small positive effect.

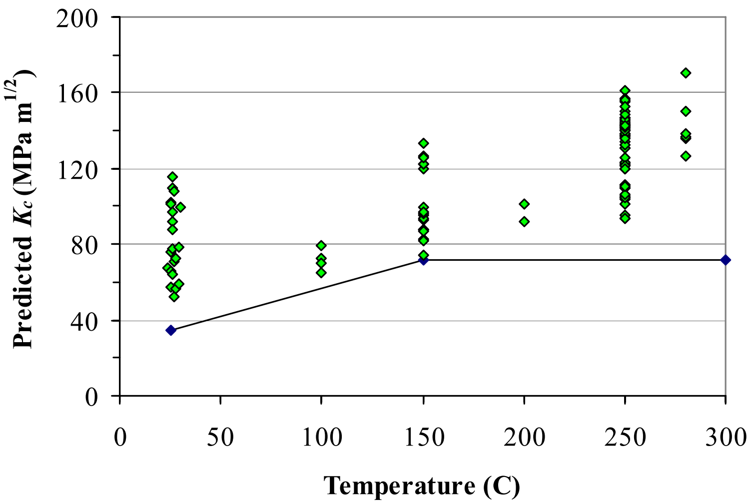

Figure 5-4: Comparison of predicted K_c with the lower bound

The lower bound estimate for *Kc* is specified in CSA N285.8 as a function of temperature (T) as [11]:

$(K_c)_{LB} = 27 + 0.3T$ for T ≤ 150 °C, and

$(K_c)_{LB} = 72$ for T > 150 °C (5.3)

This lower bound is compared with the predicted *Kc* values for 106 test samples in Figure 5-4. This plot shows that the regression model can provide a tube specific estimate of the fracture toughness which is expected to be more realistic than a generic lower bound curve. Recognizing that material chemistry and operational conditions can easily be obtained for all in-service pressure tubes, the proposed model (Eq.5.1) is amenable to use in fitness for service assessment of pressure tubes.

The predicted lower bound for *Kc* values has also been compared with statistically based fracture toughness and shown in Figure 5-5.

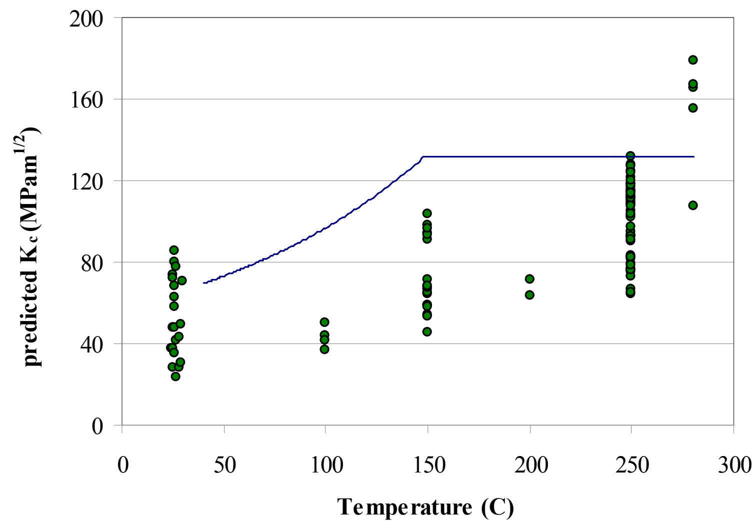

Figure 5-5: Comparison of predicted K_c with statistical lower bound

5.6 Model development for low Chlorine content data

A set of 36 data points has taken and forward selection, backward elimination and stepwise selection methods of variable selection has been performed using SPSS. The regression results using these methods are summarized in Table 5-5,Table 5-6 and Table 5-7 respectively.

