STATISTICAL MODELLING OF ECDA DATA FOR THE PRIORITISATION OF DEFECTS ON BURIED PIPELINES

By

Nik Nooruhafidzi Bin Muhd Noor

A thesis submitted to Brunel University London for the degree of

Doctor of Philosophy

College of Engineering, Design and Physical Sciences

Department of Mechanical Engineering

Brunel University London

December 2017

Abstract

Buried pipelines are vulnerable to the threat of corrosion. Hence, they are normally coated with a protective coating to isolate the metal substrate from the surrounding environment with the addition of CP current being applied to the pipeline surface to halt any corrosion activity that might be taking place. With time, this barrier will deteriorate which could potentially lead to corrosion of the pipe.

The External Corrosion Direct Assessment (ECDA) methodology was developed with the intention of upholding the structural integrity of pipelines. Above ground indirect inspection techniques such as the DCVG which is an essential part of an ECDA, is commonly used to determine coating defect locations and measure the defect's severity. This is followed by excavation of the identified location for further examination on the extent of pipeline damage. Any coating or corrosion defect found at this stage is repaired and remediated. The location of such excavations is determined by the measurements obtained from the DCVG examination in the form of %IR and subjective inputs from experts which bases their justification on the environment and the physical characteristics of the pipeline.

Whilst this seems to be a straight forward process, the factors that comes into play which gave rise to the initial %IR is not fully understood. The lack of understanding with the additional subjective inputs from the assessors has led to unnecessary excavations being conducted which has put tremendous financial strain on pipeline operators. Additionally, the threat of undiscovered defects due to the erroneous nature of the current method has the potential to severely compromise the pipeline's safe continual operation.

Accurately predicting the coating defect size (TCDA) and interpretation of the indication signal (%IR) from an ECDA is important for pipeline operators to promote safety while keeping operating cost at a minimum. Furthermore, with better estimates, the uncertainty from the DCVG indication is reduced and the decisions made on the locations of excavation is better informed. However, ensuring the accuracy of these estimates does not come without challenges. These challenges include (1) the need of proper methods for large data analysis from indirect assessment and (2) uncertainty about the probability distribution of quantities. Standard mean regression models e.g. the OLS, were used but fail to take the skewness of the distributions involved into account.

The aim of this thesis is thus, to come up with statistical models to better predict TCDA and to interpret the %IR from the indirect assessment of an ECDA more precisely. The pipeline data used for the analyses is based on a recent ECDA project conducted by TWI Ltd. for the Middle Eastern Oil Company (MEOC).

To address the challenges highlighted above, Quantile Regression (QR) was used to comprehensively characterise the underlying distribution of the dependent variable. This can be effective for example, when determining the different effect of contributing variables towards different sizes of TCDA (different quantiles). Another useful advantage is that the technique is robust to outliers due to its reliance on absolute errors. With the traditional mean regression, the effect of contributing

variables towards other quantiles of the dependent variable is ignored. Furthermore, the OLS involves the squaring of errors which makes it less robust to outliers. Other forms of QR such as the Bayesian Quantile Regression (BQR) which has the advantage of supplementing future inspection projects with prior data and the Logistic Quantile Regression (LQR) which ensures the prediction of the dependent variable is within its specified bounds was applied to the MEOC dataset.

The novelty of research lies in the approaches (methods) taken by the author in producing the models highlighted above. The summary of such novelty includes:

- The use of non-linear Quantile Regression (QR) with interacting variables for TCDA prediction.
- The application of a regularisation procedure (LASSO) for the generalisation of the TCDA prediction model.
- The usage of the Bayesian Quantile Regression (BQR) technique to estimate the %IR and TCDA.
- The use of Logistic Regression as a guideline towards the probability of excavation
- And finally, the use of Logistic Quantile Regression (LQR) in ensuring the predicted values are within bounds for the prediction of the %IR and POPD.

Novel findings from this thesis includes:

- Some degree of relationship between the DCVG technique (%IR readings) and corrosion dimension. The results of the relationship between TCDA and POPD highlights a negative trend which further supports the idea that %IR has some relation to corrosion.
- Based on the findings from Chapter 4, 5 and 6 suggests that corrosion activity rate is more prominent than the growth of TCDA at its median depth. It is therefore suggested that for this set of pipelines (those belonging to MEOC) repair of coating defects should be done before the coating defect has reached its median size.

To the best of the Author's knowledge, the process of employing such approaches has never been applied before towards any ECDA data. The findings from this thesis also shed some light into the stochastic nature of the evolution of corrosion pits. This was not known before and is only made possible by the usage of the approaches highlighted above. The resulting models are also of novelty since no previous model has ever been developed based on the said methods.

The contribution to knowledge from this research is therefore the greater understanding of relationship between variables stated above (TCDA, %IR and POPD). With this new knowledge, one has the potential to better prioritise location of excavation and better interpret DCVG indications. With the availability of ECDA data, it is also possible to predict the magnitude of corrosion activity by using the models developed in this thesis. Furthermore, the knowledge gained here has the potential to translate into cost saving measures for pipeline operators while ensuring safety is properly addressed.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, ***Alhamdullillah!*** All thanks and praise to the Almighty God for making this thesis a possibility. Special thanks also go to my parents, Dr Muhd Noor and Tuan Ruhaya, who have provided me continuously with love and support throughout my life and the motivation needed for the completion of this work. Thank you to my dearest wife, Dr Zurina Moktar who has loved and taken care of me and provided me with all the necessary support and resources which have made this journey a success.

A big thank you is directed to my supervisor Prof. Keming Yu who has tirelessly provided me with the necessary knowledge and supervision for the work presented in this thesis, without whose guidance the present research would have been much more difficult. I am also indebted to my 2nd supervisor and mentor Prof. Tat-Hean Gan for recognizing the potential I had and always feeding me with the necessary advice and direction for the completion of this work. As for my industrial supervisor, Dr. Ujjwal Bharadwaj, a special thank you is headed to you for the ideas and support you have given me within the three years I had with AFM. All of the contributions from these people has made this research a success. You are all mostly appreciated.

To all my friends from NSIRC, AFM CaMMUS and TWI Ltd., especially Hwei Yang Tan, Saiful Effendi Syafruddin, Hazwan Daut, Jazeel Chukkan, Renaud Bourga, Fairuz Shamsuddin, Saiful Tumin, Weeliam Khor, Anthony Jopen and Muhammad Terchoun, I would sincerely thank all of you for being part of this endeavour which have made this journey meaningful.

Finally, I would like to thank my sponsors Majlis Amanah Rakyat (MARA) and the Malaysian Government for providing me with this opportunity and the financial assistance needed during my stay here in the UK.

Funding Bodies and Research Hosts for this work;

List of Abbreviations

%IR	IR Drop
AC	Alternate Current
ACVG	Alternate Current Voltage Gradient
AFM	Asset Fracture (Integrity) Management
AI	Artificial Intelligence

AIC	Akaike Information Criterion
ALD	Asymmetric Laplace Densities
ANN	Artificial Neural Network
API	American Petroleum Institute
BBN	Bayesian Belief Network
BQR	Bayesian Quantile Regression
CBM	Condition Based Maintenance
CDA	Coating Disbondment Area
CIPS	Closed Interval Potential Survey
CM	Condition Monitoring
CoD	Coefficient of Determination
CP	Cathodic Protection
CRA	Corrosion Resistant Alloy
DC	Direct Current
DCVG	Direct Current Voltage Gradient
DOC	Depth of Cover
DUC	Deposits Under Coatings
ECDA	External Corrosion Direct Assessment
FBE	Fusion-Bonded Epoxy

FFS	Fitness for Service
GDP	Gross Domestic Product
HSS	Heat Shrink Sleeves
ICCP	Impressed Current Cathodic Protection
ILI	In-Line Inspection
IMG	Integrity Management Group
LASSO	Least Absolute Shrinkage Selection Operator
LQR	Logistic Quantile Regression
MCMC	Markov Chain Monte Carlo
MEOC	Middle Eastern Oil Company
MMO	Mixed Metal Oxides
NACE	National Association of Corrosion Engineers
NN	Neural Network
NSIRC	National Structural Integrity Research Centre
OLS	Ordinary Least Squares
PCM	Pipeline Current Mapper
POPD	Percentage of Pit Wall Depth
PS	Pipe Size
QR	Quantile Regression

RBI	Risk Based Inspection
RUL	Remaining Useful Life
SA	Sacrificial Anode
SAS	Sacrificial Anode System
SR	Soil Resistivity
SRB	Sulphate Reducing Bacteria
TCA	Total Corroded Area
TCDA	Total Coating Defect Area
TCV	Total Corroded Volume
TIS	Time in Service
TWI	The Welding Institute
UoF	University of Florida

Chapter 1

Introduction

1.0 Motivation

Pipelines are generally a safe and reliable medium for transporting hydrocarbon fluids [1]. In the context of oil and gas, the hydrocarbons are transported from the drilling rig (offshore or onshore) to the refining plant and on to oil depots. But, although the probability is low, as with other structures, pipelines have a tendency to fail [2] which can be detrimental with implications to the economy, society and the environment. Demonstrably good design, appropriate material selection and best practice are key to ensuring the continued safety of pipelines.

For buried pipelines, corrosion is a major concern [3]. The threat can be minimized by the application of coatings which range from spray applied coatings to polyolefin coatings [4]. By isolating the metal substrate from its environment, the process of corrosion is minimized. The coating also serves as a layer which prevents any electrochemical processes from occurring [5] in microscopic corrosion cells present on the metal surface. Pipeline coatings however can themselves fail and in the case of organic coatings, deterioration will happen over time. The failure can be due to incorrect application of the coating, soil stresses experienced by the pipe or the coating's loss of adhesion. Generally, the failure of a coating can be summarised as changes in any of its chemical, physical, or electrochemical properties [6] that will eventually result in coating discontinuities or defects. If this were to occur on pipelines, the exposed metal would be susceptible to corrosion, which in turn compromises the pipeline's overall integrity.

Pipelines which are buried are normally cathodically protected (cathodic protection – CP). This is when an external electrical source (in the case of an ICCP) is provided making the pipe cathodic in nature. The system acts as a backup to the coating system and comes into play when coating defects are present [7]. The amount of current needed for protection depends on the extent of coating failure as larger coating defects will require larger amounts of current for protection.

The monitoring of the CP system and the overall integrity of the pipeline is normally addressed by employing the guidelines given in the NACE ECDA

document particularly SP RP0502-2013 [8]. The performance of the CP system is checked using indirect measurements (measurements taken above ground) in the form of Pipe-to-Soil Potentials traversing along the pipeline using reference electrodes [9]. The assessment is sometimes called a Close Interval Potential Survey or CIPS. Any anomalies found are recorded for further investigation.

Another indirect assessment commonly used is the Direct Current Voltage Gradient (DCVG) where potential drops are monitored (%IR). The technique is used to identify the location of coating defects and to classify their severity based on Total Coating Defect Area (TCDA). Based on the defect severity, a decision can be made whether to proceed with further direct assessment requiring excavation of the defect site. Research into aboveground techniques has shown that the technique is considerably accurate in locating defect locations but lacks the ability to predict precisely the size of coating defects [9].

It is normal where excavation of defect sites proves to be fruitless that small defects are still perceived to be significant (in terms of %IR). Part of the inaccuracy in determining the location of excavation sites is due to the input of subjective judgments (on where to dig) made by experts. These misjudgements will incur higher inspection costs if unnecessary excavations are conducted. On top of this, the US federal regulation [10] has indicated that pipeline operators should produce criteria for the identification and documentation of indications from an indirect assessment technique which

will be considered for further assessment (direct assessment through excavation). The criteria also serve to define the urgency of the subsequent inspection based on the documented and identified indications. As mentioned, the current DCVG tool is, to a certain extent, inaccurate at classifying defect sizes (TCDA) and with the added subjective input of experts, misinterpretation only becomes worse. Prioritisation can only be done when the predicted size of defects (TCDA) is accurately represented and subjective input is limited.

Hence it is imperative that an objective solution is established to prioritise coating defects and corrosion accurately and also to comply with the regulations, which will promote safety and financial prudence. In this thesis, a novel technique of modelling the factors that contribute to DCVG indications and a method to predict defect size (TCDA) are presented. The modelling technique used to predict the TCDA and %IR is Quantile Regression (QR), Bayesian Quantile Regression (BQR) and Logistic Quantile Regression (LQR).

1.1 Aims and Objectives

The aim of this thesis is to derive statistical models based on the quantile regression method (QR) to better predict coating defect sizes and to interpret the indication signal (%IR) from the indirect assessment more precisely. This will help to promote the continual safety of the operation of pipelines while minimising costs. The current method of selecting indications

for further examination through excavation, is done with subjective input from an expert, i.e. it can vary from one person to another. The work done in this thesis aims to minimise such input where some of the models developed only considers continuous (numerical) variables. The level of corrosion and how it affects the size of TCDA and *vice versa* will be investigated to see if there is a correlation. If they are correlated, then by accurately predicting the size of a coating defect, one can also predict the amount of corrosion within that region. Additionally, the outcome of the model generated in this thesis will be considered for the use of future ECDA projects. The specific objectives for reaching these aims are denoted below.

- To develop of regression models based on QR to predict the size of coating defects using all available data within the MEOC ECDA indirect and direct assessment phase.
- To formulate a regularised version of the QR models that may be generalised for application to other pipelines beyond the scope of this thesis.
- To generate models using the BQR method to predict the %IR and TCDA which can be incorporated into future ECDA projects in the form of *a prior* distributions.
- To investigate %IR based on the constrained method of LQR in ensuring the %IR readings are within bounds.
- To investigate of the corrosion variable POPD using the LQR method to see how other variables which give rise to corrosion behave.

- To examine the effects of the POPD variable in other models within this thesis to see how it reacts with TCDA and %IR.
- To generate a *probability of excavation model* by using logistic regression applied to corrosion and coating defects. This will serve as additional information for pipeline operators in their decision-making process for selecting excavation sites.

1.2 Summary of Methodology

The data used for the analysis of TCDA, %IR and corrosion in this thesis comes from the Middle Eastern Oil Company (MEOC) pipeline ECDA data provided by TWI Ltd. All the variables that will be investigated in this thesis are based on this dataset. All statistical analyses done in this research was performed using the R statistical software. The packages used for the regression analyses are summarised as follows:

- Quantile Regression
    - Name of Package: Quantreg (QR)
    - Version - 5.34
- Bayesian Quantile Regression
    - Name of Package: BayesQR (BQR)
    - Version – 2.3
- Logistic Quantile Regression
    - Name of Package: Robust Logistic Linear Quantile Regression (LQR)
    - Version – 1.7

Linear correlation analysis was performed to see if the variables are connected and based on these results the linearity of the models is determined. The results found from here (linear correlation) are not applied to the BQR part of the analysis since the main objective of this section is to find an initial form of knowledge (prior distribution) to be carried over to subsequent ECDAs.

This research uses regression techniques in the investigation of the relationships between variables within the MEOC ECDA indirect and direct assessment phases. A QR method was chosen to characterise the dependent variable, e.g. the TCDA or the %IR, to give a complete picture of its distribution without relying on assumptions of the response distribution, e.g. a normal distribution. As in the case of the MEOC data, most of the dependent variable's distribution is skewed and long tailed. QR is most effective in representing these types of distribution. Different approaches to QR such as BQR and LQR were also applied to the MEOC data for separate objectives. The BQR method allows the usage of prior knowledge to be incorporated within the current assessment. This is an attractive proposition since an ECDA is iterative in nature and updating results is key in the maintenance of pipelines. The LQR method solves the issue of predicting outcomes outside of its predetermined range. The usage of LQR was applied to %IR and POPD since these two variables are both percentages. The thesis flow can be summarised by the flowchart below.

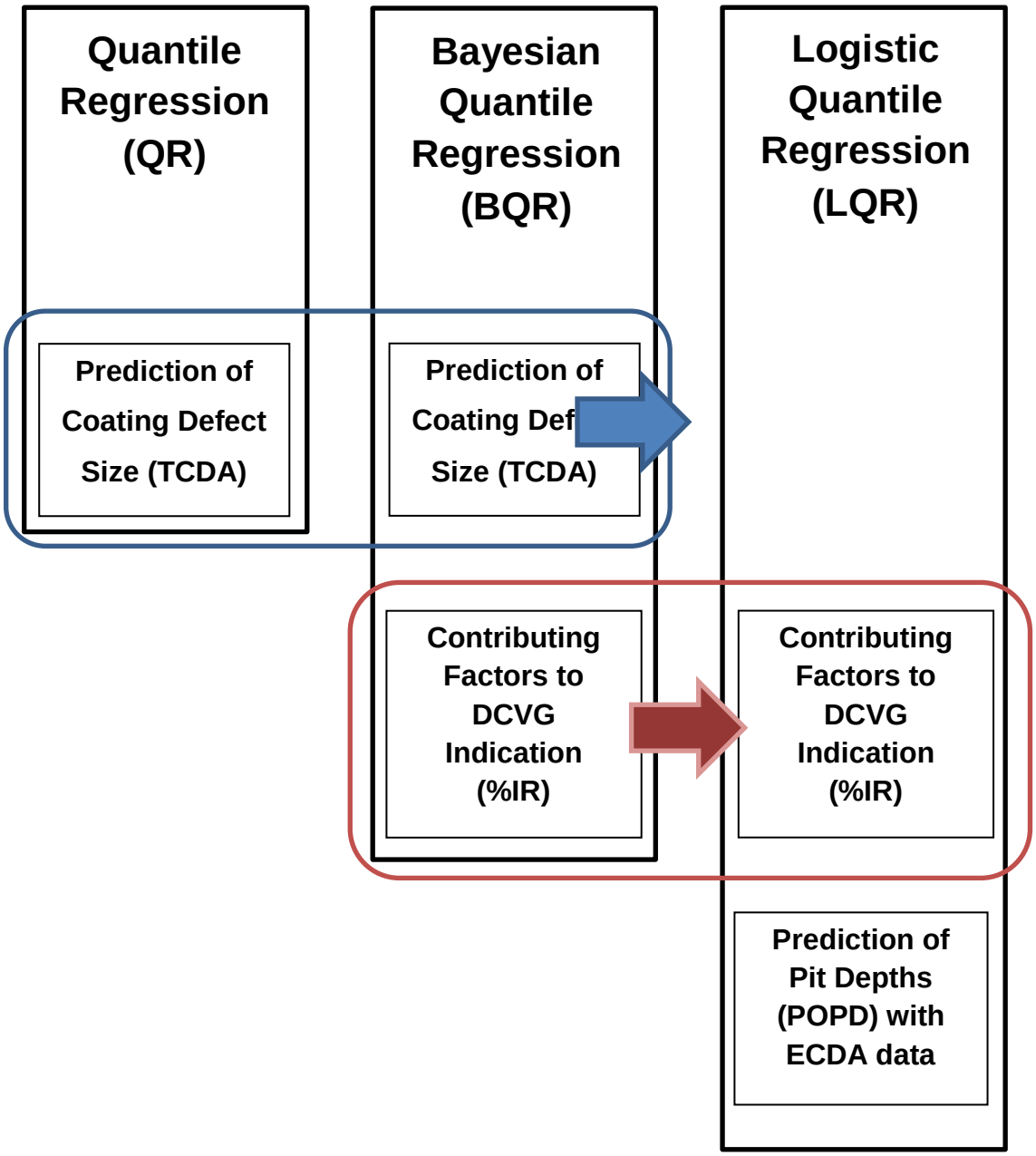


Figure 1-1: Flowchart of Statistical Techniques used for the Analyses of the MEOC data

1.3 Thesis Outline

This thesis is divided into seven chapters. The first three chapters presents a description of previous research conducted on the prioritisation of defects from an ECDA. These chapters also highlight the differences and the advantages of using the methodology presented in this thesis. Analyses using the said methodology are presented in the following three chapters ending with a conclusion and suggestion of future work in the final chapter. A brief description of each chapter is as follows.

- Chapter 2 describes an overview of the ECDA process and the tools used to conduct such assessments. It also highlights current and previous research on defect prioritisation from an ECDA which gives context and identifies the gaps and advantages in the current research presented here. Chapter 2 also touches on corrosion of pipelines which is the subject of interest in the current investigation. Additionally, the MEOC project and its dataset are described in this chapter.
- Chapter 3 gives the present research methodology. The tools used to achieve the aims are predominantly regression techniques. The QR, BQR and LQR techniques are elaborated and justified for each separate objective.
- Chapter 4 highlights the initial application of the QR technique to the MEOC ECDA data. The chapter starts from the correlation of the variables of interest followed by the construction of 2 models predicting TCDA. These two models are distinguished by the

inclusion of interaction variables. A following regularisation technique, particularly the LASSO technique, was conducted to further generalise these models.