Table 5-5 Backward Regression

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	Sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
1	(Constant)	31.88	60.20		0.53	0.60	-93.70	157.46
	Test Temp (X_9)	0.03	0.04	0.11	0.70	0.49	-0.06	0.12
	IRF (X_7)	-1.53	1.25	-0.25	-1.22	0.24	-4.14	1.09
	IRT (X_8)	0.14	0.20	0.14	0.71	0.49	-0.27	0.56
	P (X_2)	-0.37	0.15	-0.78	-2.47	0.02	-0.68	-0.06
	C (X_3)	0.07	0.10	0.23	0.71	0.49	-0.14	0.28
	O (X_4)	0.06	0.02	0.54	2.77	0.01	0.01	0.10
	Fe (X_5)	0.04	0.01	0.72	2.73	0.01	0.01	0.06
2	(Constant)	29.68	59.38		0.50	0.62	-93.80	153.16
	Test Temp (X_9)	-1.62	1.23	-0.27	-1.32	0.20	-4.18	0.94
	IRF (X_7)	0.18	0.19	0.18	0.93	0.36	-0.22	0.57
	IRT (X_8)	-0.37	0.15	-0.79	-2.54	0.02	-0.68	-0.07
	P (X_2)	0.07	0.10	0.22	0.69	0.50	-0.14	0.27

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	Sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
	C (X_3)	0.06	0.02	0.55	2.86	0.01	0.02	0.10
	O (X_4)	0.04	0.01	0.72	2.76	0.01	0.01	0.06
3	(Constant)	26.49	58.48		0.45	0.66	-94.79	147.77
	Test Temp (X_9)	-1.68	1.21	-0.28	-1.39	0.18	-4.20	0.83
	IRT (X_8)	0.21	0.18	0.21	1.15	0.26	-0.17	0.59
	P (X_2)	-0.31	0.12	-0.67	-2.66	0.01	-0.56	-0.07
	C (X_3)	0.06	0.02	0.55	2.88	0.01	0.02	0.10
	O (X_4)	0.04	0.01	0.83	4.28	0.00	0.02	0.06
4	(Constant)	88.91	21.65		4.11	0.00	44.11	133.70
	Test Temp (X_9)	-1.13	1.12	-0.18	-1.01	0.32	-3.44	1.19
	IRT (X_8)	-0.37	0.11	-0.79	-3.46	0.00	-0.59	-0.15
	C (X_3)	0.05	0.02	0.47	2.63	0.02	0.01	0.09
	O (X_4)	0.04	0.01	0.85	4.32	0.00	0.02	0.06

Table 5-6 Forward regression

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	Sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
1	(Constant)	138.39	6.95		19.92	0.000	124.1	152.7
	Fe (X_5)	0.02	0.01	0.38	2.11	0.000	0.0	0.0
2	(Constant)	135.33	6.08		22.27	0.000	122.8	147.8
	Fe (X_5)	0.04	0.01	0.745	3.83	0.000	0.0	0.1
	P (X_2)	-0.29	0.09	-0.612	-3.14	0.000	-0.5	-0.1
3	(Constant)	85.22	21.35		3.99	0.000	41.2	129.3
	Fe (X_5)	0.04	0.01	0.900	4.76	0.000	0.0	0.1
	P (X_2)	-0.41	0.10	-0.881	-4.20	0.000	-0.6	-0.2
	O (X_4)	0.04	0.02	0.410	2.43	0.002	0.0	0.1

Table 5-7 Stepwise Regression

Model	Covariate	Unstandardized Coefficients		Std Coff	t value	Sig	95% confidence interval for B	
		B	Std error				Lower Bound	Upper Bound
1	(Constant)	138.39	6.95		19.92	0.000	124.11	152.67
	Fe (X_5)	0.02	0.01	0.38	2.11	0.000	0.00	0.04
2	(Constant)	135.33	6.08		22.27	0.000	122.81	147.84
	Fe (X_5)	0.04	0.01	0.75	3.83	0.000	0.02	0.06
	P (X_2)	-0.29	0.09	-0.61	-3.14	0.000	-0.48	-0.10
3	(Constant)	85.22	21.35		3.99	0.000	41.16	129.28
	Fe (X_5)	0.04	0.01	0.90	4.76	0.000	0.03	0.06
	P (X_2)	-0.41	0.10	-0.88	-4.20	0.000	-0.62	-0.21
	O (X_4)	0.04	0.02	0.41	2.43	0.002	0.01	0.08

5.7 Results

The backward elimination method results in best regression model with only 7 covariates, namely, Phosphorous (X_2), Carbon (X_3), Oxygen (X_4), Iron (X_5), Irradiation fluence (X_7), Irradiation temperature (X_8) and Test Temperature (X_9). The model equation is estimated as,

$$y = 31.878 - .367X_2 - 0.071X_3 + 0.056X_4 + 0.036X_9 - 1.53X_7 + 0.141X_8 + 0.029X_9 \tag{5.2}$$

The model statistics is summarized in Table 5-8 shows that all the regression coefficients.

Table 5-8 Regression results for the predictive model

Covariates	Regression Coefficients		t value	Significance (p value)	95% confidence intervals for b_k	
	b_k	Standard Error			Lower Bound	Upper Bound
(Constant)	31.88	60.20	0.53	0.60	-93.70	157.46
Test Temp (X_9)	0.03	0.04	0.70	0.49	-0.06	0.12
IRF (X_7)	-1.53	1.25	-1.22	0.24	-4.14	1.09
IRT (X_8)	0.14	0.20	0.71	0.49	-0.27	0.56
P (X_2)	-0.37	0.15	-2.47	0.02	-0.68	-0.06
C (X_3)	0.07	0.10	0.71	0.49	-0.14	0.28
O (X_4)	0.06	0.02	2.77	0.01	0.01	0.10
Fe (X_5)	0.04	0.01	2.73	0.01	0.01	0.06

For this model R^2 = 0.76, which means it explains 76% of the variability in the data, i.e., sum of squares of deviation from the mean of K_c data. The standard error of the model with 20 degrees of freedom is 8.05 MPa√m, which is approximately less than the standard deviation of the K_c (10.64 MPa√m) for 28 data points. A comparison of the observed and predicted values presented in Figure 5-6 shows that 90% prediction intervals enclose the sample data quite well.

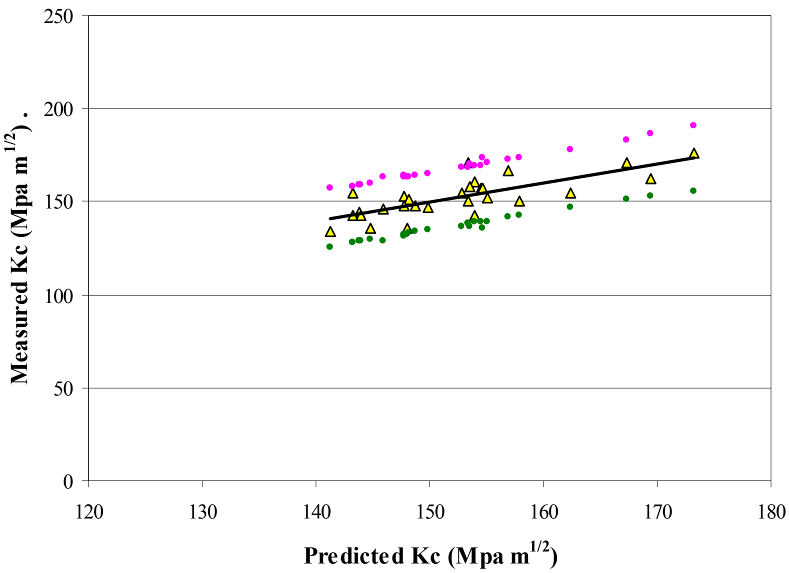

Figure 5-6 Regression model for K_c with 90% prediction interval

Figure 5-7 shows that the residuals fitted quite closely by a normal distribution, which validates the modelling assumption underlying the regression analysis.