- Chapter 5 presents the models constructed from the BQR technique. BQR was used to take advantage of the iterative nature of an ECDA. By formulating models which result in posterior distributions, one can incorporate this with the subsequent ECDA process. The BQR technique is also compared to the regular QR method which results in similar predictions.
- Chapter 6 provides an estimation of %IR based on the LQR method. The building of the model considers only continuous variables. This was done to limit the amount of subjectivity (such as one would get from using categorical variables such as soil properties). This chapter also takes lessons from chapter 5 where certain relevant variables are only needed for good interpretation of the dependent variable. Also, the dependent variable is bounded ensuring interpretation within a specified range (as opposed to the result obtained in Chapter 5). A further analysis using the same LQR technique was used to formulate models for corrosion represented by the POPD variable. The effects of contributing factors to corrosion are investigated.
- Chapter 7 summarises the results obtained from the previous three chapters and also suggests useful future research.

1.4 Contributions to Knowledge

There are a few findings conducted in this research which can be of use in the form of additional knowledge for researches, pipeline operators and regulators alike. The novelty of this research lies in the approach taken by the Author in analysing the MEOC data set in achieving its objectives. These are listed below.

- The relationship of TCDA and its contributing variables based on the MEOC data was modelled using QR and considering the non-linear relationship between the independent variables towards the dependent variable. Polynomial regression models were chosen to describe these relationships.

- The effects of having Interaction and Non-Interaction variables within the QR model were compared and analysed.
- A regularisation procedure in the form of the Least Absolute Shrinkage Selection Operator (LASSO) technique was conducted based on the Interaction and the Non-Interaction QR models. The resulting LASSO models showed that the Non-Interaction Model is more general than the Interaction model which follows closely the MEOC data structure between the %IR and TCDA.
- Modelling was done with the Bayesian approach to QR highlighting the relationship between variables to %IR. The modelling used two datasets. One dataset is the original dataset while the other has 4 points in the model removed by the author. This bases on the general

understanding of the DCVG system and the expected results one would obtain from a similar inspection assessment (higher %IR values are expected to have larger TCDA measurements and *vice versa*). Four models were generated from this procedure resulting in posterior densities that can be used for subsequent ECDA on buried pipelines.

- Two more models were developed based on the BQR with different sets of variables and with the two datasets mentioned previously. Unlike the previous four models, these estimated the TCDA. Results from these models are useful in the form of *a priori* distributions to be used in the subsequent integrity assessment.
- The formulation of the Tex and the $Texx$ variables to show the justification for excavation for the MEOC ECDA project. These variables are later used in the logistic regression of the MEOC data to produce a probability of excavation. The model could be of use as a guideline in the selection of excavation sites for future ECDA endeavours.
- The construction of a model to predict %IR with respect to its contributing factors was done using the LQR method. This overcomes the transgression of boundaries observed in Chapter 4 and 5. The resulting model is in parallel to that in Chapter 5 which indirectly supports the model's validity.
- Another model based on the LQR was also formulated but as opposed to the previous model, the intending outcome is the variable POPD.

Novel findings from this thesis includes:

- Some degree of relationship between the DCVG technique (%IR readings) and corrosion dimension exists. The results of the relationship between TCDA and POPD highlights a negative trend which further supports the idea that %IR has some relation to corrosion.
- Based on the findings from Chapter 4, 5 and 6 suggests that corrosion activity rate is more prominent than the growth of TCDA at its median depth. It is therefore suggested that for this set of pipelines (those belonging to MEOC) repair of coating defects should be done before the coating defect has reached its median size.

To the best of the Author's knowledge, the process of applying the approaches highlighted above has never been applied before towards any ECDA data. The novel findings from this thesis also shed some light into the stochastic nature of the evolution of corrosion pits. This was not known before and is only made possible by the usage of the approaches highlighted above. The resulting models are also of novelty since no previous model has ever been developed based on the said methods.

1.5 List of Publication

1. Making Use of External Corrosion Direct Assessment (ECDA) Data to Predict DCVG %IR Drop and Coating Defect Area – Paper accepted

for publication in the *Materials and Corrosion* Journal. This paper highlights the results obtained from the ECDA process which was conducted on 250km of buried pipelines. The results from the indirect and direct assessment part of the ECDA were modelled using the classical Quantile Regression (QR) and the Bayesian Quantile Regression (BQR) method to investigate the effect of factors towards the IR drop (%IR) and the coating defect size (TCDA).

1.6 Hosts for Research (Industrial): TWI Ltd.

The entirety of this research was done at TWI ltd. It is the industrial host for this Doctorate where it partners with Brunel University London through its subsidiary, The National Structural Integrity Research Centre (NSIRC). TWI Ltd. is a non-profit organisation which is membership based, championing training, research, technology and consultancy in the field of joining technology, integrity management and materials engineering. The Author was based in the Asset Fracture Integrity Management (AFM) section which is part of the larger Integrity Management Group (IMG). AFM conduct projects relating to Risk Based Inspection (RBI), Fitness for Service (FFS), Fracture Mechanics and software development.

NSIRC is a first for the UK. The centre brings together industry and academia to promote research in structural integrity and engineering. More than 20 universities and two founding sponsors (BP and Lloyd's Register Foundation) have collaborated with NSIRC since its inception in 2013.

Research conducted at NSIRC is industrially driven (working with real life projects) and is determined by the current needs of the industry. On top of this, academic supervision is provided by affiliating universities such as Brunel University London (which is the case for the Author) to ensure the research meets its intended purpose.

Chapter 2

Literature Review on Integrity Management Techniques for Buried Pipelines

2.0 Introduction

Corrosion can be defined as a naturally occurring phenomenon that involves the deterioration of metal as a result of interaction with its environment [11]. Metals that are not in their natural state corrode. Metals which are do not. Gold and silver are examples where metals are in their natural state. The process of corrosion occurs when metals revert to their natural state [12], [13]. Energy is added to metal oxides to produce industry-usable metals

such as carbon steel. Corrosion is a process where energy is “leaking out” from the metal, reverting it to its natural state. Trace elements, for example chromium, may be added in metals such as stainless steel to slow this process by forming an oxide layer [14]. Although the rate is slowed, corrosion does still happen. Records have shown that the ancient Romans have applied tar to their steel as a form of protection against its environment which stops or slows down the process of corrosion [15].

The issue of corrosion has become an increasingly important topic in recent years due to the escalating costs associated with the problem. Baker [16], has shown that there have been 1074 significant incidents since 1988 occurring in both onshore and offshore pipelines in the USA associated with corrosion. Here, significant means the occurrence of fatalities or hospitalization and property damage. These incidents equate to a total cost of 0.5 billion dollars in damages. Another study in the United States by CC Technologies Laboratories, Inc., entitled “Corrosion Costs and Preventive Strategies in the United States [17]”, conducted from 1999 to 2001 with the support of NACE, found that the direct cost of corrosion in the U.S totals $276 billion which is approximately 3.1% of the country’s gross domestic product (GDP). A recent study by NACE within its IMPACT (International Measures of Prevention, Applications, and Economics of Corrosion Technologies Study) [18] group informs that in 2013, the estimated global cost of corrosion was 2.5 trillion dollars, equating to 3.4% of the global Gross Domestic Product (GDP). If current corrosion mitigation measures are taken, there is potentially 15% to 35% (equivalent to 375 to 875 billion

dollars) of savings that could be realised. The study goes on to say that the high cost of corrosion has been known for years. One such study was conducted by Uhlig in 1949 [19] which showed that the cost of corrosion in the United States in 1949 was 2.5% of total GDP.

Knowledge of the mechanisms of corrosion is imperative for the successful implementation of a corrosion management strategy. For corrosion to happen, four elements must be present, namely an anode, a cathode, a metallic path and the electrolyte [20]. If any one of these elements is not present within the electrochemical cell, then there will be no corrosion. Chemical reactions occur between the metal and the electrolyte at the anodic and the cathodic areas of the metal. At the anode area, metals tend to give up electrons (oxidation). The electron travels through the metallic path and is then picked up by the cathode (reduction). Essentially, what happens at the anode is the dissolution of metal into metal ions which reacts with elements in the electrolyte, in the case of iron producing rust.

Different metals have different tendencies to give up electrons. For example, zinc has a higher tendency to give up electrons compared to copper. We call zinc more reactive than copper. This tendency is largely due to the amount of energy needed to extract the metal from its natural oxides. The force that drives metallic corrosion is called the Gibbs energy change. In corrosion terms, volts (V) are used to denote the driving force. Corrosion is favourable and can be exacerbated in some conditions. Examples include the exposure of carbon steel to hydrogen sulphide, H_2S, which may or may not be

detrimental based on the H_2S partial pressure [21]. Another example is where higher temperatures increase the rate of chemical reaction and the rate of the corrosion process. The effect of temperature on the corrosion rate of (e.g.) iron depends on its influence on the oxygen solubility, the solution viscosity, diffusion rates, oxidation rates and biological activity [22]. The aggressiveness of the corrosion process in saline solutions were looked at in [23] where structural steel specimens were exposed for two months to different samples of sea water with differing salinity. The study arrived at the conclusion that the aggressiveness of sea water in terms of corrosion is not only a function salinity but also of temperature, its pH, oxygenation, water flow and dissolved gases.

Consequences to corrosion can be very substantial. The book by [11] categorises some of the consequences of corrosion into the following situations:

- The shutdown of plants
    - In the event of a processing or power plant and refinery shutdown, the downstream effect will snowball to industry and the consumer.
- Loss of containment (LOC)
    - The leakage of products can be detrimental to the environment and cause hazards in the surrounding area. In any water distribution network system, 25% of the water may be lost due to corrosion.

- Loss of efficiency
    - The corrosion product from the heat exchanger tubes tends to reduce the effectiveness of heat transfer and piping capacity due to build up in the surrounding metal.
- Contamination
    - Corrosion products may contaminate products consumed by consumers and pose a health threat to society.

As stated earlier, the consequences of corrosion can have severe implications for cost as well as damage to the environment.

In modern times, corrosion is fought with a range of techniques which may be either simple and complex in nature. Simple but effective techniques such as applying coal tar, organic coatings and concrete on pipelines have been applied. Of all the approaches to preserving a pipeline from corrosion, the most important is Cathodic Protection (CP) [24]. This technique will be discussed in greater detail in the next section. Corrosion inhibitors are also used where their application is mainly for the prevention of internal corrosion. These work by forming a protective layer inside the pipe where corrosion is minimized (a form of internal coating) due to the inhibition of corrosion reaction. Apart from external techniques, the use of high resistant steel or corrosion resistance alloy (CRA) is also an option but this will likely result in higher manufacturing and construction costs (higher capital

expenditure) compared to carbon steel which will increase expenditure in term of operational cost [25].

2.1 Cathodic Protection (CP)

2.1.1 Background

The use of zinc coatings for the protection of steel started in France as early as 1742 [26]. Steel was protected by dipping into molten zinc, a method called *galvanizing*. Cathodic protection was also introduced by Sir Humphry Davy in 1824 when he successfully demonstrated the protection of copper sheathing by iron rods (anodes) [27]. In his experiment, Davy showed that by electrically connecting two pieces of metals and submerging them in an electrolyte, one of the metals deteriorated (iron or zinc) at a faster rate while the other metal (copper) would remain in good condition. Based on this knowledge, Davy proposed to the British Navy to protect copper sheathed ships by attaching blocks of iron to the ship's hull. As suspected by Davy, the result of this trial showed that copper was protected against corrosion by allowing the iron blocks to corrode. However, the cathodically protected copper is susceptible to marine fouling compared to the non-protected copper. Non-protected copper generates concentrations of copper ions (during degradation) which poisons marine life and prevents marine growth. Due to this, the British Navy rejected the proposal by Davy on the basis that fouling on the ship's hull causes a ship's speed to decrease. Davy's work was later picked up by Robert Mallet in 1840 where he produced sacrificial anodes from zinc alloys. The zinc alloys worked well locally especially when preventing the galvanic effects coming from the ship's bronze propeller.

When wood was replaced by steel hulls, all the British Navy's vessels were fitted with this anode. In the 1950s, the Canadian Navy applied cathodic protection to their vessels in a combination of an anti-corrosion and anti-fouling paint. The result from this combination showed that the usage of cathodic protection on sea vessels is possible and lowers the maintenance cost. The fuel costs for operation of these vessels was also shown to decrease due to the smoothness of the hull (without any fouling) which decreases the amount of drag exerted by the water.

In the early 20th century, CP made its way into the pipeline industry. The very first such application started in England and the United States in 1910 – 1912 [28]. The reason for this introduction was that pipelines are buried in corrosive environments. In addition, pipelines span hundreds of miles which makes protecting the entire length a challenge. To put this in perspective, in the US today, 2.4 million miles of pipeline are used to carry oil, gas and other petroleum products (the largest network of pipelines in the world) [29]. Due to these concerns, pipeline operators in the 1930's introduced pipeline coatings and CP as means of preventing corrosion. Along the length of the pipeline, the soil's characteristics were measured and the most likely locations for corrosion were identified. These hot spots received the most attention for coating and CP protection. In 1933, Kuhn proposed the first criterion for CP which states that pipelines should be polarised to -0.85V to ensure that corrosion does not occur [30]. This criterion has been accepted worldwide and is known as the protection potential criterion.

In buried pipelines, coatings are applied during construction or pre-coated before the laying of the pipe into its intended place. Coatings combat corrosion through several ways but mainly by isolating the metal of the pipe from the surrounding environment (good barrier properties which prevents the ingress of corroding species such as water and oxygen), restricting the flow of corrosion currents within a corrosion circuit and by allowing the CP current to counteract the corrosion process by supressing them (making them cathodic). [31] gives a guideline for the properties of an effective coating system for pipelines such as: possession of strong adhesion characteristics; tolerance of high operating temperatures; ability to withstand soil stresses and cathodic disbonding. A regulation in the United States, 49 CFR 192.461 – External Corrosion Control: Protective Coating [32], exemplifies this further by stating that a protective coating should be applied on a prepared surface, have sufficient adhesion to the pipe (which resists the migration of under film moisture), have ductile properties to resist cracking, having sufficient strength to withstand the stresses of the soil and have properties compatible with the supplemental CP.

Some of the best coatings are made from organic materials. The application of these coatings is not limited to pipelines but is applied to storage tanks, bridges, ships and marine structures. Organic coatings have been the choice for pipeline operators due to performance, corrosion resistance properties, strong adhesion to the structure, fast application and high abrasion resistance [33]. Types of coatings currently in use include coal tar enamel, polymeric tapes, fusion-bonded epoxy (FBE), spray-applied liquid coatings,

and two and three-layer polyolefin coatings, e.g. a three-layer polyethylene coating. Thompson and J. Saithala [34] have divided these coatings based on the time period it was developed. The first generation (since the 1920's) of coatings includes coal tar enamel, asphalt, single layer polyethylene, two-layer polyethylene, tapes and heat shrink sleeves (HSS). The second generation of coatings which began in the 1970's are the single layer FBEs while the third generation of coatings which started in 1985 includes three-layer polyolefin and dual layer FBEs. By having a CP system working with a coating system, a pipeline's continual safe operation is ensured. The protection current needed for corrosion prevention is also reduced due to the coating's ability to distribute the CP current along the pipe more evenly [35]. Additionally, some coatings allow CP currents to pass through them after the coatings have absorbed some water [36–38]. Due to this, protection current is able to reach and suppress corrosion even when the pipe is not fully exposed to the environment.

During the installation of the pipelines, there is the possibility of coating damage as a result of improper handling. Excavator arms which are usually used to position the pipes in place have the potential to scrape some of the coating off. Unintentional scraping of pipe coatings can also happen during excavation of a buried pipe [39]. This breakdown in the continuity of the coating is often referred as *coating defect* or *coating holiday*. Besides the damage acquired during the installation and the excavation process, due to their organic nature, these coatings will eventually fail with time. Coating defects lead to the loss of adhesion between the coating and the pipe.

Electrochemical mechanisms such as coating disbondment and oxide lifting can be considered as one of the reasons for coating degradation [40]. At the location of coating defects, the CP current flows underneath the coating thus generating alkaline hydroxyl groups through the reduction of the cathode. If the coating is not suitable or is unstable for the CP system, these hydroxyls along with the pipe's negative potential cause the coating to disbond (loses adhesion) from the pipe. The alkalinity generated from the cathodic reaction reacts with the organic polymer to disbond the coating at the metal coating interface [40]. The process will be amplified if the original coating defect is large. More current is drawn to the defect which exacerbates the problem. The oxide lifting phenomenon occurs due to the anodic corrosion product which accumulates underneath a coating defect. The corrosion product combined with the compacted oxides contribute to more of the coating being disbonded.

Work was done by [41] where it was showed that a CP system is required as backup for a coating system due to the accelerated corrosion found at certain areas of the coated pipe. Another study by Riemer and Orazem found that corrosion at coating defects is much more severe compared to pipes with no coating at all. [42]. At coating defects, the exposed bare metal tends to be anodic while the adjacent metal underneath the coating becomes cathodic. A corrosion cell is then developed where the cathodic part of the metal is protected but the exposed section undergoes accelerated corrosion. Due to the potential difference created, corrosion is more

pronounced at locations where coating is missing whereas the cathodic part of the metal is protected even more.

2.1.2 The Inner Workings of CP

As was stated in the previous section on corrosion, metals have a tendency to give up electrons. It is a process by which the metal is inclined to return to its natural state as metal oxides. This reaction is called corrosion. The generation of cathodic and anodic sites of a corrosion cell can be either from two different metals touching each other (metallic path) or from the same piece of metal where micro corrosion cells are created due to the variations in metallurgical properties across the surface. The availability of oxygen is also a factor. Oxygen rich areas of the surface of steel will tend to become a cathode and thus protect it from corrosion, and an anode is formed where the oxygen supply is depleted [43]. Corrosion only occurs at the anodic site of the corrosion cell. A common example of the reaction at the anodic side of iron:

$$Fe \rightarrow Fe^{2+} + 2e^{-} \qquad \textbf{2.1}$$

Iron giving up electrons is called oxidation. The free electrons are then channelled through a metallic path to the cathode area where a cathodic reaction takes place. There are two cathodic reactions based on the pH of the electrolyte. In neutral solutions or seawater, the reaction is:

$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- \quad \textbf{2.2}$$

Hydroxyl ions are produced while in acidic conditions the reaction is:

$$2H^+ + 2e^- \rightarrow H_2 \quad \textbf{2.3}$$

Hydrogen gas is produced.

If there is a change to a structure, either by supplying or withdrawing the number of electrons in equation 2.1 and 2.2, the reaction will also change. It is a well-known principle, that whenever there is a change in an equilibrium system, the system will try to adjust itself back to its original equilibrium state [44], [45]. For example, if electrons are withdrawn in a structure, the reaction rate in equation 2.1 will increase due to the system trying to compensate for the loss of electrons. Similarly, if electrons are added to a structure, equation 2.1 will decrease and equation 2.2 and 2.3 will increase based on the pH of the electrolyte [44], [45].

An electrode is said to be not in its equilibrium state when there are currents (electrons) flowing to or from its surface. In a galvanic cell, the flow of these currents alters the potential between the electrodes where the anode tends to be more cathodic and the cathode becomes more anodic. As the flow of currents increases, the potential difference between the electrodes decreases. This change of potential either for the anode or the cathode is called polarization.

Cathodic protection builds on these 2 aforementioned concepts. The cathode is polarized by supplying current (electrons) from an external source so that the potential equals the thermodynamic potential of the anode. At this stage, the anode and the cathode are said to be equipotentials. Since the anode is at its thermodynamic potential, corrosion will not occur. A cathodic protection system works by bombarding the corroding metal (whose protection is intended) with current using an external power source. An auxiliary anode is connected to the pipe through a metallic path which is normally of low resistance. Auxiliary anodes can be in the form of a galvanic anode or an impressed current anode. The current in a galvanic anode stems from the potential difference between the two metals (the pipe and the auxiliary anode) and the current from the impressed system comes from an external DC power normally through a transformer rectifier unit [43]. Current then leaves the auxiliary anode, passes through the soil (electrolyte), polarises the metal (that needs protection) and returns to the current source (external). If conditions are favourable, the system can be 100% effective in combating corrosion since the anode has returned to its original thermodynamic potential. Metal ions are prohibited from entering the electrolyte because of the current supplied by the external source which supresses the ions to remain at the anode. As long as the metal (anode) is kept at this potential, then corrosion is impossible [24].

2.1.2.1 Sacrificial Anode System (SAS)

SAS, also known as galvanic systems, uses sacrificial anodes (SA) for the generation of protective currents. When the SA goes through an anodic reaction, electrons are released (oxidation) producing current which flows in the electrolyte to the cathode (reduction reaction). As the SA undergoes this reaction it is essentially sacrificing itself to protect the structure. The rate of degradation of the anode depends on the potential difference between the anode and the cathode. Figure 2-1 below shows a schematic of a typical layout of a SAS system.