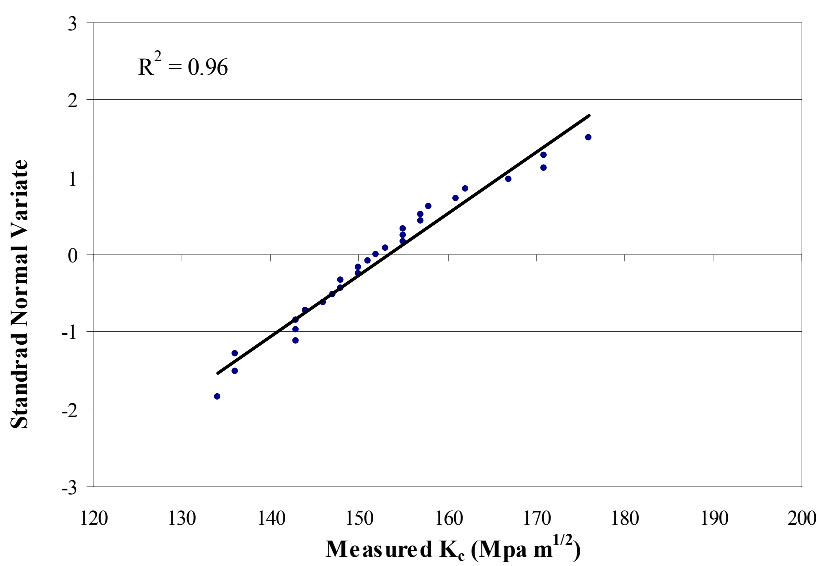

Figure 5-7 Residuals of the regression plotted on the normal probability paper

5.8 Discussion

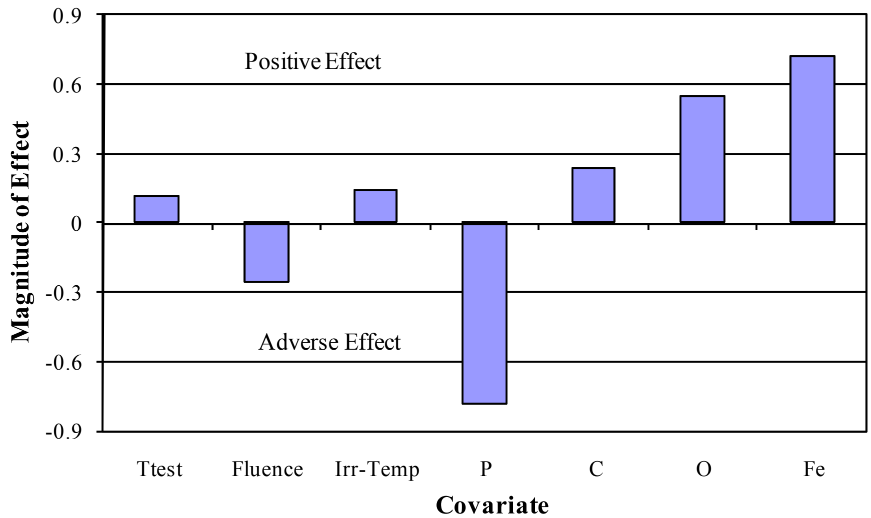

Figure 5-8: Effect of covariates on the fracture toughness

The importance and nature of the effect of a covariate on Kc can be investigated by examining the sign and magnitude of standardized regression coefficients.

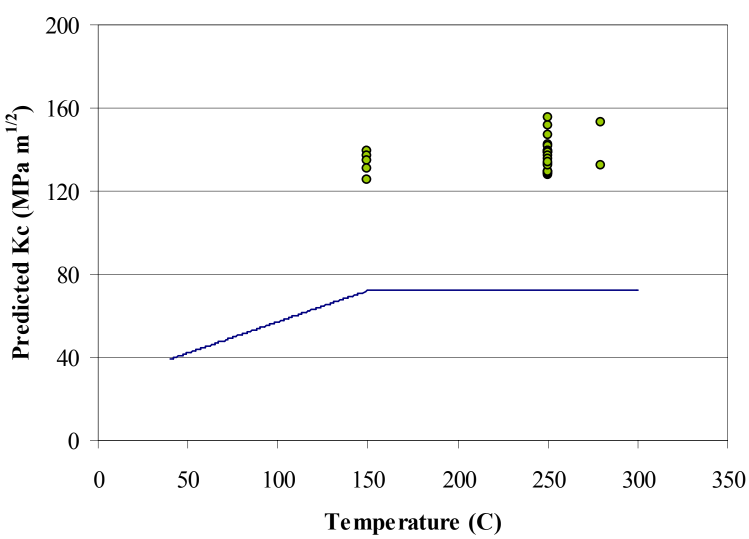

Figure 5-9: Comparison of predicted K_c with the lower bound

Figure 5-8 shows that the Phosphorous has the highest influence (negative), followed by [Fe] which has a positive effect on *Kc*. The effect of the [O] (positive), [C] (positive) Irradiation Temperature (positive) Test temperature (positive) and Irradiation Fluence (adverse) has a small effect.The lower bound obtained from code is compared with the predicted *Kc* values for 36 test samples in Figure 5-9. This plot shows that the regression model can provide a tube specific estimate of the fracture toughness which is expected to be more realistic than a generic lower bound curve.

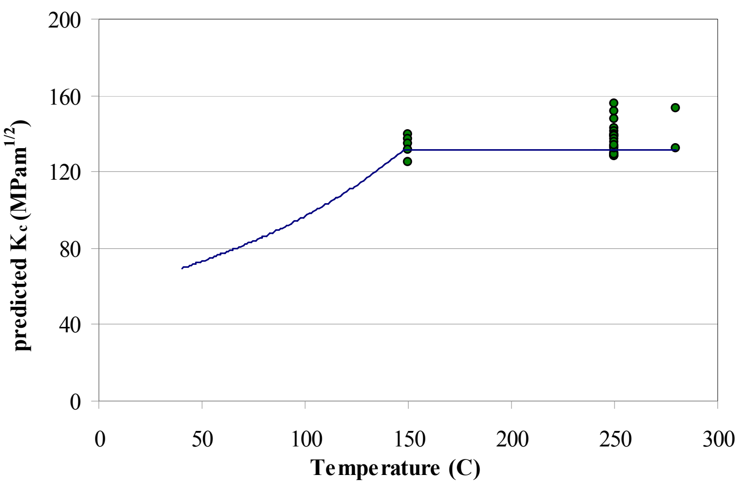

Figure 5-10: Comparison of predicted K_c with statistically based lower bound

The predicted lower bound for *Kc* values has also been compared with statistically based fracture toughness and shown in Figure 5-10.

5.9 Concluding Remark

The backward elimination method results in best regression model with only 5 covariates, namely, chlorine (X_1), oxygen (X_4), irradiation fluence (X_7), irradiation temperature (X_8) and test temperature (X_9). For this model $R^2 = 0.747$, The standard error of the model with 100 degrees of freedom is 16.90 MPa√m, which is approximately half of the standard deviation of the *Kc* (32.83 MPa√m). Test temperature has the highest influence (positive), followed by [Cl] which has an adverse effect on *Kc*.

The effect of the irradiation temperature (positive) and irradiation fluence (adverse) is of similar order, and oxygen has a small positive effect. The lower estimate for *Kc* is specified in CSA N285.8 is compared with the predicted *Kc* values for 106 test samples in Figure 5-4. This plot shows that the regression model can provide a tube specific estimate of the fracture toughness which is expected to be more realistic than a generic lower bound curve.

Chapter 6
Conclusions

6.1 Conclusions

The obtained data has been checked for best fit distribution. Normal, Weibull and Gumbel probability paper has been plotted and has R^2 value 0.94, 0.93 and 0.91 respectively. Finally normal distribution has chosen best fit for the given data set.

Next, dependence of temperature on fracture toughness has been analyzed. For temperature less than equal to 150^0 C, linear regression model shows a close resemblance with clause 13.2.2 CAN/CSA – N285.4-94. For temperature greater than 150^0 C, linear regression model does not show resemblance with clause 13.2.2 CAN/CSA – N285.4-94. An increase in data point can improve the model.

To develop the multivariate regression model correlation between dependent variable and independent variable has been considered. Figure 4-1 shows that *Kc* has high negative correlation with chlorine (ρ = -0.65) and flow stress (ρ = -0.49) and high positive correlation with the test temperature (ρ = 0.65). *Kc* exhibits modest correlation ($0.2 \leq |\rho| \leq 0.5$) with irradiation temperature, iron and grain size parameters, Fr and Ft. Also correlation between two dependent pair has been considered to control collinearity problem. A three variable regression model (Test Temperature, chlorine, and Irradiation Temperature) has been developed. This model explains 71% variability of observed data.