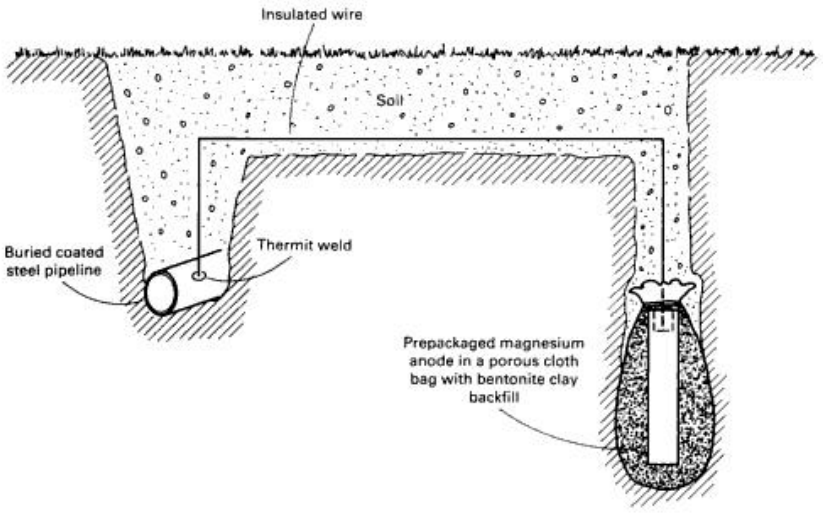


Figure 2-1: Typical SAS layout (image taken from corrosion-in-rod-pumped-wells.wikispaces.com)

The size and physical characteristics of the SA are based on the amount of surface area of the structure that needs protection. The types of SA in common use to protect carbon steel are magnesium, zinc and aluminium. All of these have different current densities and potential. Selection of a certain type depends on a number of factors such as the electrolyte's resistivity, required current density and the length of protection. Advantages of SAS is that it is normally cheap to install with little maintenance thereafter.

Moreover, it is portable since it does not require an external source of power. The main disadvantage of a SAS system is that the driving force is low compared to an impressed system. Additionally, with higher soil resistivity the lack of output becomes even more of a problem. However, in certain situations, the low voltage output can be an advantage compared to an impressed system in that the threat of overprotection is reduced. Also, due to its low output value, the possibility of interfering with adjacent pipelines through stray currents is also reduced.

2.1.2.2 Impressed Current Cathodic Protection (ICCP)

In an ICCP system, an external current is supplied to the cathode through the auxiliary anode. This external current is supplied through a transformer/rectifier which lowers the voltage and converts alternating current (AC) power to direct current (DC). The current produced by the transformer rectifier creates a potential difference and current is passed from the anode (oxidation) through the electrolyte to the cathode (pipe to be protected). Currents from the cathode then travel through the low resistant wire back to the transformer/rectifier which completes the circuit. The accumulation of current from the external source at the pipe surface makes it negatively charged (polarized), up to the point where only cathodic reaction (reduction) takes place. If too much current is supplied to the structure, the pH of the *catholyte* changes and will eventually lead to hydrogen evolution. Normally, a limit of -1200mV polarization is applied to pipelines to prevent this phenomenon. The evolution of hydrogen has the potential to initiate

cracks and also cause the coating to disbond (at coated pipes). In water with high dissolved chlorides (high salinity), chlorine can be evolved at the anode. Chlorine gas poses a serious health threat to humans (if inhaled) and working in such an environment should be performed with care [45]. Figure 2-2 shows the arrangement for a typical ICCP setup for buried pipes.

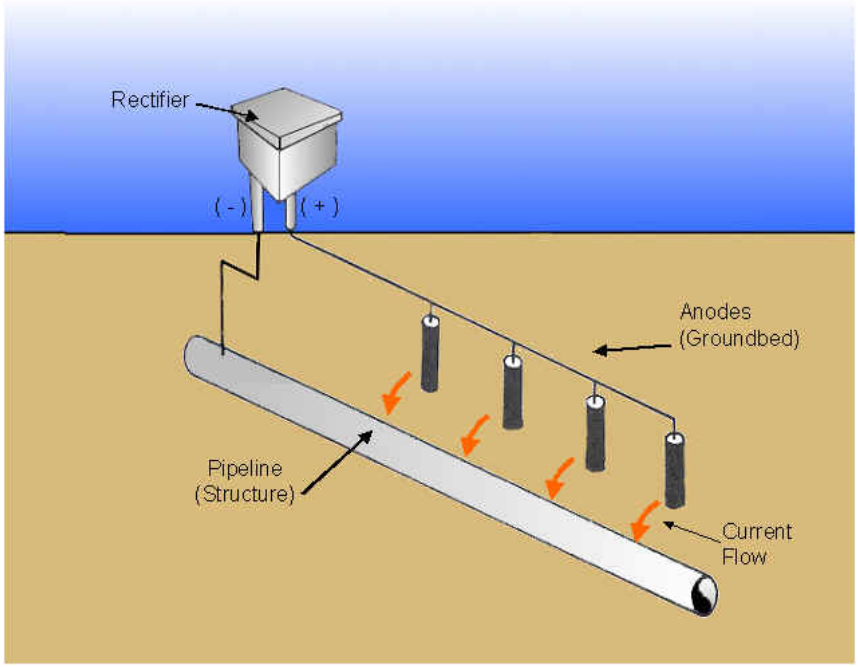


Figure 2-2: ICCP System for a pipeline (image taken from http://encyclopedia.com.my/category/technology/)

In a typical buried pipeline arrangement, the anode is made from materials such as graphite, lead alloys, high silicon iron, mixed metal oxides (MMO) and scrap iron [45]. If scrap iron is used as the anode, then the anode follows equation 2.1. The anode is then categorised as a consumable anode. But if the anode is electrochemically inert the electrolyte will be oxidised (lose electrons) and follow the following equation:

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^- \quad \textbf{2.4}$$

The biggest advantage of having an ICCP system installed is that it can last a long time (for electrochemically inert anodes) and provide greater driving force in terms of providing protection current to the structure. The application is most suited where high soil resistivity is encountered. An ICCP system can also cover great lengths of pipe having one system every few kilometres. One ICCP system can also cover several pipes from one location. Moreover, the user can control the amount of current being applied to the pipeline therefore regulating the amount of protection needed based on the characteristics of the environment. As opposed to SAS, once the system is installed, there is no way of modulating the amount of current it produces other than the initial current that the operator has previously determined. There are downsides to the usage of an ICCP system which were highlighted earlier. Examples such as hydrogen evolution, chlorine evolution and coating disbondment are the result of incorrectly applying too much protection (over protection). This however can be avoided by following international standards for proper usage and understanding the mechanisms of corrosion in a given environment.

2.1.3 Coatings as Means of CP Current Distribution

Pipelines with no coating receiving cathodic protection will be protected from corrosion but the location of such protection is not distributed evenly. The improvement of the distribution of protective currents along the length of the

pipeline can be achieved by coating the pipeline. Coatings can range from organic coatings (for normal temperatures) or glass coatings (at elevated temperatures). Protective currents are drawn to coating defects which ensures protection of the pipeline where the metal substrate has been exposed to the corroding environment. Additional benefits of protecting coated pipes are that the current requirement for complete protection is less and thus more coverage of protection is possible. Moreover, the lower required protection current leads to fewer auxiliary anodes being used, which translates into reduction of cost. Soil is regarded as an electrolyte with good electrical conductivity. Because of this, one magnesium anode is able to protect 8km of coated pipeline compared to only 30m of protection for uncoated pipeline. With the addition of an impressed current system, one magnesium anode can cover 80km of pipeline, 10 times more than SAS. The governing limit to the coverage of one anode is based on the metallic resistance of the pipe and not on the soil resistivity.

2.2 External Corrosion Direct Assessment (ECDA)

The safe continual operation of a pipeline depends on its structural integrity. External Corrosion Direct Assessment (ECDA) is intended to maintain this (mandated) level of safety by having a structured process which assesses and reduces the threat of external corrosion. ECDA seeks to identify and rectify corrosion activity proactively by repairing corrosion defects and eradicating its causes. By doing so, the ECDA process limits the growth of defects so that they do not affect the overall pipeline integrity. The

application of ECDA must be continuous. Iterative assessments are done periodically to ensure continual improvements to pipeline integrity. From the continuous process of ECDA an assessor is able to locate corrosion which has happened, is currently happening and the locations of possible future corrosion.

Initially, ECDA was introduced as an alternative method for assessment of the integrity of pipelines. It was intended to be used on pipelines that were not piggable and were not subjected to pressure testing. Nowadays, the method is becoming ever more popular and most pipeline operators use the method to assess the integrity of buried pipes. Traditionally, pipeline operators relied on inspection techniques such as Closed Interval Potential Survey (CIPS), Direct Current Voltage Gradient (DCVG), Pipeline Current Mapper (PCM) and Alternating Current Voltage Gradient (ACVG) to identify locations of possible corrosion activity. ECDA takes these methods and organizes them in a more systematic way. Characteristics and operating history of the pipeline (pre-assessment) are integrated with the field investigation (indirect assessment – the aforementioned techniques) and inspection of the pipeline's surface (direct assessment). After this is done, an evaluation (post assessment) on the condition of the pipeline is conducted based on all the findings from the previous steps. Thereby, a more comprehensive outlook on the "health" of the pipeline is achieved [8]. The following is the detailed description of the steps of an ECDA.

2.2.1 Pre- Assessment

In this stage of the assessment, a thorough study on the pipeline's characteristics such as the physical dimensions of the pipe, its operating pressures, temperature, the material of the pipe, etc. is carried out. The environment of the pipeline is also looked at and parameters such as soil resistivity, terrain features and other nearby structures are analysed and annotated. Other factors to consider are the pipeline's operating history and previous ECDA results (if available). This information helps the assessor to determine if an ECDA is suitable for the assessment of a particular pipeline. The selection of indirect techniques to be used is also determined at this stage. If the assessment shows that the pipeline is buried underneath a river or is situated on private property, for example, techniques such as ACVG, and access permission from landlords are required.

2.2.2 Indirect Assessment

This stage of the assessment uses the indirect techniques previously mentioned. Locations of where to conduct such an assessment are determined from the pre-assessment stage. The purpose of inspection at this stage of the ECDA is to identify if there are any coating defects (through DCVG or ACVG) where corrosion is likely to occur. Also, the level of protection that the pipeline is currently receiving is determined (by conducting a CIPS assessment). By combining these two streams of information, an assessor is able to determine the most probable locations of

corrosion activity. Locations which show inadequate cathodic protection coupled with the absence of coating lead to a high possibility of corrosion. The ECDA requires both of these inspection techniques to be done one after the other with little time in between. Some specifications advise on using three assessment techniques rather than two. The advantage of using three methods is that the prioritisation of locations which needs the most attention can be determined. The determined locations are normally reserved for the slowest technique which is normally the DCVG.

2.2.3 Direct Assessment

Findings from the Indirect Assessment step is considered for this step of the ECDA. Locations of where corrosion is suspected to be worst are determined (from the indirect assessment step) and excavations of these sites is conducted to investigate the severity of defects. The excavation sites are termed *bell hole* sites since the excavation is shaped like a bell. At these excavation sites, a direct inspection of the pipe surface is possible and the accuracy of the results from the indirect techniques is determined. To see whether the results obtained from the indirect methods completely correspond to the actual physical findings, random excavations are conducted at locations where there were no indications indicated from the indirect assessment. Direct examination within these bell holes includes measurements of the soil resistivity, corrosion dimensions, coating defect dimensions, deposits accumulated underneath coatings, the type of soil within the pipe's environment, the pH of water (if present) in the bell hole and

the pH of water underneath a defected coating. Also at this step, corrosion and coating defects are repaired before moving on to the next step.

2.2.4 Post Assessment

This step of the assessment determines the effectiveness of the previous three steps. Consequent iteration of the inspection process is also determined to ensure that defects found would not reach such a critical size as to jeopardize the pipe's integrity. Also, an estimate of the remaining life of the inspected pipeline is determined and recommendations for pipeline integrity management are made.

2.3 Indirect Assessment Techniques

The above ground indirect assessment is used for the 2nd step of the ECDA process. Assessments are done above ground where no physical contact (except for CIPS – contact to the pipe is achieved through a trailing copper wire) exist between the inspection instrument and the pipeline. The inspection is based on the CP system giving off currents to the pipeline thus producing voltage drops (%IR – coating defects) and the pipe-to-soil potential which is distributed along the length of the pipe. The following are the most popular indirect techniques that are practised in industry today.

2.3.1 Current Mapper (Attenuation)

The method is sometimes referred to as the Pipeline Current Mapper (PCM). This indirect technique is aimed at determining the condition of the pipeline's coating. The technique works by transmitting current from one end of the pipe and receiving it at distances further down the length of the pipe. In a way, it tries to replicate the current received by the pipeline in the CP system. The receiver takes measurement at incremental steps away from the transmitter. These measurements are then plotted on a graph to see the behaviour of the current attenuation with respect to distance. As the current travels along the pipeline, a steady rate of attenuation is observed indicating a pipe which has good coating conditions. At locations of coating defects, the drop in current attenuation is sudden due to the large amounts of current being consumed to protect the exposed metal substrate – much like a CP system where currents are drawn to the location of coating defects. The rate of this drop can be used as an indicator to the severity of the coating defect. Patterns and trends from the plotted graph also show the distribution of current for the entire length of the pipe under inspection. This can be advantageous since the trend is similar to that one would expect when the pipe is under the protection of cathodic current. Therefore, locations of underprotection can be determined and rectification is possible. The CIPS procedure also has this ability. This will be discussed in detail in the next paragraph.

2.3.2 Close Interval Potential Survey (CIPS)

The main aim of a Close Interval Potential Survey is to assess the working capacity of the CP system. Traditionally, measurements of the pipe-to-soil potential are taken at test posts where an electrical cable (providing electrical contact) is attached to the pipe. The measurements taken at a test post is a measure of the CP performance only at that location and do not include potential readings in between the test posts. For this reason, CIPS was developed to assess CP performance in-between test posts so a potential profile can be developed to identify faults of under protection of the system and possibly locate coating defects.

The survey is referred as *close interval* due to the closeness of each interval when making potential measurements. These measurements are taken using reference electrodes which themselves have their own potential value. For the purpose of standardising the measurements, electrodes are calibrated based on the reference electrode potential at the start of each survey. The selection of the type of electrode to be used depends largely on the type of electrolyte in which the pipe is residing in. Normally-used electrodes range from copper – copper sulphate ($Cu/CuSO_4$) to silver – silver chloride (Ag/AGCl).

CIPS is conducted by a surveyor who uses reference electrodes attached to the bottom of a walking cane to take potential measurements by stabbing it to the ground along the traverse of the pipeline route. For the completion of the electrical circuit, the surveyor carries a trailing copper wire attached to a test post at one end and attached to the electrodes at the other. Current

supplied by the ICCP system is interrupted for the surveyor to take two measurements – the ON potential and the OFF potential. The reason why two measurements are taken is due to IR error caused by the indirect technique itself. Soil in-between the reference electrode and the pipeline will contain the generated voltage gradient due to the current being supplied by the CP system. Because of this voltage gradient (error) accurate readings are not obtained during the ON potential. Therefore, an OFF potential reading which eliminates the voltage gradient is preferred as it is a truer representation of the pipe-to-soil potential. Well-developed criteria specify that a cathodically protected pipeline should have a potential of -850mV and a maximum of -1200mV OFF potential for sufficient protection [9] [46]. During CIPS, an interrupter is used to interrupt the current supplied to the pipe. However, if the CP system is left in the OFF mode too long, there is concern about depolarising the pipe where corrosion might initiate. To prevent this from happening, a 4 second cycle is introduced, 3 seconds of which will be in the ON mode and 1 second in the OFF mode. This has shown to be a very efficient way to collect measurements of the OFF potential without significantly depolarizing the pipe. After taking close interval measurements of pipe-to-soil potential, the data are plotted to produce a potential profile along the length of the pipe. An example of the potential profile can be seen in Figure 2-3. The plot can be categorised as three different types namely Type 1, Type 2 and Type 3 [47], [48].

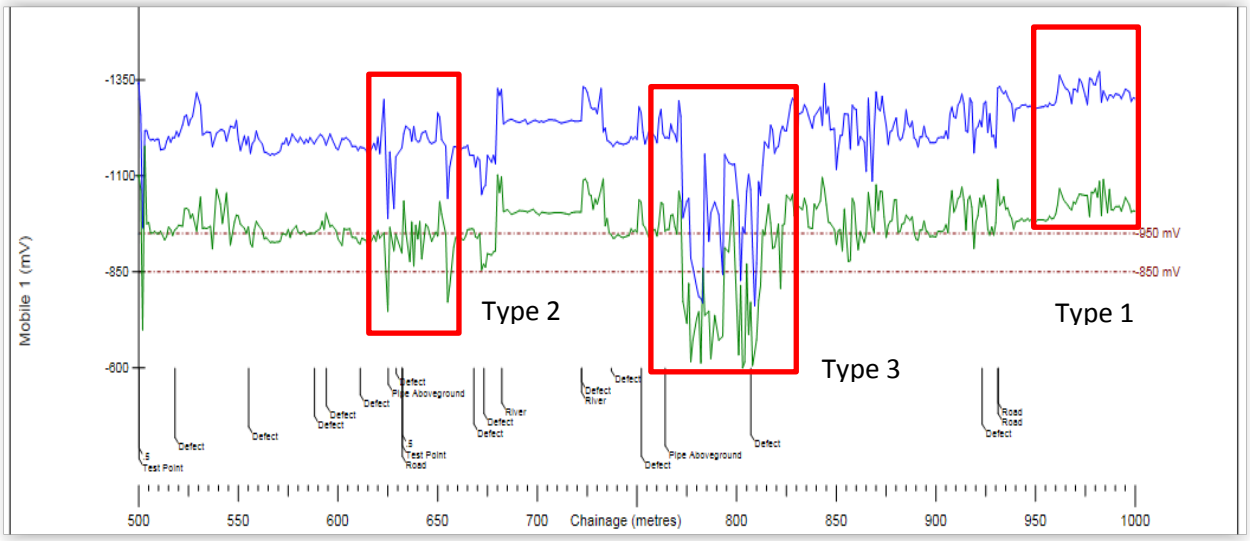


Figure 2-3: CIPS Data from the Field Showing Different Types of Defect

A type 1 defect can be considered a minor defect where the ON and OFF potential is still below (more negative) than the -850mV criterion (NACE SP0169-2007). CP protection here is unsatisfactory and a further drop in potential (more positive) relates to an unprotected segment of the pipe. This area is susceptible to corrosion if the pipe coating is also unsatisfactory. Type 2 describes the OFF potential to more positive beyond the set criterion while the ON potential is still within limits. It represents an unprotected area and could be detrimental to the pipeline integrity. A drop in potential (more positive) in type 2 areas can also mean the pipe coating or the soil has a high electrical resistance at this location. The drop in potential can be rectified by increasing the amount of current supplied to the pipe at the nearest test station. Type 3 relates to locations where both the ON and OFF potential has gone beyond the -850mV criterion. The protection here is inadequate and more current is needed to overcome this drop. The indication can also be used to infer a coating or pipeline defect [48]. The CIPS primary purpose is to assess the performance of the CP. While it can

give indications of coating defects and possibly corrosion, it needs additional techniques for validation.

2.3.3 Direct Current Voltage Gradient (DCVG)

In the previous section, it was highlighted that CIPS uses pipe-to-soil potential as the variable for measurement. In a DCVG survey, the main focus is on the soil-to-soil potential survey. The technique is also conducted above ground along the pipeline but in contrast to CIPS, no electrical connection is needed to the pipeline. The main purpose of a DCVG survey is to locate coating faults or defects. A DCVG survey is not only used to find locations of defects but to also measure their severity in terms of their size. The survey is conducted in close intervals much like the CIPS. Due to this, it is ideal to conduct both techniques on the same pipeline as part of the fulfilment of an ECDA phase 2. The requirement is that for an ECDA, at least two aboveground inspection techniques are to be applied for indirect measurements [8].

Consider a scenario where a pipeline (bear with no protective coating) is cathodically protected by an ICCP system. Currents are flowing from the anode to the cathode through the soil and onto the pipe. The movement of currents in the pipeline (cathode) produces voltage gradients in the pipe itself and the surrounding soil. The voltage is at its peak at the pipe and gradually reduces as one moves away from the pipe (into the soil). The gradual reduction in voltage is called the voltage gradient. This voltage

gradient is regarded as the difference between the pipe-to-soil potential relative to remote earth. Voltage gradients are only seen at coating defects (resembling the bare pipe mentioned above) along the pipeline. This is because the current from the anode flows to these imperfections thus increasing the local potential. In a perfectly coated pipe, currents still flow to the pipe through the coating (all coatings are organic hence not 100% insulating) in small quantities but these potentials are negligible. The DCVG reading of voltage gradients on top of a perfectly coated pipe will indicate zero (very close to the remote earth potential). The DCVG equipment consists of a voltmeter and normally two walking canes (similar to CIPS). Attached at the bottom of these walking canes are two copper – copper sulphate, Cu/Cu SO4, reference electrodes, one on each side.