An analysis for low chlorine content and high temperature data has been also performed. From the database, a smaller sample of size = 28 data points having [Cl] $\leq$ 1 ppm was extracted for analysis. A regression model has developed using 7 variables including only ingot chemistry and operational conditions Table 4-7, resulting in the following model:

$$y = 31.878 - .367X_1 - 0.071X_2 + 0.056X_3 + 0.036X_4 - 1.1526X_5 + 0.141X_6 + 0.029X_7$$

This model explains 57% variability in data points.

The backward elimination method results in best regression model with only 5 covariates, namely, chlorine (X_1), oxygen (X_4), irradiation fluence (X_7), irradiation temperature (X_8) and test temperature (X_9). The model equation is estimated as,

$$y = -92.203 - 4.287X_1 + 0.05X_4 - 2.869X_7 + 0.595X_8 + 0.194X_9$$

The model statistics is summarized in Table *5-4* show that all the regression coefficients are significant at 5% level, except the intercept (b_0) for which the p value is 0.08. The statistical analysis presented in this thesis showed that a significant portion (~ 75%) of the burst test fracture toughness variability can be addressed. Since the covariates in the model are readily available for in-service pressure tubes (chlorine concentration, oxygen concentration, irradiation fluence, irradiation temperature, and operating temperature), it can be used to can be used to predict the fracture toughness of in-service pressure tubes, and define a suitable probabilistic lower-bound. The proposed approach and resulting model will improve the understanding of fracture toughness variability for in-service tubes, and may provide a basis for positive (less conservative) changes to guidelines for fitness-for-service assessment.

Appendix A

Correlation of Measured K_C with all 12 covariates

Table 6-1: Correlation Matrix

	Fracture Toughness	Test Temp	Irradiation Fluence	Irradiation Temp	Flow Stress	Cl	P	C	O	Fe	OFFCUT Avg Fr	OFFCUT Avg Ft	OFFCUT Avg Fl
Fracture Toughness	1	0.65	0.10	0.37	-0.48	-0.65	0.17	-0.03	0.16	-0.19	0.28	-0.23	-0.15
Test Temp	0.65	1.00	0.31	0.42	-0.74	-0.21	-0.02	-0.13	-0.09	-0.13	0.13	-0.10	-0.06
Irradiation Fluence	0.10	0.31	1.00	0.42	-0.23	-0.05	-0.01	0.04	0.12	-0.11	0.16	-0.12	0.06
Irradiation Temp	0.37	0.42	0.42	1.00	-0.31	-0.10	-0.27	-0.23	-0.16	-0.21	0.10	-0.04	-0.10
Flow Stress	-0.48	-0.74	-0.23	-0.31	1.00	0.23	0.07	0.15	0.07	0.11	-0.08	0.06	0.00
Cl	-0.65	-0.21	-0.05	-0.10	0.23	1.00	-0.34	-0.03	-0.23	0.07	-0.40	0.36	0.05
P	0.17	-0.02	-0.01	-0.27	0.07	-0.34	1.00	0.55	0.30	0.23	-0.03	-0.01	-0.02
C	-0.03	-0.13	0.04	-0.23	0.15	-0.03	0.55	1.00	0.50	0.30	-0.01	-0.12	0.34
O	0.16	-0.09	0.12	-0.16	0.07	-0.23	0.30	0.50	1.00	-0.34	0.26	-0.26	0.04
Fe	-0.19	-0.13	-0.11	-0.21	0.11	0.07	0.23	0.30	-0.34	1.00	-0.45	0.38	0.18
OFFCUT Avg Fr	0.28	0.13	0.16	0.10	-0.08	-0.40	-0.03	-0.01	0.26	-0.45	1.00	-0.95	0.12
OFFCUT Avg Ft	-0.23	-0.10	-0.12	-0.04	0.06	0.36	-0.01	-0.12	-0.26	0.38	-0.95	1.00	-0.39
OFFCUT Avg Fl	-0.15	-0.06	0.06	-0.10	0.00	0.05	-0.02	0.34	0.04	0.18	0.12	-0.39	1.00

Appendix B
Periodic inspection of CANDU Components
(CAN/CSA-N285.4-94)

Clause: D.13 Critical Crack length (CCL)

D.13.1 General

The length of a through-wall crack at which crack instability occurs is used in the flaw stability and leak-before-break (LBB) analysis procedures described in Clauses C.2 and C.4, respectively.