At defect locations the pipe to soil potential will be high because of the current flowing to it. The DCVG surveyor will pick this up with his voltmeter due to the fact that there are potential differences between the soil potential and the remote earth potential. This is what is meant by soil-to-soil potential. The defect will produce a spherical voltage gradient which radiates away from that defect. The surface of these voltage gradient spheres (also called equipotential lines) represents a constant potential. Voltage gradient which are close to the defect are closely packed together but as the distance increases away from the defect the voltage gradient tends to be spaced much further apart. This indicates the rate of voltage drop is abrupt near a defect but slowly levels off further away from the centre. See Figure 2-4.

Due to the voltage gradient produced by the defect, a DCVG surveyor is able to read the voltage drop and pinpoint the location of defects. The above ground surface is considered to be the in-plane voltage gradient as it cuts through the spherical equipotential lines. Figure 2-4 shows an illustrative example of these equipotential lines.

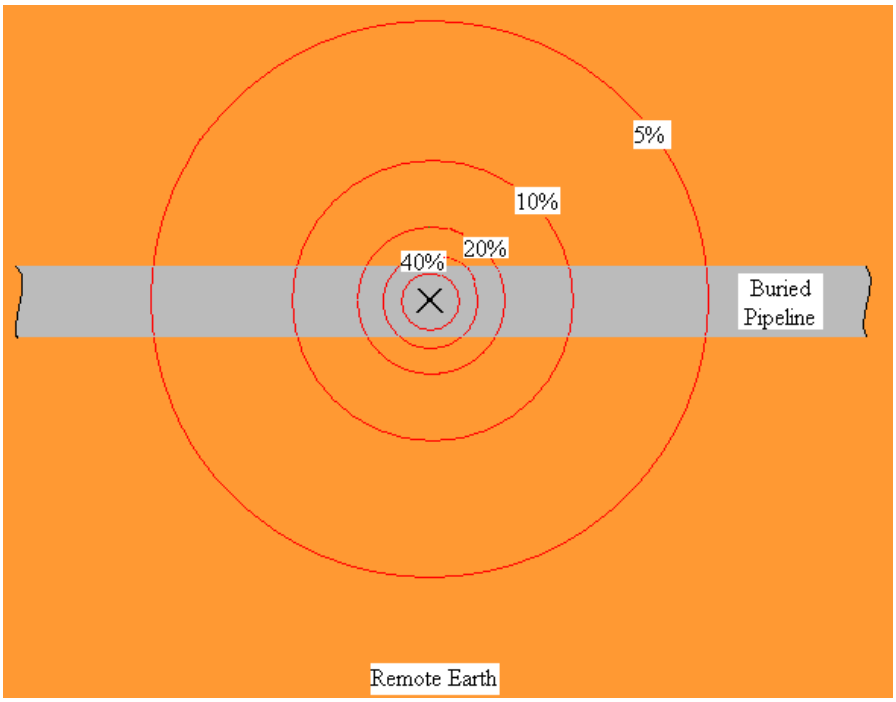


Figure 2-4: Plan View of The In-plane Voltage Gradient Spheres (equipotential lines) Generated at the Ground Surface (image taken from DC-Voltage Gradient (DCVG) Surveys Using MCM's Integrated Pipeline Survey Test Equipment and Database Management Package)

The percentages in Figure 2-4 represent the potential as a percentage of the maximum potential directly above the pipe. As one moves away from the centre, the potential percentage drops quite quickly near the centre then drops more slowly.

A DCVG survey is conducted by first interrupting the current flow with an *interrupter*. The interruption should have an asymmetric cycle which normally consists of a longer ON potential and shorter OFF potential. The ON potential during the interruption ranges from 4000mV to 6000mV. The high potential is used to aid the DCVG surveyor in detecting defect equipotential lines and to have ample time to read the movement of the pointer on his voltmeter. Also, the interrupter is used to capture the DCVG voltage which is the true voltage generated by the CP current flowing to a defect. DCVG voltages are useful in assisting the surveyor to interpret the severity of defects. More on this will be discussed in the following. The survey is done along the pipeline route by initially setting the voltmeter to zero. The surveyor walks the pipe route with the two walking canes and their reference electrodes touching the ground surface one after the other. The left-hand probe will be placed directly on top of the pipe while the right-hand probe will be placed 5 feet away. As the surveyor approaches a defect, the needle on the voltmeter will start to swing either in the positive or negative direction based on the direction of the flow of current. This swinging effect is due to the ON and OFF cycle applied by the interrupter with the OFF potential showing a zero reading and the ON potential illustrating the soil-to-soil potential for that location. The closer the surveyor is to the defect the higher the voltage reading will be. When the defect is located directly underneath the surveyor, the needle movement will stop swinging and will indicate the maximum potential reading. As he walks on, the needle will start swinging again but with different polarity. This is also due to the direction of current flow which flows in a different direction from before the defect. If the

surveyor walks away from the defect, the reading will go back to zero. The swinging action and change in polarization allow the surveyor to determine the exact location of the coating defect. It has a precision of up to 6 inches from the epicentre.

After the defect is located, it is essential to determine its severity. The first step is to quantify the total millivolt (total mV) across the equipotential lines. This is done by initially recording the maximum millivolt (max mV) which is the potential directly above the defect where the left-hand cane is directly (ideally) above the defect and the right-hand cane is 5ft away (laterally). After the max mV has been recorded, the left-hand cane is then positioned at the previously positioned right hand cane while the right-hand cane is again spaced 5ft away from the left cane. This is repeated several times until the potential reaches zero. All the different potential readings from the different positions are then summed (Total mV) and divided by the average potential reading at the defect location between the test posts (if the defect is midway between test posts). If the defect location is not midway, linear interpolation is used as an approximate for the defect's IR drop. The calculated %IR value is given in the following equation.

$$\%\boldsymbol{IR} = \frac{\boldsymbol{Total\ mV\ (Soil - to - Soil)}}{\boldsymbol{IR\ Drop\ at\ location\ (Pipe - to - Soil)}} \bullet \mathbf{100} \qquad \mathbf{2.5}$$

Depending on the percentage IR obtained, the surveyor may classify the percentages into categories. The categories serve as basis whether further

examination is required which may include excavating the site where the defect was discovered. Category 1 is considered a small defect. The percentage range runs from 0 to 15%. Category 2 then follows classifying the defect as medium. The percentage range is from 16% to 35%. Lastly the medium-large category is category 3 with range from 36% to 60%. For defects found greater than 61%, immediate action is to be taken involving excavating the site and conducting a thorough inspection of the defect. It must be kept in mind during a DCVG survey that the technique is only an indicative method. The calculated percentage IR is only an indication to the surveyor and should not be taken as absolute. The categories mentioned earlier are further explained in the following:

- Category 1: 1 to 15%IR – The criticality of coating defects in this category is low which suggests repair is not needed. The CP protection current should be sufficient to protect the affected areas.
- Category 2: 16 to 35%IR – Coating defects in this category are similar to the previous category which considers it as low threat. Repair of defects within this category is recommended but not necessary. Due to the fluctuations of the output current coming from the CP system, defects are monitored for its growth as it might end up in a situation where the amount of protective current supplied from the CP are insufficient.
- Category 3: 36 to 60%IR – Unlike the previous two categories, defects in this category are considered serious and repair is recommended. Monitoring of these defects is highly recommended as

they are large consumers of CP current. Additionally, the fluctuation of the supplied CP current has the potential to undermine the structural integrity of the pipe.

- Category 4: 61 to 100%IR – This category of defects requires immediate attention for repair. It is a very high consumer of CP current and coupled with the possibility of its supply being low, corrosion will develop, and which will jeopardise the safe operation of the pipeline.

2.3.4 Alternate Current Voltage Gradient (ACVG)

The ACVG technique borrows the same principle from the DCVG assessment. Its objective, similar to the DCVG inspection, is to detect coating defects. The main difference is that alternating current is used in place of direct current. Alternating currents are generated by a transmitter of high or low frequencies. Other differences are that instead of having electrodes handled by the inspector, the electrodes of an ACVG inspection are mounted on an A-Frame. The width of the A-Frame is approximately half the distance of the manually handled electrodes. The DCVG and ACVG techniques are the most accurate methods for locating coating defects compared to other indirect techniques such as PCM. For the sizing of defects, DCVG is the only method which can reliably indicate a defect's severity.

2.4 Research on External Corrosion of Buried Pipelines under Cathodic Protection

Soil has the ability to affect the integrity of underground structures such as pipelines, foundation pilings and water mains. The primary cause of deterioration of integrity for these structures is soil corrosion. According to Yahaya et.al [49] soil corrosion is the deterioration of metal due to the chemical, mechanical and biological action of the soil's environment. Research has been conducted [50–53] which investigated and identified the factors that contributes to the external corrosion of buried pipelines. Factors such as the soil texture, moisture content, the soil's pH, the degree of aeration, temperature and the resistivity were all found to be a contributing corrosion factor.

An investigation by A.I.M. Ismail [54] showed that corrosion occurs in buried structures due to the unfavourable interaction between soil and water. The paper also highlights that the particle size of the soil, swelling, shrinkage and clay mineral content all having contributions towards corrosion. Also on the theme of soil characteristics, Benmoussa 2006, [55] found that corrosion in buried pipelines is influenced by the soil resistivity, pH value, moisture content, the temperature of the surrounding environment and the chemical composition of the soil itself. The experiments showed that steel corrodes more readily in acidic environments and at higher temperatures. Although pipelines are inhibited against corrosion by cathodic protection with a -0.850 mV potential, corrosion still occur.

The variable soil resistivity has become an important factor in determining the corrosion rate of a buried pipeline. Investigative work has been done by [56] which shows that corrosion reaches its maximum at 65% of the moisture content of the soil. This can be termed the *critical soil moisture content* (with respect to corrosion). Further work by [57], [58] has classified soil corrosivity by the values of the respective soil resistivity. However, in the work reported by [59] it is highlighted that the soil resistivity factor is not the only factor in determining corrosion. Pipeline corrosion is a random phenomenon and could be governed by more than this parameter. Apart from soil resistivity, corrosion can also be due to microbes such as sulphate reducing bacteria (SRB). This phenomenon was first observed by [60]. After this discovery, the pipeline industry started to recognise that microorganisms have an effect on corrosion. The corrosion term is sometimes referred to microbial induced corrosion or MIC for short.

Various research has also investigated the corrosion underneath coatings at locations of coating defects and disbondment. The phenomenon is termed as crevice corrosion and can also lead to stress corrosion cracking of the pipeline. Crevice corrosion is a localised corrosion defect where water or other solution is stagnant within the crevice gap. Within these gaps, a restricted oxygen reduction (cathodic) reaction thus produces an anodic environment instead. This creates a highly corrosive microenvironment which is conducive for further metal degradation [61]. At coating defects, the edges of the defect tend to peel away causing the detachment of the coating

from the pipeline. If excessive protection current is applied to the pipeline, the disbondment will grow as the alkaline state of the area interacts with the adhesive resulting in its ineffectiveness. Disbondment around a coating defect seems to be of little concern since the coating is still present. However, sometimes the CP current is unable to reach these crevices due to the shielding effect of the coating. The size of disbondment is also said to be dependent on the level of cathodic protection current, the type of coating and the species present within the affected area [62]. Additionally, when there is no coating defect but disbondment of the coating is present, the shielding effect of the coating reduces the effectiveness of the CP protection current but no corrosion is observed due to the coating resistance isolating the metal substrate from the surrounding environment [5].

2.5 Previous Research on Prediction of Coating Defect Area

US federal regulations [10] have stipulated that pipeline operators should produce criteria for the identification and documentation of indications from an indirect assessment technique which will be considered for further assessment (direct assessment through excavation). The criteria also serve to define the urgency of the subsequent inspection based on the documented and identified indications. One of these criteria could be based on the defect's severity. As mentioned before, to a certain extent the current DCVG tool is inaccurate in classifying defect sizes. Prioritisation can only be done when the predicted size of defects is accurately represented. Hence it

is imperative that a working method is established to comply with the regulations and to promote safety and financial prudence.

The determination of coating defect areas on pipelines has not yet been a popular research theme in the pipeline industry. Although this is crucial, the industry tends to rely on expert judgement on where to excavate based on DCVG indications, a risk-based profile of the likelihood of corrosion activity and on pre-assessment data. Efforts were made in correlating indirect assessment and the direct assessment data in the form of a statistical model. The most notable of this is [63]. In this paper, a linear quantile regression was used to model the relationship between the coating defect area and its possible contributors. The paper also sheds light on the challenges faced by pipeline operators when interpreting DCVG indications. The contributing factors are also influenced by other uncontrollable variables such as stray currents and interference from overhead AC lines.

J.P,Mckinney [64] has produced a model which estimates the coating defect area based on simulated data. The approach taken is deterministic in nature using the Finite Element Method (FEA). The model was constructed with the help of in house software developed by the University of Florida, called CP3D. The software modelled pipelines under cathodic protection environment along with the surrounding electrolyte. Most notable findings include DCVG indications increasing as soil resistivity and defect size increase. This shows that in theory, %IR depends on the nature of soil

resistivity where higher soil resistivity values correspond to higher %IR readings.

Moghissi, [65] has identified that there is no simple solution to prioritizing coating defects for further assessment. Based on the indirect assessment of the ECDA process, Moghissi, [65] has collected data from the CIPS, DCVG and current attenuation assessments. These data were used to derive basic formulations to model the relationship between coating defect area and its possible contributing factors. The approach taken here uses similar methods (FEA) as of those found in an earlier work by McKinney [64].

On the other hand, as was pointed out by [9], indirect assessment indications, particularly the DCVG, are a result of unknown factors besides coating defect area. These factors play an important role in interpreting DCVG indications. Hence, the interpretation of DCVG signals can be erroneous if the factors are not determined accurately.

A report produced by CC Technologies [66] Inc for the PHMSA - US Department of Transportation Pipeline and Hazardous Materials Safety Administration reports on an investigation into the accuracy, resolution and the limitations of techniques being used in the pipeline industry for their indirect inspection. Coating defects and disbondments on typical pipeline coatings were looked at and were analysed based on the differing techniques. The different aboveground indirect techniques were the DCVG, ACVG, PCM, C-Scan and the Pearson survey. Inspections were done over 3

sites with pipe size ranging from 22 to 32 inches. Results from these techniques were analysed and compared. It was determined that the DCVG technique is the most accurate amongst all other methods. It was said that the technique is accurate in locating defects, but considerable effort was made in sizing these flaws. Another interesting finding with regards to DCVG is that the %IR reading did not show any proportional relationship with the area of coating (metal substrate exposed to the environment). The researchers thought that this may be due to the linear assumption of the attenuation of the signal between test posts – when the distance between test posts is wide and the coating condition is generally poor, the linear assumption does not hold as well.

An article by Marcel Roche et.al [67] describes the author's and the author's company (Total) experiences in coating failures. Based on these case studies, the article proposes an explanation why these failures occur, and it investigates the parameters that may give rise to coating failures such as the temperature, coating type and the application procedure of the coating (surface preparation). It also highlights the comparison of a DCVG technique compared to an In-Line Inspection (ILI) for the detection of corrosion of pipelines. The results showed that DCVG indications correspond to 59% of the total corroded areas found by the ILI technique suggesting that some of these areas received some form of effective cathodic protection. It concludes by saying DCVG is able to locate dangerous situations resulting from coating disbonding. However, the technique needs further evaluation to verify its effectiveness.

CorrPro inc. [68] has submitted a white paper to the PHMSA highlighting possible improvements to the ECDA process. The document is produced in three parts namely the improvement of an ECDA for cased pipes, severity ranking of ECDA's aboveground indirect inspection techniques and the voltage drops for paved areas. The one that is of particular interest here is the second part which relates to the accuracy in classifying and prioritising indirect inspection techniques. This report bases its argument on the NACE standard [8] and says that the classification and prioritisation for indirect inspection i.e. DCVG, ACVG, CIPS etc. indications are intentionally made general for initial guidance. The reason for this is to get pipeline operators to further refine the criteria based on the ongoing ECDA process (which is iterative in nature). However, when no refinement is done, this will lead to incorrectly identifying and prioritising indirect inspection indications under a wide variety of conditions. The report continues by saying that the classification and the prioritisation criteria set out in the current NACE SP0502-2008 "are imprecise and ambiguous". They are subject to different interpretations and the resulting classification and prioritisation are different from "operator to operator, pipeline to pipeline and sometimes location to location". Due to this, it is also quite normal to see inconsistencies between the indirect inspection's indication and the result found during direct assessment. For example, higher %IR readings are paired with low or zero coating defect area and *vice versa*. In numerous cases, necessary excavations were not done due to the lack of accuracy of the classification and prioritisation scheme. This will only get noticed when failure occurs or

when for some reason (alternative ways) it is found out that excavation should have been performed. There are some pipeline operators who do follow the route of refinement. However, the methods employed by these operators are not well understood, documented and consistently applied. The following illustrates the daunting task of classifying and prioritising indirect inspection's indications which was summarised in the report.

- The determination of the accuracy of the aboveground indirect inspection techniques requires excavation to be done at every defect found (from the indirect inspection) to see whether the indication proves to be true (based on defect location and size).
- The actual defect found during direct examination (excavation) tends to be smaller than initially thought based on the classification procedure. This leads to the classification scheme to be conservative in terms of predicting the size of defects.
- The inevitable effects of stray currents will cause uncertainty in prediction of size of defects. An indication showing minor DCVG indications with added interference from stray currents is far more problematic than a severe DCVG indication having the same kind of interference from stray currents. This is due to the smaller defect being treated as minor priority and will go "unnoticed" whereas the effect of stray currents can be catastrophic.
- Corrosion rate is not a function of coating defect size (not always).
- Technology is needed for the interpretation of the indirect inspection indications which must lead to tools which are state of the art in terms of technology.

- The need of mathematical tools to consistently classify and prioritize indications from indirect inspection techniques.

In 2012, [69] researched on the aboveground indirect inspection techniques and how it correlates with ILI data and direct assessment data. All the data available were juxtaposed to see how corrosion relates to coating defects. By doing so, making an assessment on the overall condition of the pipe's integrity is possible. The indirect inspection techniques considered were DCVG, ACVG and CIPS and were set against the results of the ILI. Alignments made here were later compared with results of the direct assessment. The result of the alignment of data found that ACVG performed better than DCVG for coating defect detection. This is due to ACVG having higher detection sensitivity which makes detection of smaller defects possible. Enhanced sensitivity of locating defects is also due to the unique signal (Alternating Current) used by ACVG which is easily discernible from other types of signal. The DCVG technique on the other hand relies on the CP current for its signal. Based on the paper's finding, these signals do not generate large enough potentials for the technique to detect. The locations of coating defects found by the ACVG and the DCVG techniques and the metal loss identified during direct examination points to low correlation. This shows that the CP protective current is effective in preventing corrosion. However, the findings also highlight that detecting coating anomalies alone is insufficient in proving a pipe's integrity, and that a corrosion study is needed to assure complete protection. It was also found that future prioritisation of excavation sites is highly questionable since the criteria for

excavation of the anomalies found by the ACVG technique (severe category) did not align with the results found by the ILI. Observations from the direct examination also did not align with the results from the ACVG. Therefore, future excavation of those sites found by ACVG is not warranted. Further findings suggest that DCVG is better suited than ACVG for classifying severe coating defects since the signal it uses is directly generated from the CP current. The report concludes that combining all the available above ground inspection techniques (namely the DCVG, ACVG and CIPS) is crucial in providing a comprehensive picture in determining the condition of a particular pipeline.

Work was done by [70] which illustrates the usage of DCVG as a quality control tool during the construction of new pipelines. Coatings are susceptible to damage as the result of improper handling and the effect of excavators scraping some of the coatings off during installation. The investigation of coating defects was done on 2 newly constructed gas pipelines in South Africa. Different criteria for excavation were set due to the likelihood of small coating defects appearing (as opposed to larger defects occurring throughout the service lifetime of the pipe). The investigation concluded that the DCVG technique determines the locations of coating defects quite accurately and should be used as a control quality tool for newly constructed pipes. However, the paper does not explicitly mention the accuracy of defect sizing.

2.6 Gap in Literature

Most of the work done in the context of coatings under cathodic protection revolves around the study of the coating itself and the corrosion mechanism. Research effort has been poured into the phenomenon of corrosion underneath disbonded coating both experimental and theoretical (through modelling). The scope of these studies does not take into account real life datasets and the method of arriving at a conclusion is deterministic (not probabilistic). Moreover, no quantification of pit depths at a coating defect (not underneath) and how it relates to contributing factors was ever modelled. Additionally, these studies are very “localised” in the sense that they cannot be readily applied in the field, e.g. where to excavate based on the results obtained. Investigations were confined to the lab or to producing theoretical simulated data.

As one can observe, literature on predicting the size of coating defects is sparse. Considerable work has been done by the team at the University of Florida (UoF) under the guidance of Professor Orazem on modelling of the cathodic protection system and the prediction of the size of these defects. The approach taken is highly theoretical where analysis was done using in-house finite element software. The data for analysis was also simulated (computationally) and the approach taken for modelling is deterministic (through FEA). Other such efforts were made by [65] which used the same modelling technique as the UoF approach. The simulated method lacks the ability to accurately mimic real-life phenomenon. This and the deterministic

approach gives highly constrained results which lead to conclusions that allow for only small margins of error.

To counter the deterministic method highlighted above, a statistical approach was used for coating size prediction in [63] which employed the quantile regression method to characterise the distribution of the coating defect more comprehensively. In this work, the relationship between variables was assumed to be linear. However, based on most of the findings above and after screening the data used for this thesis, it is suggested that the relationship of coating defect area and the corresponding IR drop is not linear. This is further corroborated by the findings from investigative work by CC technologies [66]. The research also did not investigate the contributing factors to the generation of the IR drop. Additionally, with a linear approach, the predictions can sometimes land in areas which are outside the bounds of the predicted dependent variable, e.g. %IR values of more than 100%. Other gaps include inconsideration of interaction between the variables of an ECDA indirect and direct assessment procedure, a simplified approach that can be readily used for inspectors in the field and finally the incorporation of current results for the subsequent ECDA inspection.

This thesis intends to address these gaps and issues by developing models which:

- use data from real life ECDA projects;
- establish a method in estimating pit depths based on contributing factors by using a statistical approach – the study of corrosion pits

seems to suggest the evolution of its growth are stochastic. Statistical methods allow for the introduction of uncertainties which will better explain the uncertain nature of pitting;

- explain the relationship between coating defect size and its contributing factors through a non-linear approach;
- include variables which has interactions between them and see how they affect the final predicted value;
- treat the current finding as prior knowledge to be included in subsequent ECDA projects on the same pipeline;
- predict the dependent variable within a specified range (outcome);
- can be used as guidelines for pipeline operators in the decision-making process of selecting the location of excavation for further direct assessment.

2.7 Middle Eastern Pipelines Data

A Middle Eastern Oil Company (MEOC) has awarded a contract to TWI Ltd. to conduct an External Corrosion Direct Assessment work on its network of pipelines. There is a total of nine (9) pipelines to be looked at, all of which are non-piggable, hence the ECDA approach. The ECDA work shall comply with the ANSI/NACE SP0502-2010: Standard Practice Pipeline External Corrosion Direct Assessment Methodology. Based on this standard, the ECDA was carried out according to the sections and data were extracted for the purpose of this research:

2.7.1 Pre-Assessment

The data in this section includes the design data of the pipe which include its philosophy, material selection, pipe characteristics etc. Historical operations activity is also included in this section. It was found that the total length of the nine pipelines covers over 300 km. The pipe sizes are from 26 to 42 inches. Operating pressure is from 8 to 17 Bar. The grades of these pipes are as accordance with the American Petroleum Institute (API) which is API5L-X52 and X60. Working pressure of the pipes ranges from 40 to 60 degrees Celsius with a 400 m^3h^{-1} to 1520 m^3h^{-1} fluid flow rate. Coatings were applied on all the nine pipes which is cold wrap, coal tar or polyethylene. These coatings are organic in nature.

2.7.2 Indirect Assessment

This section of the ECDA process specifies the indirect tests that should be performed on any given pipeline. Techniques such as the CIPS, DCVG, ACVG and Pipe Current Mapper (PCM) were conducted in a series of tests on the MEOC pipelines to gain information on the condition of the pipeline and to identify locations of coating defects. The DCVG technique was identified as the most potent and reliable source of information and was used for the work presented in this thesis. The values of the %IR based on the DCVG indications found for the MEOC pipelines was annotated and later paired with its associated excavation site (direct assessment).

2.7.3 Direct Assessment

Direct assessment of defects provided a lot of useful data. After the identification of coating defects and calculation of their severity (based on %IR), decisions were made on where to excavate to further analyse the defects. The decisions were based on the magnitude of the %IR and the pre-assessment data. As can be seen here, the decision relies on expert judgment of the engineers (subjective) which can be erroneous if not all the contributing factors are considered. At excavation sites of the MEOC pipelines, data collected are the soil resistivity (based on the NACE 4 pin method), the depth of buried pipe, the material of cover, the pH of the soil and the pH of water underneath the coatings. For locations where corrosion activity is observed, the depth of the corrosion pits was measured using ultrasonic thickness measurements and pit gauges. The size of the coating defects was also measured and summed (at 1 excavation site) to become the Total Coating Defect Area (TCDA). Photographs were taken of the coating defects as backup for later verification. Deposits underneath coatings were also annotated where present. The amount of deposit underneath the coating in terms of area is divided with the TCDA to obtain a percentage value. All of the highlighted data collected at this phase was used as potential variables for the model.

Figure 2-5: Examples of Direct Assessment. Picture on the top showed a coating defect detected at PR10. Picture on the bottom illustrates an example of an excavated site

Chapter 3

Statistical Methods Used for Analyses of the MEOC ECDA Data

3.0 Introduction

Diagnostics and prognostics algorithms are used to determine the maintenance schedules for a condition-based maintenance (CBM) strategy. CBM can be considered an efficient and cost effective way of maintaining structures [71]. The development of the prognostic algorithms is largely due to the outcome from the diagnostic assessment. The diagnostics informs the user on the current state of the structure, whether a defect is found and if so, characterises it. Prognostics are simply a projection on the future good or

bad behaviour of a structure which gives a value based on the future anticipated damage and its remaining useful life (RUL).

Prognostics modelling can be divided into three categories. The first approach is the data driven approach. For this approach to be successful data on the current state of structure, i.e. condition monitoring data (CM data) and event data, i.e. run to failure data, are used. These data are normally termed the training data. The approach consists of characterising the damage based on the CM data and predicting future trends, e.g. future behaviour of damage based on the training data. The physics-based approach relies on the availability of a physical model and using measured damage data (CM data) in combination with the physical model to predict future trends and behaviour. The third approach involves a hybrid combining the previous 2 approaches [72].

The observation of data in a data driven approach is for defect characterisation in terms of its evolution. Once a trend is established, future extrapolation predicts future defect growth. No physical model is used for the prediction. The prognosis is essentially following what the data is saying. Data driven approaches are divided into two main categories, an artificial intelligence (AI) approach and a statistical approach. AI employs techniques from machine learning algorithms such as neural networks (NN) or fuzzy logic. In a statistical approach, mathematical models such as multiple regression, quantile regression, Bayesian Quantile Regression, etc. use the weighting of coefficients to extrapolate the trend for future predictions.

Many mathematical models have been developed in recent years in an attempt to describe and predict failures of pipelines. Engineers tend to look at most of these models with scepticism because of the inconsistencies found between predicted results and the actual field data [73]. Possible reasons for them to be inconsistent are:

- no mathematical model is accurate in mimicking the true mechanisms of pipeline failure.
    - These models were developed based on conditions that they were developed for, i.e. they do not account for other unknown factors which might contribute to failures. This in turn will limit each model to a particular scenario (overfitting). With the addition of more data, this problem will be minimized. However, the engineer should have sound knowledge to these limitations and where to apply it;
- the data obtained from the field are never exact.
    - Data from the field are never exact. This might be due to operator error and factors unknown to the operator conducting the test. Sometimes, these data are incomplete which makes using the model impossible. Similar to the above, with more data collected this issue will taper off as more and more data will concentrate on the true parameter value.
    - Uncertainties that come with the collected data should be well understood by the engineer. These uncertainties on occasion will have an effect on the final outcome of the model's result;

- operator knowledge of the system is often lacking.
    - The researchers developing the models often do not possess the same (or required) knowledge as the field engineer. This will limit the model's scope.

In this research, the data collected is from an ECDA procedure in the form of indirect and direct assessment measurements. Due to the lack of a physical model for the prediction of coating defect size, a statistical regression approach is taken for the analyses of trends and prediction of defect growth for the MEOC data.

The MEOC dataset consists of only 200 plus data points (excavation sites) with the number of variables being large (after the introduction of polynomial and interaction terms) which makes dividing the data into a training and test set for validation purposes seems impractical. Additionally, the %IR reading is inconsistent (more on the inconsistencies is discussed in the following) with the TCDA measurements, which has the potential to contribute to errors if the dataset is divided. To counter this inadequacy, more data is needed (from future inspections or other similar ECDA projects) to increase the number of data points which can outweigh the number of variables and also to "drown out" the lack of trend observed in the MEOC dataset. Therefore, a method of incorporating new data for future inspection results was modelled through the Bayesian technique so that possibly, better prediction of the variable of interest will be achieved when new data comes in.

As was discussed in chapter 2, the determination of the severity of the defect found with a DCVG survey is relative. The excavation that follows does not necessarily produce positive results, as sometimes the defect is found to be the opposite of what was predicted by the aboveground indirect inspection technique. This is what was observed in the MEOC dataset. The inconsistency is perhaps due to its (DCVG) reliance on voltage gradient equipotential lines which in turns rely on the current flowing to coating defects. The current flow is thought to be based on a number of factors such as the level of CP current supplied to the pipeline, soil resistivity, soil compactness, the depth of the pipe buried, type of soil, type of coating, pH of soil, etc. To model this phenomenon, ideally all the values for these factors are made available to the modeller. With the MEOC dataset however, factors such as the CP current level and the soil compactness were not available (not taken during inspection). This is normal for an ECDA project such as this where some measurements were not taken due to cost or time. Modelling had to make do only with the available data in the project's report.

Interaction effects between these variables could also contribute significantly to the prediction of the variable of interest. Due to this, the interaction effects were considered to see if there are any valuable findings which could be of significance. Another factor to consider is the introduction of subjective interpretations from the inspectors inspecting the pipelines. Subjectivity was used particularly to describe categorical variables such as the characteristics of soil, water content or the amount of deposits found underneath a disbonded coating. With wrongly attributed characteristics of these variables

to properties, there is the possibility of incorrect conclusions. The research begins with the introduction of all the variables made available to the author and gradually reduces the amount of subjectivity within these variables by not including them in the later models (BQR and LQR).

Uncertain data is also included in the models constructed within this thesis. Variables such as TCDA which has a set of readings with a value of %IR gives an idea of the apparent uncertainty within the data set. Taking the mean of such variables is inaccurate and the variation in these values will produce different outcomes. Quantile Regression (QR) solves this issue by capturing the whole of the TCDA's distribution and characterising the distribution more comprehensively. A Bayesian approach to QR (BQR) is also included due to the iterative process of the ECDA and the possibility of incorporating future inspection results into the current findings. Lastly the Logistic Quantile Regression (LQR) is applied to make sure the predicted outcome is within its intended bounds. The following sections highlight all of the statistical techniques used for the analyses of the MEOC data.

3.1 Statistical Methods used in This Research

3.1.1 Correlation

The Pearson's correlation measures the linear relationship's strength between two variables. It also measures the trend of the variable in terms of direction. The correlation coefficient value ranges from -1 to 1. As this value approaches -1 or 1, the degree of relationship is said to be strongly

correlated. Otherwise, if the correlation value approaches 0, then the relationship is regarded as weaker. The negative and positive signs of the correlation value illustrate the direction of the relationship with negative values sloping downwards and positive values climbing upwards. The formula for the Pearson's correlation coefficient is

$$\boldsymbol{r} = \frac{\boldsymbol{N}\sum \boldsymbol{XY} - \sum(\boldsymbol{X})(\boldsymbol{Y}))}{\sqrt{[\boldsymbol{N}\sum \boldsymbol{X^2} - \sum(\boldsymbol{X^2})][\boldsymbol{N}\sum \boldsymbol{Y^2} - \sum(\boldsymbol{Y^2})]}} \qquad \textbf{3.1}$$

where r is the Pearson's correlation coefficient, N is the sample size and X and Y are the variables of interest. The coefficient of determination (CoD) is similar to the Pearson correlation. All one needs to do is to square the correlation coefficient value to obtain the coefficient of determination. The squared value can be thought of as a percentage illustrating the portion of the data points which the regression line is able to explain. If for example, the CoD value is calculated to be 50%, the interpretation would be that the regression line would be able to explain only half of the data points while the other half is attributed to error. CoD values also represent the goodness of fit of the model. The higher the coefficient value, the better the fit (explains more of the data points).

3.1.2 Quantile Regression (QR)

Regression is a statistical technique to examine the existence and extent of relationship between a dependent variable and other independent variables [74]. In its simplest (bivariate) form, regression shows the relationship between one independent variable (X) and a dependent variable (Y):

$$\boldsymbol{Y = \beta_0 + \beta_1 X + \varepsilon} \qquad \textbf{3.2}$$

The regression parameter β_1 shows the magnitude and direction of that relation, and an intercept term β_0 captures the status of the dependent variable Y when the independent variable X is absent. A final error term ε captures the amount of variation that is not predicted by the slope and intercept terms. More sophisticated forms of regression allow for more independent variables Xs, interactions between the independent variables and other complexities in the way that one variable affects another.

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X + \ldots\ldots\ldots\ldots \beta_p X_p + \varepsilon} \qquad \textbf{3.3}$$

One usually assumes that the error term ε has zero mean and uses models (3.2) or (3.3) to model or predict the conditional mean of Y given Xs, and then uses the ordinary least squares (OLS) to estimate regression parameter $\hat{\beta}$,

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{argmin}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left[\boldsymbol{Y}_{\boldsymbol{i}} - \boldsymbol{\beta}_{\boldsymbol{0}} - - \ldots \ldots \ldots \ldots \boldsymbol{\beta}_{\boldsymbol{p}} \boldsymbol{X}_{\boldsymbol{p}}\right]^{\boldsymbol{2}} \quad \textbf{3.4}$$

However, if the distribution of Y is asymmetric or skewed as in the case of our data set (see Figure 3-1) or even heavy tailed, QR is better than OLS-based mean regression for the relationship measurement and model-based prediction [75]. As in this case, the mean is no longer the best representative of the underlying distribution. Instead, quantiles such as median are of interest. In particular, extreme quantiles, which measure the tails, are often used to assess risk or remaining life.

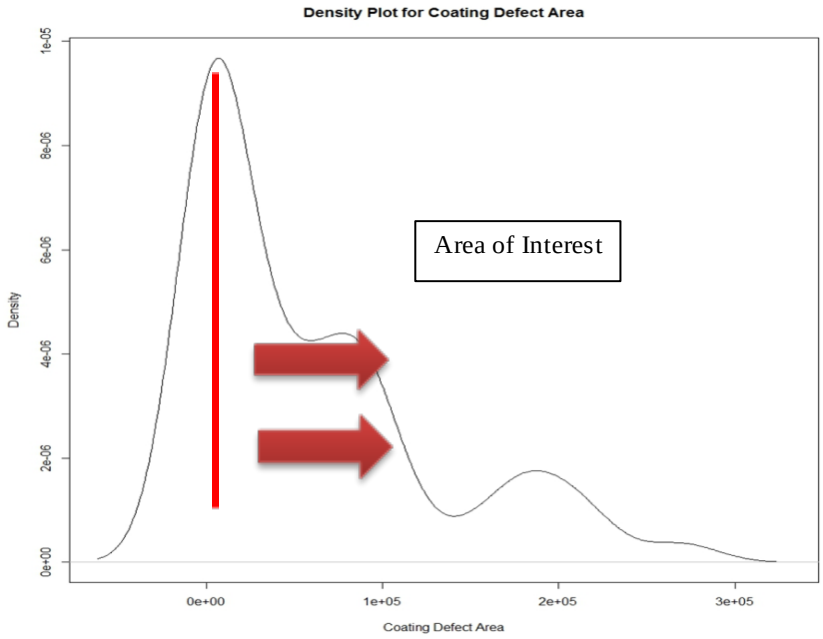


Figure 3-1: Probability Density Plot for TCDA

Figure 3-1 visualizes the TCDA in its probability density function (PDF) form. Most of the readings are skewed to the right side indicating the bulk of the density is at zero. The right-hand side of the PDF is where the QR is

effective as it can model certain high quantiles of interest (as opposed to the OLS-based mean regression) which can be used to investigate and predict the effect of some factors on large/high TCDA.

The definition and estimation of QR are given below. Let the functions Q_Y,

$$\boldsymbol{Q_Y(\tau \mid X_1, X_2, \ldots . X_p) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2 + \ldots \ldots \beta_p(\tau)X_p} \quad \textbf{3.5}$$

be the τ^{th} $(0 < \tau < 1)$ quantile of Y given $X_1, X_2, \ldots . X_p$, then $Q_Y(\tau \mid X_1, X_2, \ldots . X_p)$ is equivalent to,

$$\boldsymbol{F_Y(Q_Y(\tau \mid X_1, X_2, \ldots . X_p)) = \tau} \quad \textbf{3.6}$$

Where $F_Y(y \mid \bullet)$ is the conditional distribution of Y. The regression variables $\beta_0(\tau), \beta_1(\tau), \beta_2(\tau). \ldots \ldots \beta_p(\tau)$ may depend on τ and can be estimated by,

$$\boldsymbol{\hat{\beta}(\tau) = argmin_{\beta} \Sigma_{i=1}^{n} \left[\tau \cdot I\left(Y_i > Q_Y(\tau \mid X_1, X_2, \ldots . X_p)\right) + (1 - \tau)\right.} \quad \textbf{3.7}$$

$$\boldsymbol{\left. \cdot I\left(Y_i < Q_Y(\tau \mid X_1, X_2, \ldots . X_p)\right)\right] \cdot}$$

$$\boldsymbol{\mid Y_i - Q_Y(\tau \mid X_1, X_2, \ldots . X_p) \mid}$$

where I is termed as the indicator function. Examples of quantiles are, $\tau = 0.5$, $Q_y(0.5 \mid X_1, X_2, \ldots . X_p)$ which is the median, $\tau = 0.25$, $Q_y(0.25 \mid X_1, X_2, \ldots . X_p)$ which is the 1st quartile and $\tau = 0.75$, $Q_y(0.75 \mid X_1, X_2, \ldots . X_p)$ which represents the 3rd quartile.

3.1.3 Least Absolute Shrinkage Selection Operator (LASSO)

The selection of independent variables remains a crucial problem in the case of building a QR model [76]. To have better prediction from the constructed model and hence better prediction accuracy, one has to select the most appropriate set of variables. This can be done through selection of independent variables which have the greatest meaning to the predictive dependent variable selected. This is highly subjective and may also leave out vital information as some of the lesser known meaningful variables are discarded leaving the model potentially less precise. One way to solve this issue is to use a penalized parameter (imposing constraints) which shrinks the less important variables to zero (or close to zero). The constraints are there to limit the model's flexibility in having multiple solutions in a given dataset as can be observed in our MEOC data, thereby reducing the model's ability to overfit. Although this idea was thought to be counterproductive in the sense that it sacrifices the model's unbiasedness (as in the case with the OLS) and reduces the model's variance, the overall model predictive capabilities is improved. The interpretation of the model is also improved by having fewer variables and a much more concise model to deal with.

In the case of the constructed models in Chapter 4, the Least Absolute Shrinkage Selection Operator (LASSO) proposed by Tibshirani [77] was used as the selection operator. The LASSO can simultaneously estimate the parameter value and perform variable selection. The LASSO estimate is the solution to minimize:

$$\min_{\beta} \|\vec{y} - A\vec{\beta}\|_1 + \lambda\|\vec{\beta}\|_1 \quad \textbf{3.8}$$

where λ is the regularization/penalized parameter of the minimized function. Major advantages of the LASSO include eliminating (variable selection) or downsizing the value of coefficients which makes the model more interpretable and accurate. In the case of the MEOC data, coefficients with values less than 1 down to zero after they have gone through the LASSO process are discarded for further generalization. The main goal of this endeavour is to obtain a general rule (simpler but still accurate) for prioritisation of coating defects. The rule or guideline is hoped to be supplementary in assisting pipeline operators to make decisions on where to excavate for further direct assessment of the condition of the pipe.

3.1.4 Bayesian Quantile Regression (BQR)

In classical statistics, assumptions are made on the estimated variables where the value is considered fixed, but the quantity is unknown. We can estimate a parameter using samples from a population and calculate the parameter of interest based on that sample. However, different samples will give different estimates. The distribution of the different estimates is referred to as the sampling distribution. Uncertainty is represented by the confidence interval associated with the estimate. Before the collection of data, the parameter will be within the (1-r) level confidence interval (which is random) with the probability of 1-r. When data is collected, and a new calculation of

the confidence interval is made, whether the parameter is within this new interval or not is unknown and there is no possible way of knowing [78].