D.13.2 Fracture toughness for axial through-wall flaw

13.2.2 Lower-bound fracture toughness

For temperature less than or equal to 150^{0}C, the lower-bound fracture toughness is given by

$$K_c = 27 + 0.30T \quad MPa\sqrt{m} \tag{D.13-1}$$

and for temperature greater than 150^0 C, the lower-bound fracture toughness is given by

$$K_c = 72 \quad MPa\sqrt{m} \tag{D.13-2}$$

K_c = fracture toughness, defined as the critical stress intensity factor at the onset of flaw instability,

$MPa\sqrt{m}$ $\quad$ T = temperature (^{0}C)

13.2.3 Statistically based fracture toughness

Temperature less than or equal to 150^0 C.

Foe temperatures less than or equal to 150^{0}C, the relation for the statistically based fracture toughness is given by

$$K_{c1} = \exp(A_{kc1} + B_{kc2}T + \varepsilon_{kc1}) \tag{D.13-3}$$

K_{c1} = statistically based fracture toughness, $MPa\sqrt{m}$

A_{kc1} = 3.762 ; $\quad$ B_{kc1} = 5.8849 X 10^{-3} ; $\quad$ T = Temperature, ^{0}C

$$\varepsilon_{kc1} = t(d.o.f. = 29)sd_{kc1}\left[1 + \left(\frac{1}{31}\right) + \frac{(T - 87.742)^2}{1.3653.9}\right]^{1/2}$$

t (d.o.f.=29) = student's t-distribution with 29 degree of freedom

$sd_{kc1} = 0.174$

The value of $A_{kc1} = 4.14$, $B_{kc1} = 3.9 \times 10^{-3}$

To verify the above statistical equation 43 data point ($T \leq 150\ ^0C$) has been taken from a data set of 106 samples. A linear regression between ln(Kc) and temperature results in following model with R^2 value 0.47.

$$Kc = \exp(0.0039T + 4.14) \tag{B-1}$$

Figure 6-1 shows comparison between equation (D.13-3) and equation (B-1) with observed data points shown as dots.

Temperature greater than 150^0 C.

For temperature greater than 150^{0}C, the relation for the statistically based fracture toughness is given by $K_{c2} = exp\ (A_{kc2} + \varepsilon_{kc2})$

K_{c2} = statistically based fracture toughness, $MPa\sqrt{m}$

$A_{kc2} = 4.6495$

$$\varepsilon_{kc2} = t(d.o.f. = 34)sd_{kc2}\left[1 + \left(\frac{1}{35}\right)\right]^{1/2}$$

t (d.o.f. = 34) = student's t-distribution with 34 degree of freedom

$sd_{kc2} = 0.1809$

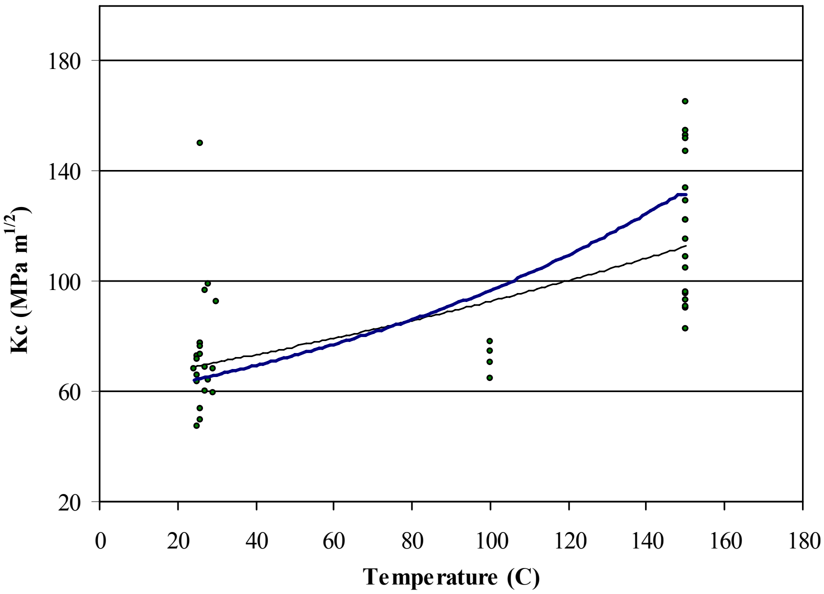

Figure 6-1: Comparison of predicted K_c with statistically based lower bound

References:

[1] Holt RA, Wong HW. Life management of CANDU fuel channels. Nuclear energy, 2002,41(1):69-74

[2] Langford, W.J. and Mooder, L.E.J., “Fracture Behaviour of Zirconium Alloy Pressure Tubes for Canadian Nuclear Power Reactors”, International Journal of Pressure Vessels and Piping, V.6, pp. 275-309, 1978.

[3] Davies, P.H. and R.S.W. Shewfelt, “Link Between Results of Small- and Large-Scale Toughness Tests on Irradiated Zr-2.5Nb Pressure Tube Material”. *Zirconium in the Nuclear Industry*: *Eleventh International Symposium, ASTM STP 1295,* E.R. Bradley and G.P. Sabol, Eds., American Society for Testing and Materials, pp.492-517, 1996.

[4] Chow, C.K. and Simpson, L.A., Determination of the fracture toughness of irradiated reactor pressure tubes using curved compact specimens. *Eighteenth Symposium on Fracture Mechanics, ASTM STP 945,* American Society for Testing and Materials, pp 419-439, 1988.

[5] Aitchison, I. and Davies, P.H., “Role of Microsegregation in Fracture of Cold-Worked Zr-2.5Nb Pressure Tubes”, Journal of Nuclear Materials, 203, pp.206-220, 1993.

[6] Cheadle, B.A, Coleman, C.E and Licht, H., “CANDU-PHW Pressure Tubes: Their Manufacture, Inspection, and Properties”, Nuclear Technology, Vol. 57, pp. 413-425, 1982.

[7] Theaker, J.R., Choubey, R.,Moan, G.D., Aldridge, S.A., Davis, L., Graham, R.A., and Coleman, C.E.” Fabrication of Zr-2.5nb Pressure Tubes to Minimise the Harmful Effects of Trace Elements”, *Zirconium in the Nuclear Industry*: *Tenth International Symposium, ASTM STP 1245,* A.M. Garde and E.R. Bradley, Eds., American Society for Testing and Materials, Philadelphia, pp 221-242, 1994.

[8] Kiefner, J.F., Maxey, W.A., Eiber, R.J. and Duffy, A.R., *Progress in Flaw Growth and Fracture Toughness Testing, ASTM STP 536*, American Society for Testing and Materials, West Conshohocken, PA, 1973, pp. 461-481.

[9] Chatterjee, S., and Price, B., *Regression Analysis by Examples*. John Wiley & Sons, New York, NY, 1977.

[10] McCuen, R.H. Statistical Hydrology. Prentice Hall, Englewood Cliffs, NJ, 1993.

[11] CSA N285.8-05. Technical requirements for in-service evaluation of zirconium alloy pressure tubes in CANDU reactors. The Canadian Standards Association, 2005.

[12] Joshef F. Hair, JR., Rolph E. Anderson, Ronald L. Tatham, and William C. Black, *Multivariate Data Analysis with readings.* Prentice Hall, Englewood Cliff, NJ, 1995

[13] Draper N.R. and Smith H. *Applied Regression Analysis.* John Wiley & Sons, New York, NY, 1966.

[14] Miller Irwin and John E. Freund *probability and Statistics for Engineers.* Prentice Hall, Englewood Cliff, NJ, 1977.

[15] John O. Rawlings, Sastry G. Pantula and David A. Dickey *Applied regression analysis: A Research Tool.*Springer-Verlag NY, 1998.

[16] Pandey, M.D., Radford D. D. (2007). A statistical approach to the Prediction of Pressure Tube Fracture Toughness Annual Canadian Nuclear Society Conference, 2007.