Unlike the classical approach, Bayesian inference is a new way of thinking about statistics. The parameter of interest is not fixed but a random variable and we are sure that this parameter will fall within the computed credible interval.

Based on the paper by Yu and Moyeed [79], the τth regression quantile $(0 < \tau < 1)$ can take on any solution, $\hat{\beta}(\tau)$, and is associated to the aforementioned quantile regression minimization problem (minimization β)

$$\boldsymbol{min} \sum_{\boldsymbol{t}} \boldsymbol{\rho_{\tau}}(\boldsymbol{y_t} - \boldsymbol{x_t'\beta}), \tag{3.9}$$

the *loss function* being

$$\boldsymbol{\rho_{\tau}}(\boldsymbol{u}) = \boldsymbol{u}(\boldsymbol{\tau} - \boldsymbol{I}(\boldsymbol{u} < \boldsymbol{0})) \tag{3.10}$$

Yu and Moyeed [79] also show that the minimization of the loss function above is exactly the same as maximizing the likelihood function which is formed by joining two asymmetric Laplace densities (ALD).

The probability density function of the asymmetric Laplace distribution is given as follows,

$$\boldsymbol{f(y;\ \mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma}\ exp\left\{-\rho\left(\frac{y-\mu}{\sigma}\right)\right\}} \quad \textbf{3.11}$$

and based on the y observations $y = (y_1, \ldots., y_n)$, the distribution of the posterior of β, $\pi(\beta \mid y)$ is in the form of the Bayes theorem

$$\boldsymbol{\pi(\beta \mid y) = Likelihood(y \mid \beta) \times g(\beta)} \quad \textbf{3.12}$$

The $\boldsymbol{g(\beta)}$ is considered as the prior distribution of β and $Likelihood(y \mid \beta)$ is the likelihood function. Since minimizing the loss function highlighted above is exactly the same as maximizing the ALD, the likelihood can be written like this

$$\boldsymbol{Likelihood(y \mid \beta) = \tau^n(1-\tau)^n exp\left\{-\sum_i \rho_\tau(y_i - x_i'\beta)\right\}} \quad \textbf{3.13}$$

As for the priors, one can use any prior. But in the absence of a prior (as in the case of this thesis due to the lack of expert opinion and the limited amount of data) Yu and Moyeed [79] proved that a non-informative improper prior yields a proper posterior distribution. In this method, there are no known conjugate priors but with the relative ease in using MCMC with the Metropolis Hastings algorithm, one is easily able to produce the posterior distribution of the parameter(s).

3.1.5 Markov Chain Monte Carlo with Metropolis-Hastings Algorithm (MCMC)

MCMC is used to calculate the posterior distribution from the joint distribution of the likelihood and the prior. The joint posterior from the Bayesian process $\boldsymbol{\pi}(\boldsymbol{\beta} \mid \boldsymbol{data})$, is equivalent to the equilibrium distribution of a Markov Chain which is generated by the MCMC scheme. The validity of this distribution is based on the convergence of the MCMC after the burn in period is surpassed.

An algorithm such as Metropolis-Hastings has become popular as a method in the construction of Markov chains. The underlying technique of this algorithm is based on the simulation technique of *rejection sampling*. A new sample is generated based on the proposal distribution, $t(\beta, \beta^c)$, and is accepted or rejected based on some probability.

The proposal distribution used is a normal distribution which is the default setting within the BQR package of the R software. A new sample, β', is generated (via the proposal distribution which is based on the current state, β^c) and is accepted or rejected based the acceptance probability, $\alpha(\beta', \beta^c)$. This is given by,

$$\boldsymbol{\alpha(\beta', \beta^c) = min}\left[\frac{\boldsymbol{\pi(\beta')\pi(data \mid \beta')t(\beta^c, \beta'),}}{\boldsymbol{\pi(\beta^c)\pi(data \mid \beta^c)t(\beta', \beta^c),}}, \mathbf{1}\right] \qquad \textbf{3.14}$$

The random process of the random walk is used for the generation of β' from β^c and for the assessment in this thesis, the proposal distribution is symmetric. This will in turn make the ratio of $\frac{t(\boldsymbol{\beta^c},\boldsymbol{\beta'})}{t(\boldsymbol{\beta'},\boldsymbol{\beta^c})} = \mathbf{1}$. The resulting acceptance probability is thus,

$$\boldsymbol{\alpha(\beta', \beta^c) = min}\left[\frac{\boldsymbol{\pi(\beta')\pi(data \mid \beta')}}{\boldsymbol{\pi(\beta^c)\pi(data \mid \beta^c)}}, \mathbf{1}\right] \qquad \textbf{3.15}$$

$\pi(data \mid \beta')$ is the likelihood function while the $\pi(\beta')$ is the prior distribution. If the new value β', has a higher probability (in terms of $\pi(\beta')\pi(data \mid \beta')$) than the current value β^c, then we will always accept the new value. If this is not the case, a value, u, has to be generated from a uniform distribution $U(0,1)$, and will be compared to the acceptance probability $\alpha(\beta', \beta^c)$. If $u < \alpha(\beta', \beta^c)$ then we accept the new value. Otherwise we reject it and keep the old (current) value. This process is repeated until convergence is reached where the stationary distribution is an approximate of the target distribution.

The burn in period, which is the time taken for the algorithm to converge, is unknown and thus could be substantial. Normally the first 2500 iterations are discarded but with regards to the work in this research, a 5000 iteration burn-in period is assumed. This is due to the high number of iterations needed to achieve convergence. The author relied on the graphical representation of the trace plot (time series) to indicate whether convergence has been achieved.

3.1.6 Logistic Quantile Regression (LQR)

It is important to define the working knowledge of the LQR to see the full benefits of the technique. The following definition is taken from [80]. Suppose we have a continuous dependent variable Y, with a set of p covariates, $X_1, X_2, \ldots . X_p$, where the dependent variable is bounded by a particular interval, Y_{min} and Y_{max}. Quantile regression is represented by the functions Q_Y,

$$\boldsymbol{Q_Y(\tau \mid X_1, X_2, \ldots . X_p) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2 + \ldots \ldots \beta_p(\tau)X_p} \quad \textbf{3.16}$$

is the $\tau^{th}\ (0 < \tau < 1)$ quantile of Y given $X_1, X_2, \ldots . X_p$, then $Q_Y(\tau \mid X_1, X_2, \ldots . X_p)$ is defined as,

$$\boldsymbol{F_Y(Q_Y(\tau \mid X_1, X_2, \ldots . X_p) = \tau} \quad \textbf{3.17}$$

where $F_Y(y \mid \bullet)$ is the conditional distribution of Y. An example of the most popular quantile is the median where $\tau = 0.5$, $Q_y(0.5 \mid X_1, X_2, \ldots . X_p)$ and divides the dependent variable's distribution into two equal parts with the same probability.

It is assumed that for any quantile τ, there exist a set of variables $\beta_0(\tau), \beta_1(\tau), \beta_2(\tau). \ldots \ldots \beta_p(\tau)$ which is fixed and a known non decreasing function, h (known as the link function) from the interval Y_{min} and Y_{max} to the

real line (the reason for this transformation is to comply with the constraints). This is further defined as,

$$h\{Q_Y(\tau \mid X_1, X_2, \ldots . X_p)\} = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2 + \ldots \ldots \beta_p(\tau)X_p \quad \textbf{3.18}$$

Due to the nature of the dependent variable being constrained, it resembles a probability. This is more apparent with a dependent variable which has binary outcomes (values of 0 or 1). Due to the pre-specified range of the dependent variable, variables of interest such as the %IR and POPD fit this definition perfectly (the values are in percentages). The transformation of the dependent variable is thus,

$$h(Y_i) = \log\left(\frac{Y_i - Y_{min}}{Y_{max} - Y_i}\right) = logit(Y_i) \quad \textbf{3.19}$$

The inverse transform is,

$$Q_Y(\tau \mid X_1, X_2, \ldots . X_p) = \frac{e^{(\beta_0(\tau)+\beta_1(\tau)X_1+\beta_2(\tau)X_2+\ldots\ldots\beta_p(\tau)X_p)}Y_{max} + Y_{min}}{1 + e^{(\beta_0(\tau)+\beta_1(\tau)X_1+\beta_2(\tau)X_2+\ldots\ldots\beta_p(\tau)X_p)}} \quad \textbf{3.20}$$

After the transformation of $h(Y_i)$, the regression coefficient can be estimated through the normal steps of a quantile regression,

$$Q_{h(Y_i)}(\tau \mid X_1, X_2, \ldots . X_p) = Q_{logit(Y_i)}(\tau \mid X_1, X_2, \ldots . X_p) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2 + \ldots \ldots \beta_p(\tau)X_p \quad \textbf{3.21}$$

The inference of $\boldsymbol{Q_Y(\tau \mid X_1, X_2, \ldots . X_p)}$ after the regression coefficients have been found can be made through the inverse of the transform stated above. Bottai et.al. [80] have found that quantiles are not affected by a monotone transformation which makes inference a possibility.

$$\boldsymbol{Q_{h(Y)}(\tau \mid X_1, X_2, \ldots . X_p) = h\{Q_Y(\tau \mid X_1, X_2, \ldots . X_p)\}} \quad \textbf{3.22}$$

It was shown that this is due to $P(Y \leq y) = P\{h(Y) \leq h(y)\}$ for any random variable Y and any non-decreasing (link) function, h. This special property of quantile regression was exploited by [80] and defined the LQR to accommodate dependent variables which has bounded outcomes.

3.2 Summary

This chapter highlighted the various statistical techniques used for the analyses of the MEOC data presented in this thesis. Initial assessments were made with the correlation technique to see if there is any linear correlation between the variables. The results from this assessment were used to determine the approach (linear or nonlinear) taken for the following model. A QR model was later constructed considering the results obtained from the correlation assessment. The interaction effects were also included within the quantile regression model. Due to the high number of variables present in the quantile regression model, a regularisation technique, i.e. LASSO, was used to shrink the models to make them more generalised and keep them simple and concise. The BQR was later opted for the modelling

technique to ensure continuity of the current model in becoming more accurate. Future inspection data is then possible to incorporate making the model more precise. Lastly, the LQR technique was chosen as the method of modelling to ensure that the predicted values are within its intended boundary while keeping the advantage of the QR method. As was mentioned earlier, the set of variables used for each modelling technique are not the same. Starting with the general QR, all of the available variables were used for the assessment. However, the number of variables used for the subsequent methods, i.e. BQR and the LQR, is fewer. This is to reduce the amount of subjectivity which is present within most of the categorical variables. Additionally, the variables used for the last two methods include the variables that are, according to the Author's knowledge, the most suited for the estimation of the dependent variable in question.

Chapter 4

Coating Defect Size Prediction with Quantile Regression

4.0 Introduction

The motivation for this Chapter is due to address the issues highlighted in Chapter 2 and the needs of the industry to find a meaningful solution in predicting the severity of coating defects to justify subsequent excavation for direct examination. With the availability of real life field data in the form of indirect and direct assessments which was provided by TWI, this has become possible by constructing a mathematical model in the pursuit of understanding the inner workings of the system and hence predicting the size of coating defects. The analyses conducted in this Chapter only

consider the results found in the DCVG method and the subsequent direct assessment done at excavation sites. This Chapter is also intended to identify and quantify the correlation between the various factors which affect DCVG indications. Research conducted here has shown some useful insights which might be of use for further research.

Previous research by Anes-Arteche et.al [63] showed some interesting results through linear QR. However, it lacks the ability to cope with the non-linear relationship which exists between variables. The approach taken in this Chapter is similar, but the relationship is assumed to be non-linear, based on the initial observation of the collected data. The number of variables considered in this Chapter is also greater which gives better depth of resolution in terms of the estimation of the dependent variable. Finally, the effects of interaction between certain variables are taken into account when modelling the regression model.

4.1 Middle Eastern Oil Company Pipelines Data

TWI Ltd. was appointed as contractor by the Middle Eastern Oil Company (MEOC) to conduct integrity assessment on nine of its pipelines. The assessment work covers “External Corrosion Direct Assessment” (ECDA) based on the ANSI/NACE SP0502-2010: Standard Practice Pipeline External Corrosion Direct Assessment Methodology. The data used for analysis in this Chapter was taken from the MEOC project by analysing the indirect and direct assessment of the ECDA process. The pipelines were

unpiggable therefore the ECDA approach was chosen as the most appropriate method for the assessment. Description of the data is given in Chapter 2.

The relationship of variables from indirect and direct assessment is presented through regression techniques. Variables considered for assessment are listed in Table 4-1 below. The variables TIS and PS are treated as continuous variables due to the interest of the research in trying to predict the effect of age and size on the size of coating defects. If these were treated in a categorical manner, prediction of defect size based on values within the range of the different age and size would not be possible. Objectives were formulated to better explain the relationship between variables and the prediction of TCDA based on the DCVG indication;

1. Correlation between variables through the Pearson correlation test. This is to see whether there exists linear relationship between variables. Based on this result, the linearity approach towards modelling may be decided.
2. Correlation of SR to corrosion dimensions. The reason for this is to find out whether SR has any role in corrosion (based on the MEOC data set).
3. Special correlation between TCDA and POPD. In theory, large metal exposed area will likely initiate corrosion and increases its activity. Additionally, pitting corrosion is a threat to structural integrity which makes it an important variable to consider.

4. Prediction of the size of TCDA using QR. Through this, better prioritisation of coating defect is achieved and will be useful for the selection of excavation sites.

Symbol	Variables Considered	Type of Variable / Summary Statistics	
		Quantitative	
		Min. Value	0
		1st Quantile	17.87 %
α	IR Drop (%IR)	Median	37.8 %
		Mean	38.48 %
		3rd Quantile	56.7 %
		Max. Value	98.9 %
		Quantitative	
		Min. Value	75.36 Ω-cm
		1st Quantile	560.25 Ω-cm
β	Soil Resistivity (SR)	Median	1282 Ω-cm
		Mean	2722.11 Ω-cm
		3rd Quantile	2508.14 Ω-cm
		Max. Value	43332 Ω-cm
		Quantitative	
		Min. Value	0 %
		1st Quantile	0 %
γ	Percentage of Pit Depth to Wall Thickness (POPD)	Median	2.537 %
		Mean	10.451 %
		3rd Quantile	17.471 %
		Max. Value	100 %
		Quantitative	
		Min. Value	0 %
		1st Quantile	3 %
δ	Deposits under Coatings (DUC)	Median	30 %
		Mean	35.4 %
		3rd Quantile	60 %
		Max. Value	100 %
		Quantitative	
		Min. Value	0 cm
		1st Quantile	100 cm
ε	Depth of Cover (DOC)	Median	110 cm
		Mean	109.5 cm
		3rd Quantile	130 cm
		Max. Value	210 cm

ζ	Time in Service (TIS)	Quantitative	
		Min. Value	19 years
		1st Quantile	20 years
		Median	36 years
		Mean	32.5 years
		3rd Quantile	39 years
		Max. Value	39 years
η	Pipe Size (PS)	Quantitative	
		Min. Value	26 inches
		1st Quantile	36 inches
		Median	36 inches
		Mean	35.3 inches
		3rd Quantile	36 inches
		Max. Value	42 inches
$TCDA$	Total Coating Defect Area	Quantitative	
		Min. Value	0 cm^2
		1st Quantile	1200 cm^2
		Median	9985 cm^2
		Mean	44893 cm^2
		3rd Quantile	77865 cm^2
		Max. Value	269894 cm^2
Backfill Type			
θ	Rock	Qualitative	
κ	Sand + Clay	Qualitative	
λ	Stones + Clay	Qualitative	
Coating Type			
μ	Coal Tar	Qualitative	
ξ	Polyethylene	Qualitative	
CW	Cold Wrap	Qualitative	
Backfill Geometry			
ρ	Angular	Qualitative	
σ	Round + Angular	Qualitative	
R	Rounded	Qualitative	
pH Of Water in Soil			
φ	Acidic	Qualitative	

χ	Alkaline	Qualitative
ψ	Neutral	Qualitative
pH Of Water Underneath Coating		
ω	Acidic	Qualitative
ü	Alkaline	Qualitative
ï	Neutral	Qualitative

Table 4-1: Types of Variables Considered for Assessment

4.2 Linear Correlation of Variables

It is known that the DCVG technique and %IR is used to locate coating defects and quantify their severity. This is supported by [81] which states that the DCVG technique is used to locate and establish coating defect sizes (TCDA) on buried pipelines. Although this is the case, the %IR values obtained are not affected solely by TCDA but are also attributable to other factors [9]. To investigate this further, analyses of the correlation between the numeric variables listed in Table 4-1 were conducted. An additional correlation analysis was also carried out with the addition of the coefficient of determination using new variables (not listed in Table 4-1). This separate analysis was necessitated by the lack of data for calculating the new variables and hence the impossibility of pairing them up with other variables. The new calculated variables are the Total Corroded Area (TCA) and the Total Corroded Volume (TCV). The independent variables considered for this second analysis were the SR, TCDA, Coating Disbondment Area (CDA), Total Corroded Area (TCA) and the Total Corroded Volume (TCV).

SR was also correlated against the TCA and TCV. According to [82], “the resistivity of the soil is one of many factors that influence the service life of a buried structure”. Therefore, an investigation was conducted into the SR variable’s correlation with corrosion.

4.3 Special Linear Correlation between TCDA and POPD

The correlation between TCDA and POPD was also investigated. As the size of TCDA increases, corrosion activity is expected to increase due to the larger exposed area’s interaction with the environment. If the coating is properly applied (no coating defect present), then corrosion activity on the interface of the pipe is non-existent. Additionally, the level of CP protection is thought to play a key role in preventing the pipeline from corroding. However, as was pointed out by [83], the corrosion process at coating defects is latent even though there are adequate levels of protection coming from the CP current.

After both systems are in place, the extent of external corrosion activity is controlled by the coating and the CP system [83]. This is echoed in [5] which states that one of the most important factors for corrosion prevention in oil and gas pipelines is that the coating needs barrier properties to prevent the ingress of corroding elements and sufficient mechanical strength to resist coating breakdown. This can be summarised as an important factor to

corrosion (POPD) is coating breakdown (TCDA). Hence, an analysis of their correlation is necessary to validate this claim.

4.4 Size Prediction of TCDA

Multivariate QR was chosen to estimate the size of the TCDA. Non-linear terms were also introduced to account for the non-linear relationship between variables observed in the MEOC data set. Some of the variables used in this model are of a qualitative (categorical) nature as compared to the previous correlation method where only numerical variables were chosen for assessment. To obtain a clearer understanding of the phenomenon of coating defects, the variables included in this research are greater than those proposed in [63]. A total of 12 variables were taken from the indirect and direct assessment phase of the ECDA and were identified as the contributors to coating defects. These variables are all the information that was collected during the indirect and the direct assessment phase of the ECDA.

For ease, computation was done using the R software. With the developed model, investigating the effects of each independent variable (as listed in Table 4-1) over the range of TCDA sizes becomes possible. This is important since it enables characterization of coating defects more systematically (as opposed to categorisation) and will assist in prioritizing potential excavation sites for further assessment.

4.4.1 Non Interaction and Interaction model

[84] highlights that an effect of interaction between variables is the effect of combinations of independent variables on the resulting dependent variable. According to [85], it is recommended to investigate the interaction effect of the independent variables within a regression model as opposed to studying its isolated effect which could otherwise lead to inaccurate conclusions.

The relationship between %IR and TCDA is not straightforward [9] and relies on other factors to yield meaningful TCDA results. Thus, two models (Non-Interaction and Interaction) were proposed with the intention of observing the interaction (or the lack of) between variables, particularly the %IR variable. All possible combinations of the variables are considered and later reduced for the final model. Only continuous independent variables and a two way interaction is considered for simplicity.

4.4.2 Akaike Information Criterion (AIC)

The Akaike Information Criterion or AIC is an index used to compare and choose between two competing models according to quality (in terms of model fit). Regression models almost always suffer from information loss in their attempt to represent the true regression. In model selection, it is ideal to choose the model which minimizes this loss as well as improving the fit. AIC provides an index as a relative measure of the loss generated by the model. It is defined as

$$AIC = -2Lm + 2m \quad \textbf{4.1}$$

where:

Lm = Maximized log likelihood (for QR – Asymmetric Laplace Distribution (ALD) is used for the likelihood function)

m = Number of variables in the model

As the number of variables increases for a certain model, its goodness of fit also increases. AIC considers the increase in the goodness of fit of the model by applying a penalty when the number of variable increases. The model that uses the fewest variables but still achieves a decent goodness of fit is preferred with a lower value of the index [86].

In this Chapter, two regularised competing models were generated based on the Interaction and the Non-Interaction models. A simple and concise model is desired for the practical purpose of application during the actual DCVG survey (in the field) where quick and informed decisions are required. The AIC determines the better model and supports the context for the application of each model.

4.5 Data analyses

4.5.1 Correlation of Variables

A total of nine pipelines were considered for assessment. All the pipelines' contents were crude oil, and all were subject to impressed current CP and coated with either polyethylene, cold wrap or coal Tar. The date of construction of these pipes ranges from as early as 1969 to the latest in 1991. Initial correlation assessment was performed for all the numeric variables. The results are presented in Table 4-2 below.

	PS	TIS	%IR	SR	TCDA	DUC	POPD	DOC
PS	1.0000	-0.0406	-0.1873	0.0001	0.0603	-0.0521	-0.1949	0.0872
TIS	-0.0406	1.0000	-0.1406	-0.0327	-0.3587	0.5561	0.3149	0.0032
%IR	-0.1873	-0.1406	1.0000	-0.0782	0.1310	-0.1442	0.0058	0.0617
SR	0.0001	-0.0327	-0.0782	1.0000	-0.0500	-0.0229	-0.0229	0.1306
TCDA	0.0603	-0.3587	0.1310	-0.0500	1.0000	-0.1760	-0.0736	-0.0519
DUC	-0.0521	0.5561	-0.1442	-0.0229	-0.1760	1.0000	0.3261	-0.0130
POPD	-0.1949	0.3149	0.0058	-0.0229	-0.0736	0.3261	1.0000	-0.1841
DOC	0.0872	0.0032	0.0617	0.1306	-0.0519	-0.0130	-0.1841	1.0000

Table 4-2: Pearson's Correlation of Numeric Variables

Table 4-2 shows the results of the Pearson's correlation based on the numeric variables obtained from the MEOC data. Generally, the analysis highlights a weak linear relationship between the variables which suggests a non-linear approach may be better suited for modelling. The highest

correlation can be seen between the variables DUC and TIS with a value of 0.56 with a positive relationship while the lowest is the correlation between PS and SR also with a positive relationship. In the negative direction, the TCDA and TIS variables showed some degree of correlation with a value of -0.36. The pairing of TCDA and POPD illustrates a negative linear relationship with a value of -0.0736. This is somewhat surprising since it was previously thought this relationship is quite strong. The size of coating defect (TCDA) weakly corresponds to the %IR variable with a correlation value of only 0.1310. This further proves that the idea of relying solely on readings from a DCVG assessment can be misleading. Another important finding is the correlation between DUC and POPD which suggests some degree of relationship with a value of 0.3261. The correlation between POPD and TIS has similar magnitude with a value of 0.3149.

	%IR	SR	TCDA	TCA	TCV
%IR	1.0000	-0.2546	0.0631	0.2070	0.1846
SR	-0.2546	1.0000	-0.0100	-0.0651	0.1418
TCDA	0.0631	-0.0100	1.0000	0.0270	0.1045
TCA	0.2070	-0.0651	0.0270	1.0000	0.3030
TCV	0.1846	0.1418	0.1045	0.3030	1.0000

Table 4-3: Pearson's Correlation of Newly Calculated Variables

Table 4-3 highlights the values of the Pearson's correlation coefficient of the newly calculated variables. The calculated variables are derived by analysing the POPD and relating it to the MEOC data. The data points

available for the assessment are very few compared to the MEOC data since corrosion is not present at every excavation site. Overall, the correlation values between the variables are not high. This is another example of the inadequacy of adopting a linear correlation approach for this kind of dataset. The highest correlation occurs between TCA and TCV with a value of 0.3030. This is expected since the two calculations are related. An interesting relationship appears between %IR and TCA and also TCV with correlation values of 0.2070 and 0.1846 respectively. This is rather unexpected as the DCVG technique is not known for its corrosion detection ability. However, the interpretation of its capability based on correlation results alone must be treated with caution since the data gathered here are quite limited. Hence, the results should be recognised as relative rather than absolute. The variables %IR and TCDA yield a lower relationship value compared to the previous assessment. This is expected since the data has been “trimmed down” to facilitate the limited number of corrosion points.

A further analysis was done for the new variables to derive the coefficient of determination, R^2. The coefficient of determination quantifies the percentage of plots that are described by the regression line versus the plots that are not. Dependent variables identified for analysis are %IR, TCA and TCV. The independent variables identified are CDA, TCDA, TCA and TCV. TCA and TCV are both in the dependent and independent variable category because they are viewed both as factors of other variables and also as responses by themselves.

No	Pipe Designation	SR vs %IR	TCDA vs %IR	CDA vs %IR	TCA vs %IR	TCV vs %IR	SR vs TCA	SR vs TCV
1	PR1	0.0005	0.4433	0.0148	0.1863	0.0870	0.0040	0.0121
2	PR2	0.0007	0.0622	0.0622	0.3426	0.3338	0.0785	0.0878
3	PR8-1	0.0395	0.0724	0.0122	0.0122	0.0001	0.1867	0.0926
4	PR9-1	0.0454	0.0171	0.0171	0.1246	0.2374	0.0119	0.0074
5	PR9-2	0.2301	0.0116	0.0116	0.0700	0.0002	0.0126	0.0779
6	PR10	0.0328	0.0897	0.0897	0.3167	0.2458	0.0749	0.1234
7	PR11	0.1992	0.0028	0.0028	0.0004	0.0711	0.0552	0.5697
8	PR12-1	0.0803	0.0455	0.0455	0.2208	0.3836	0.0055	0.0522
9	PR12-2	0.0424	0.0025	0.0025	0.2796	0.2796	0.8683	0.9999

Table 4-4: R^2 Values Corresponding to the Nine Assessed Pipelines

Table 4-4 shows the coefficient of determination for each pipeline for each of the variables concerned. Generally, the coefficient values do not show strong correlation. This can be seen more clearly in Figure 4-1. For pipeline PR1, the coefficients got as high as 0.4 which means that only 40% of the variations in %IR are explained by the regression line. This is expected since %IR is used as an indication for coating defect sizing. The remaining 60% are due to other environmental factors. However, if one looks at the other

pipelines, the TCDA does not exhibit a good indicator of %IR. As far as SR is concerned, pipelines PR 9-2 and PR 11 showed some degree of correlation when aligned with %IR. All remaining pipelines showed poor correlation. For pipeline PR12-2, a good correlation between SR and corrosion dimensions is observed. It is known that as a structure is immersed in electrolyte which has resistivity of up to 10000 Ω-cm and above, then corrosion is halted [87]. This might be the case for PR12-2 but the low correlation value for other pipelines indicated otherwise. The high correlation is also due to the low number of data plots (which is 3 in this case). Pipeline PR11 also showed some correlation between SR and corrosion. As opposed to pipeline PR12-2, there are eight data plots available here. Due to this uncertainty, pipeline PR12-2 can be considered as a one-off case and cannot be regarded as a benchmark, although it fits well to established theories. From all the results of the nine pipelines, pipeline PR12-2 has the best outcome with respect to correlation, except for the comparison of %IR and the TCDA where the correlation is almost non-existent. This again is a result of minimal data plots which resulted in good correlation. The apparent high correlation is mainly due to the sparsity of data points which does not represent the true underlying relationship.

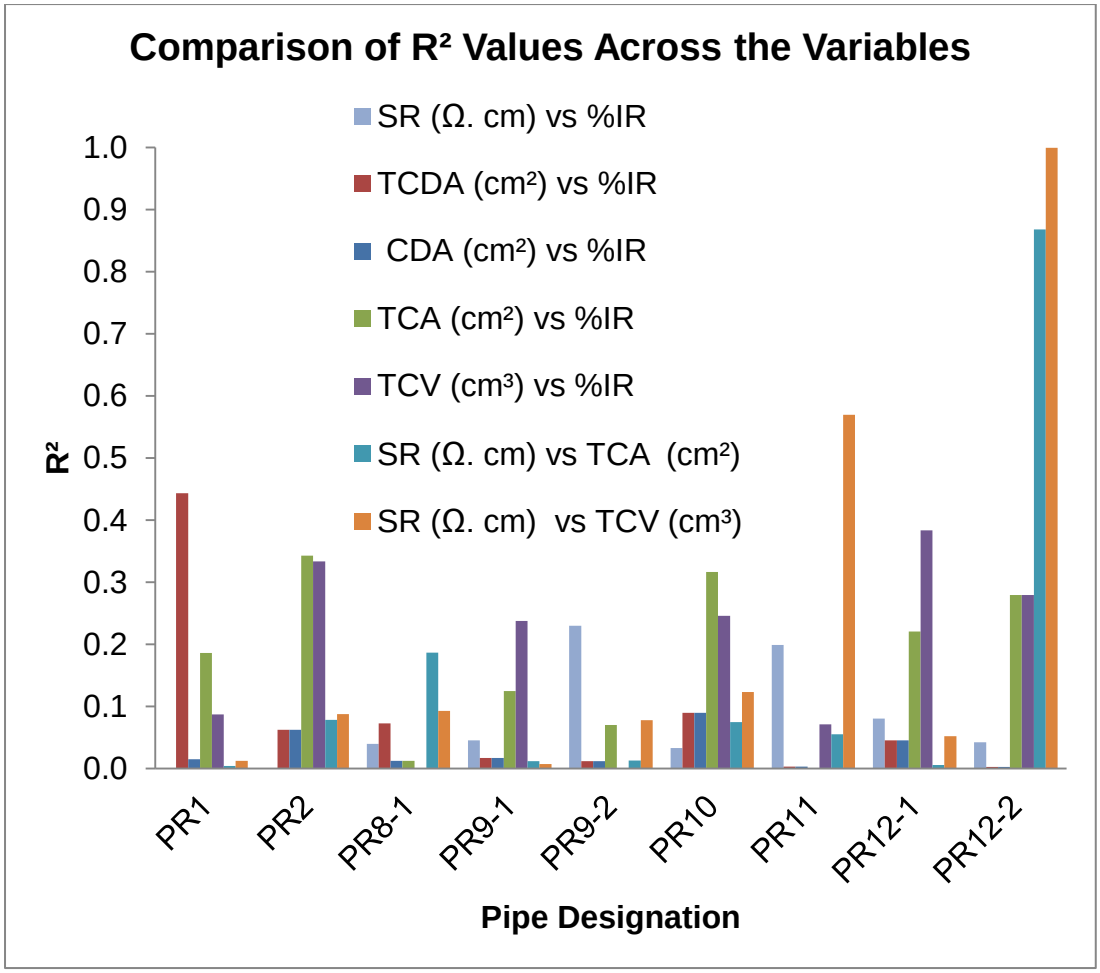


Figure 4-1: Comparison of R² Values across the Variables

There is some evidence for a correlation between %IR and the corrosion dimension where Pipelines PR2, PR9-1, PR10, PR12-1 and PR12-2, all showing some degree of correlation. A DCVG survey is intended to identify any coating defect present on the pipeline and later to categorise its severity based on size. It is unknown for the technique to identify corrosion severity directly. However, the results show some degree of relationship between the two variables. Based on this, it is perhaps possible to use the technique to get an idea of the corrosion condition of the pipeline under investigation.

4.5.2 TCDA vs POPD

The relationship between TCDA and POPD was investigated to establish any correlation between the two variables. POPD is an important variable since it represents a threat to structural integrity. Thus, with the ability to predict its presence with the use of TCDA through DCVG, operators could take proactive steps in managing the risk of corrosion. An initial correlation assessment was done using the Pearson's correlation resulting in a value of -0.0736. Hence a negative trend exists between the two variables, but the magnitude of the coefficient tells us that the linear relationship is weak. The negative trend says that as TCDA increases, the depth of pits decreases. Furthermore, Figure 4-2 shows the linear regression of the variables and indicates poor linear correlation. Additionally, the data seem to be clustered to the left side in the region of 100,000 cm^2 and below. There is also the presence of a punch through at 48,000 cm^2 of TCDA. This suggests that most deeper pits occur at smaller TCDA. At larger TCDA, the pit depth is 40% or less. Since, in theory the severity of TCDA is inspected based on the %IR indication, the chances of smaller TCDA staying unnoticed are high which increases the risk of failure of the pipeline. However, looking at Table 4-4, the relationship between %IR and TCDA is not straightforward, some pipelines indicate better readings than others. A lower reading of %IR represents only a relative indication of the size of TCDA and does not guarantee a smaller coating defect area. In the next section, a QR technique is used to investigate this further by quantifying the size of TCDA based on its contributing factors. Also, non-linear terms are added to the model to

compensate for the poor performance of the linear approach taken previously.

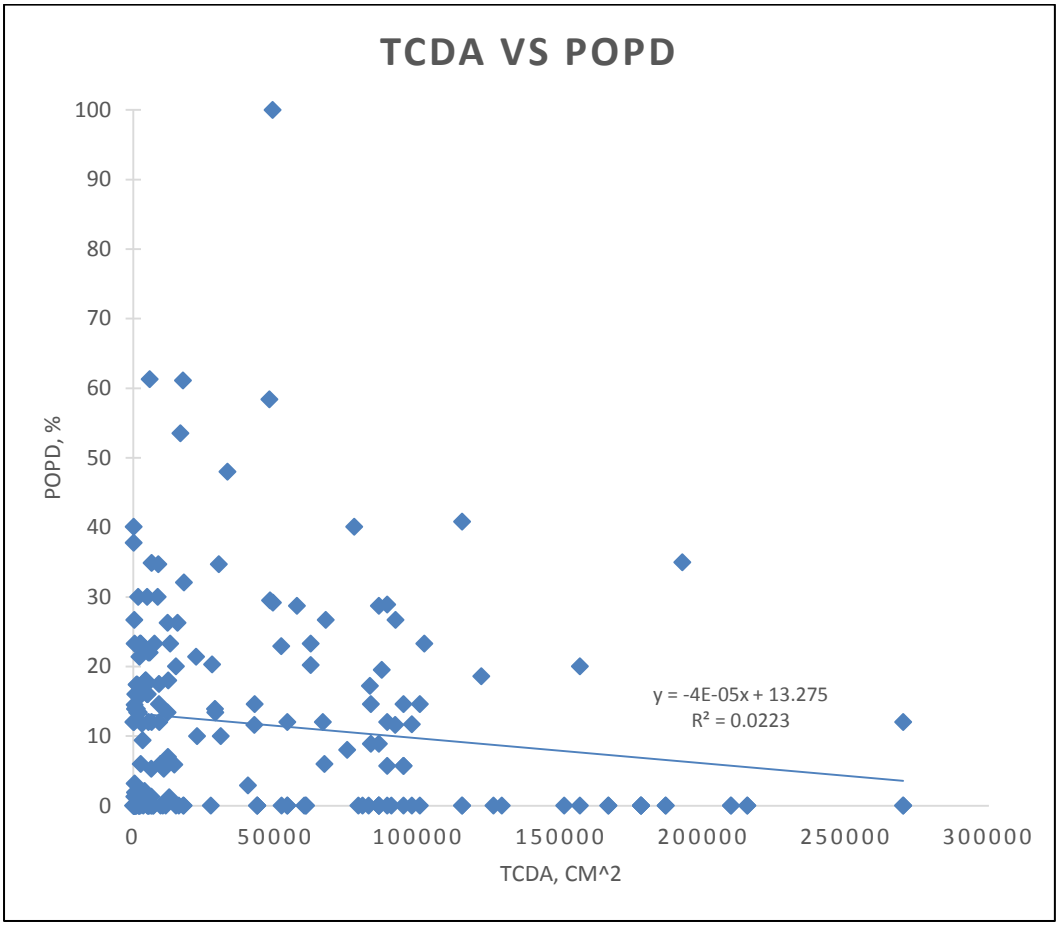


Figure 4-2: Linear Regression of TCDA vs POPD

4.5.3 Size Prediction of TCDA with QR

A more in depth study of the effect of an individual dependent variable can be performed using QR. In this section, the variable TCDA was identified as the dependent variable since this is the variable that needs predicting for prioritization purposes and represents the key stone to which other variables are related to.

The previous analyses showed that correlation between the variables is generally poor. This might be attributable to the linear approach taken in the analyses. Thus, the relationship approach taken here is non-linear. Quantile regression models were constructed based on the non-linear relationship between %IR and TCDA and is represented in Figure 4-3. The models were constructed using the quantreg package of the R software. Please note the functions in Figure 4-3 represent the relationship between %IR and TCDA only. No other variables were considered during the construction of these functions. The polynomial of these functions is at the 5^{th} order. It can be seen this relationship is better represented by non-linear functions. The 0.05 quantile (smaller TCDA) is represented by a “flat” (black) line. This is because most of the measurements for TCDA indicate close to zero readings irrespective of the ascending %IR values. Furthermore, as the quantile increases, the relationship becomes more complex, confirming the indirect correlation between %IR readings and TCDA.

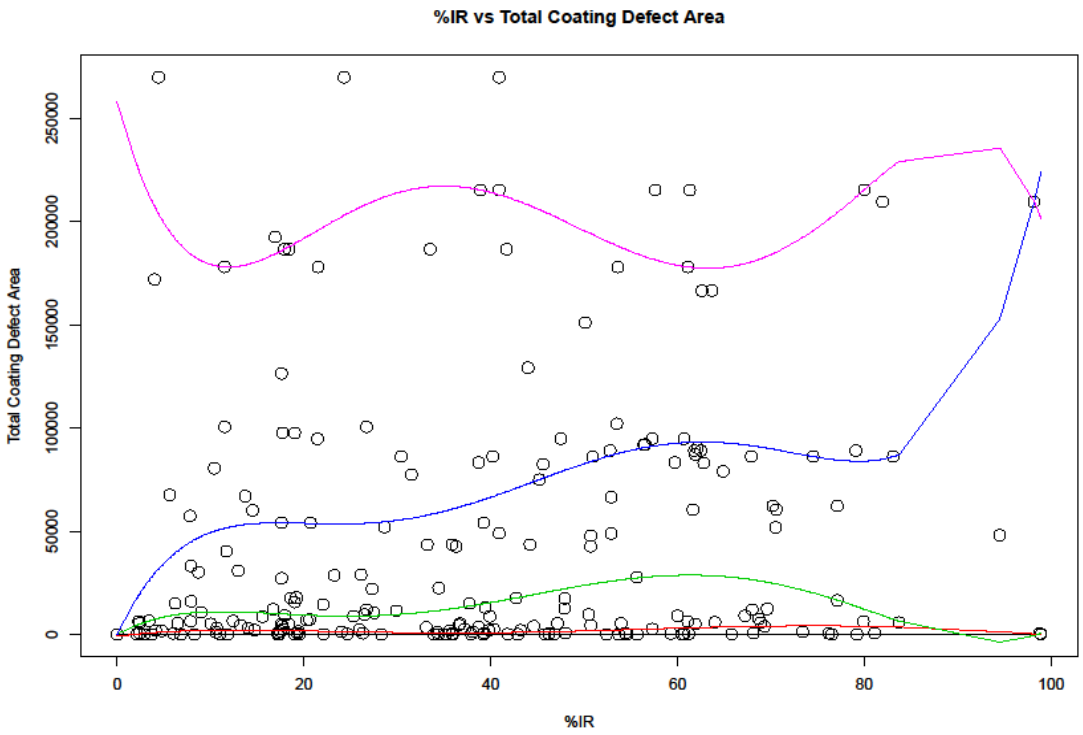


Figure 4-3: Quantile lines showing different quantiles for the TCDA. The black, red, green, blue and magenta each represent 0.05, 0.25, 0.5, 0.75 and 0.95 quantiles respectively

4.5.3.1 Non-Interaction and Interaction model

Two models were constructed using the R software. The quantreg package was utilized for the construction of these models. Both the models assumed an initial 5^{th} order polynomial for curve fitting.

The first is the Non-Interaction model which assumes there is no interaction between the independent variables and the response variable. All the independent variables have no combination effect towards the value of the dependent variable (TCDA).

In the second (Interaction) model, all possible combinations of the independent variables are considered, applied only to those variables which are numeric in nature. The variable %IR, is of particular interest since a DCVG technique relies on this reading for categorising defect size. The interaction of %IR with other variables will give a better understanding of its effect on the value of TCDA. Theoretically, more than two variables can interact but for the purpose of simplicity, only pair-wise interaction of variables were considered.

Both of these models considered the variables listed in Table 4-1.

The intercept in both the Non-Interaction and Interaction models represents a pipe covered with Clay soil, with round geometry and with a Cold Wrap material coating. These variables were taken as reference variables when considering qualitative parameters. The coefficients generated for each model (Non-Interaction and Interaction model) were later selected if their value was beyond the range of -1 and 1. All other values that are within this range were discarded. This process is to simplify the model as otherwise too many variables are considered which may lead to overfitting.

The following section shows the Non-Interaction equation generated based on the different variables which can be used to estimate the TCDA. This is followed by the Interaction model which highlights the QR model with the addition of the interaction variables.

4.5.3.1.1 Non-Interaction model

The Non Interaction fitted model generated from the ECDA report is presented below. Five quantiles were chosen namely the 0.05, 0.25, 0.5, 0.75 and 0.95 quantile to characterise the TCDA distribution. The model highlights the final outcome of the analysis after manual variable selection was done (lies outside the -1 and 1 range).

0.05 quantile

$$TCDA\ (cm^2) = 12703100 + 861.7\boldsymbol{\alpha} + 4.8\boldsymbol{\beta} + 84.3\boldsymbol{\gamma} + 522.4\boldsymbol{\delta} - 1466.2\boldsymbol{\varepsilon} - 2116880\boldsymbol{\zeta} + 9015.3\boldsymbol{\eta} + 12320.5\boldsymbol{\theta} + 1125.2\boldsymbol{\kappa} + 2130.9\boldsymbol{\lambda} - 78\boldsymbol{\mu} - 21795.3\boldsymbol{\xi} + 1834.2\boldsymbol{\rho} - 274.5\boldsymbol{\sigma} + 26292\boldsymbol{\varphi} + 13558.3\boldsymbol{\chi} + 3285.3\boldsymbol{\psi} + 5747.2\boldsymbol{\omega} + 917.9\ddot{\upsilon} + 8665.3\ddot{\iota} - 37\boldsymbol{\alpha}^2 + 7.9\boldsymbol{\gamma}^2 - 15\boldsymbol{\delta}^2 + 16.3\boldsymbol{\varepsilon}^2 + 124401.9\boldsymbol{\zeta}^2 + 123\boldsymbol{\eta}^2 + 2808.2\boldsymbol{\zeta}^3 \quad \textbf{4.2}$$

0.25 quantile

$$TCDA\ (cm^2) = 35572230 + 181.9\boldsymbol{\alpha} + 2.6\boldsymbol{\beta} + 1142.1\boldsymbol{\gamma} + 1095.7\boldsymbol{\delta} \quad \textbf{4.3}$$
$$- 1398.3\boldsymbol{\varepsilon} - 6310216\boldsymbol{\zeta} + 10392.9\boldsymbol{\eta} + 19929.1\boldsymbol{\theta}$$
$$- 3484.9\boldsymbol{\kappa} + 3192.5\boldsymbol{\lambda} + 441.3\boldsymbol{\mu} - 20913.2\boldsymbol{\xi} - 2067.8\boldsymbol{\rho}$$
$$+ 714.5\boldsymbol{\sigma} + 27141.3\boldsymbol{\varphi} + 21982.5\boldsymbol{\chi} + 5552.8\boldsymbol{\psi}$$
$$- 5418.8\boldsymbol{\omega} + 521.3ü + 1746.2ï - 20.6\boldsymbol{\alpha}^2 - 120.2\boldsymbol{\gamma}^2$$
$$- 45.4\boldsymbol{\delta}^2 + 15.7\boldsymbol{\varepsilon}^2 + 395452.7\boldsymbol{\zeta}^2 + 138.5\boldsymbol{\eta}^2 + 3.6\boldsymbol{\gamma}^3$$
$$+ 9493.4\boldsymbol{\zeta}^3$$

0.5 quantile

$$TCDA\ (cm^2) = -329032700 - 278.3\boldsymbol{\alpha} - 8.9\boldsymbol{\beta} + 2404.3\boldsymbol{\gamma} + 1333.5\boldsymbol{\delta} \quad \textbf{4.4}$$
$$- 734.3\boldsymbol{\varepsilon} + 53838800\boldsymbol{\zeta} + 20891.2\boldsymbol{\eta} + 6024.5\boldsymbol{\theta}$$
$$- 19521\boldsymbol{\kappa} - 5665.7\boldsymbol{\lambda} + 2365.7\boldsymbol{\mu} - 21441.7\boldsymbol{\xi} - 3.4\boldsymbol{\rho}$$
$$+ 1523.5\boldsymbol{\sigma} + 5497\boldsymbol{\varphi} + 20853.9\boldsymbol{\chi} + 6827\boldsymbol{\psi} - 62674.6\boldsymbol{\omega}$$
$$- 10720.7ü - 16424ï + 7.9\boldsymbol{\alpha}^2 - 231.8\boldsymbol{\gamma}^2 - 34.3\boldsymbol{\delta}^2$$
$$+ 9.2\boldsymbol{\varepsilon}^2 - 3186093\boldsymbol{\zeta}^2 + 289.4\boldsymbol{\eta}^2 + 6\boldsymbol{\gamma}^3 + 73576.3\boldsymbol{\zeta}^3$$

0.75 quantile

$$TCDA\ (cm^2) = -497471900 + 1580.2\boldsymbol{\alpha} + 2704\boldsymbol{\gamma} - 3282.9\boldsymbol{\delta} - 915.3\boldsymbol{\varepsilon} \quad \textbf{4.5}$$
$$+ 81806480\boldsymbol{\zeta} + 49842.8\boldsymbol{\eta} - 18513.7\boldsymbol{\theta} - 2895.8\boldsymbol{\kappa}$$
$$+ 4099.9\boldsymbol{\lambda} - 6416.5\boldsymbol{\mu} - 46743.9\boldsymbol{\xi} - 6124.1\boldsymbol{\rho} - 257.2\boldsymbol{\sigma}$$
$$- 21310.5\boldsymbol{\varphi} + 27315.5\boldsymbol{\chi} + 12453\boldsymbol{\psi} - 6944.7\boldsymbol{\omega}$$
$$- 703.6ü + 1037.5ï - 92.2\boldsymbol{\alpha}^2 - 291.8\boldsymbol{\gamma}^2 + 156.1\boldsymbol{\delta}^2$$
$$+ 20.4\boldsymbol{\varepsilon}^2 - 4867627\boldsymbol{\zeta}^2 + 715.3\boldsymbol{\eta}^2 + 1.7\boldsymbol{\alpha}^3 + 8.5\boldsymbol{\gamma}^3$$
$$- 2.3\boldsymbol{\delta}^3 + 112994.9\boldsymbol{\zeta}^3$$

0.95 quantile

$$TCDA\ (cm^2) = -416595500 + 570.5\boldsymbol{\alpha} - 24.4\boldsymbol{\beta} + 7009.8\boldsymbol{\gamma} \quad \textbf{4.6}$$
$$- 11886.1\boldsymbol{\delta} + 8040.9\boldsymbol{\varepsilon} + 68523450\boldsymbol{\zeta} + 72857.2\boldsymbol{\eta}$$
$$- 103939.6\boldsymbol{\theta} - 10942.1\boldsymbol{\kappa} - 9451.7\boldsymbol{\lambda} + 32010.5\boldsymbol{\mu}$$
$$- 63691\boldsymbol{\xi} + 5088.4\boldsymbol{\rho} - 8454.4\boldsymbol{\sigma} - 59023.8\boldsymbol{\varphi}$$
$$+ 20243.6\boldsymbol{\chi} + 25815\boldsymbol{\psi} - 79117.8\boldsymbol{\omega} + 3213.6ü$$
$$- 51594.2ï - 49.6\boldsymbol{\alpha^2} - 878.7\boldsymbol{\gamma^2} + 681.6\boldsymbol{\delta^2} - 180\boldsymbol{\varepsilon^2}$$
$$- 4086668\boldsymbol{\zeta^2} + 1052.8\boldsymbol{\eta^2} + 1.1\boldsymbol{\alpha^3} + 26.7\boldsymbol{\gamma^3} - 11.3\boldsymbol{\delta^3}$$
$$+ 1.2\boldsymbol{\varepsilon^3} + 95158.3\boldsymbol{\zeta^3}$$

The %IR (without considering other %IR nonlinear terms) variable for the Non-Interaction model shows a trend which is inconsistent. Starting from the 0.05 quantile, the coefficient values indicate a decreasing trend up to the median. It then peaks at the 0.75 quantile and drops again at the 0.95 quantile. To elaborate further, %IR has a positive linear effect on all the quantiles of the TCDA except for the median quantile. However, %IR squared has a negative effect on all the quantiles except the median.

Examples of other variations across the quantiles include the variable SR which has no effect on TCDA at the 0.75 quantile, the DUC which has different directional values starting at the 0.75 quantile, and the TIS variable which starts having a positive effect at the median. With normal regression, one would look only at the mean. Based on equations 4.2 - 4.6, they would miss the lower and upper tails of the distribution which clearly demonstrate a different scenario. Selected variables such as POPD, PS and TIS are later derived to estimate the size of TCDA based on different quantiles.

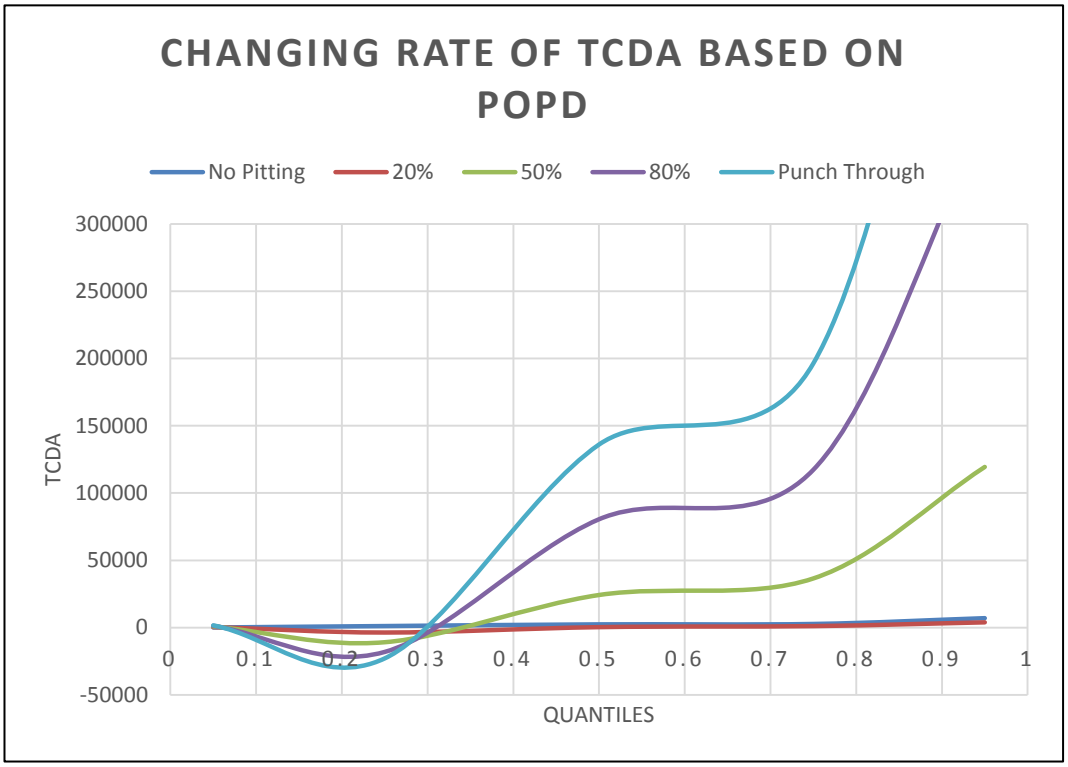


Figure 4-4: Changing Rate of TCDA, Computed as The Derivative of The Quantiles With Respect To POPD for the Non Interaction Model. The colored lines correspond to the different pit depths

Figure 4-4 shows the partially derived quantile plots with respect to the variable POPD from equation 4.2 to 4.6. Several limits of corrosion (pit) depths were used from no corrosion to pipeline failure (punch through) to identify effects of varying levels of corrosion on TCDA. The figure points to an increasing trend signalling an increase in coating defect correlated to an increasing pit depth. Lower quantiles (below 0.2) highlight a decreasing trend where deeper pits are paired with smaller TCDA. The results shown here seem to echo the trend n in Figure 4-2 where deeper pits are found in smaller areas of TCDA. This effect is more pronounced in pits that have made a hole in the pipeline wall.

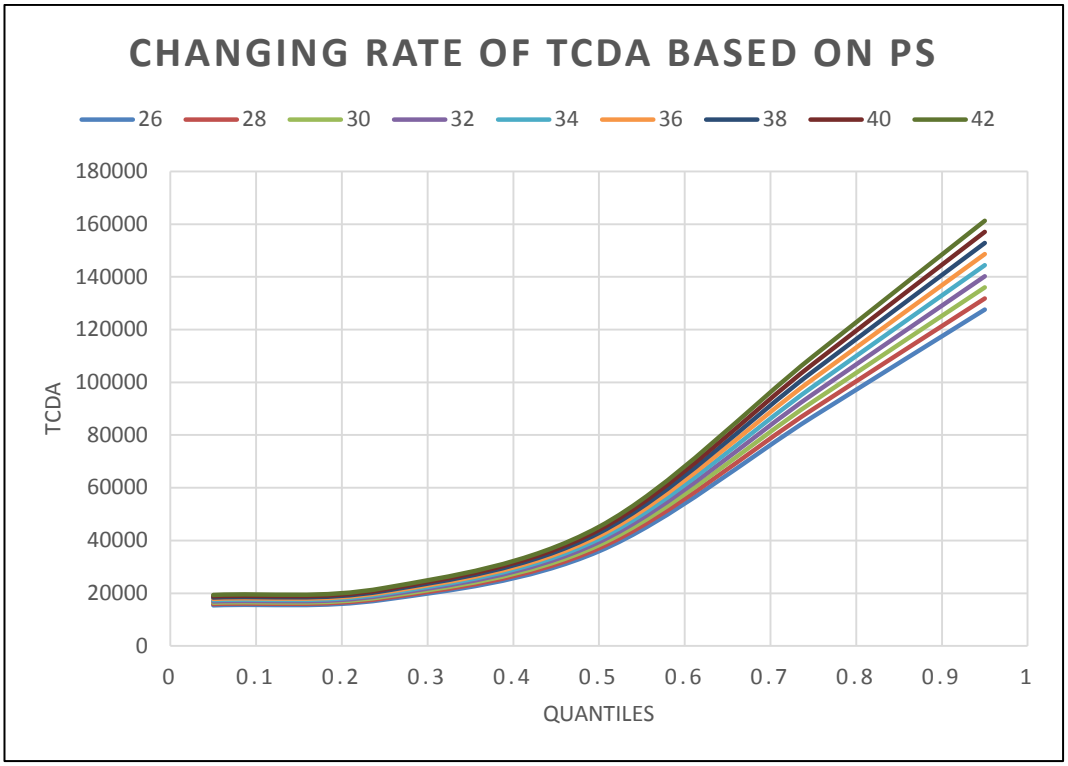


Figure 4-5: Changing Rate of TCDA, Computed as The Derivative of The Quantiles With Respect To PS for the Non Interaction Model. The colored lines correspond to pipeline sizes (in Inches)

Figure 4-5 illustrates the changing rate of TCDA based on the dimensions of the pipeline. Pipeline sizes ranging from 26 inches to 42 inches (based on the MEOC data) were treated as inputs to the partially derived models (with respect to PS) on the different quantiles (equation 4.2 to 4.6). The upward trend is seen across the range which translates to larger pipes having an increasing effect to larger TCDA. The 0.05 quantile up to the 0.2 quantile shows a flat trend which signals the non-effect of pipe sizes at smaller TCDA. It can also be said that, for smaller TCDA, an effect will continue to occur on pipelines irrespective of how big the pipeline is. At larger TCDA the apparent effect of pipe size on coating defects is more pronounced. This makes sense since larger pipelines have more surface for the coating to interact with the surrounding environment – and thus are more likely to degrade. Overall it can be said - the larger the pipe size, the more likely we are to find large coating defects.

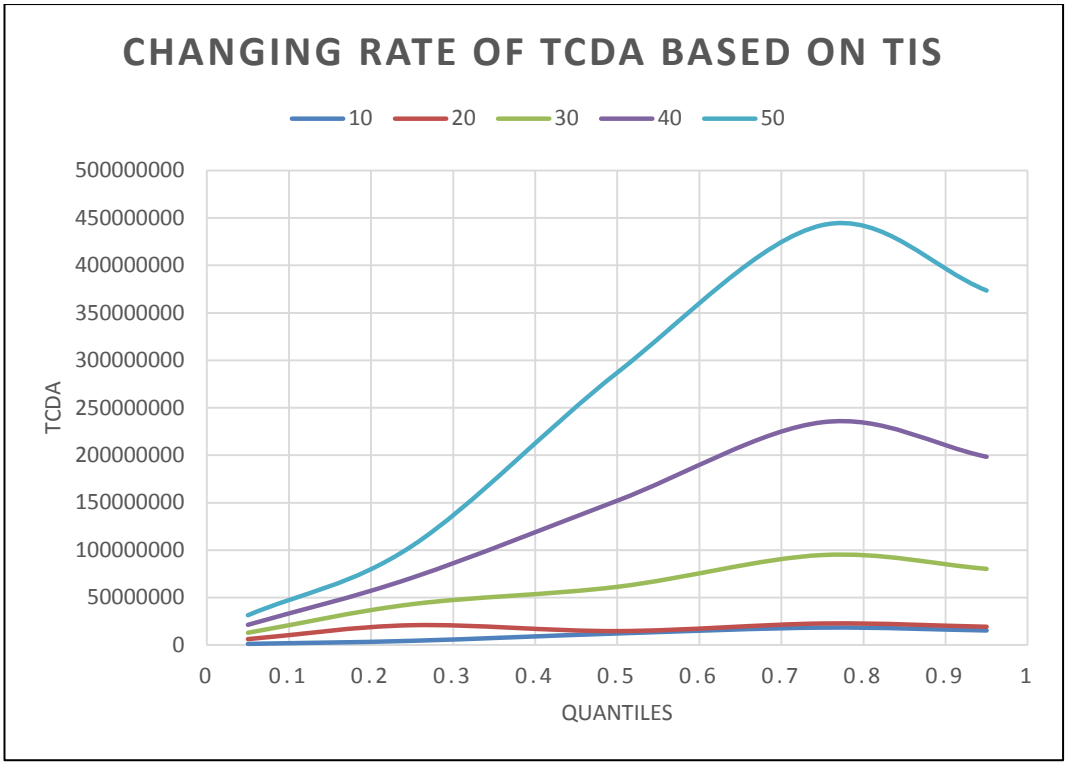


Figure 4-6: Changing Rate of TCDA, Computed as The Derivative of The Quantiles With Respect To TIS for the Non Interaction Model. The colored lines correspond to the age of pipelines (in years)

Figure 4-6 summarises the partially derived quantiles (equation 4.2 to 4.6) with respect to the variable TIS. Age as old as 50 years down to 10 years were treated as inputs to these equations. Generally, it can be seen that as the pipe ages, its effect on the size of TCDA also increases. The maximum effect of the pipe age occurs at TCDA sizes corresponding to the 0.75 quantile. After this, the effect of age on size on larger coating defects seems to taper off. This might be due to the evolution of the coating defects. Age has the biggest influence on defect sizes corresponding to the 0.75 quantile. Larger coating defects than this can only occur due to other dominant factors and not due to age alone (thus the lesser effect). For the oldest pipeline, the effect of age seems to be pronounced over the whole range of TCDA sizes.

The overall trend highlighted in Figure 4-6 seems to fit well with current industry understanding that as the pipeline ages, coatings will deteriorate and that coating defects will increase in size.

4.5.3.1.2 Interaction model

The Interaction model generated from the ECDA report is given below. Interaction terms are added to the already complex non-linear models we see in the Non-Interaction models. Although a preliminary linear correlation assessment was done between the independent variables, interaction terms were added to see if there is any combined effect coming from the variables towards the TCDA and to investigate whether these variables have any significance (in terms of meaning and magnitude). Variables were paired with every possible combination and those having the most significant values (outside the -1 to 1 range) were chosen in the final model.

0.05 quantile

$$TCDA\ (cm^2) = 28324.6 - 10756.3\alpha + 59.5\beta + 280.7\gamma - 2409.9\delta \quad \textbf{4.7}$$
$$- 705.7\varepsilon + 9733.6\zeta - 7828.1\eta + 56389.9\theta - 13772\kappa$$
$$- 4967\lambda + 13415.8\mu - 9051.5\xi - 17481.5\rho + 4041.7\sigma$$
$$+ 45549.1\varphi + 29967.8\chi - 7415.6\psi - 22459.9\omega$$
$$- 2835.1ü - 975.7ï - 41.5\alpha\gamma + 9.8\alpha\delta - 37.5\alpha\varepsilon$$
$$+ 50.8\alpha\zeta + 374.3\alpha\eta - 2.1\beta\eta - 40.2\gamma\delta + 8.8\gamma\varepsilon$$
$$- 106.8\gamma\zeta + 160.6\gamma\eta - 17.6\delta\varepsilon + 62\delta\zeta + 42.4\delta\eta$$
$$+ 52.1\varepsilon\zeta - 1.4\varepsilon\eta + 218.5\zeta\eta + 182.6\alpha^2 - 188.7\gamma^2$$
$$+ 138.7\delta^2 + 15.5\varepsilon^2 + 329.6\zeta^2 + 15.2\eta^2 - 1.2\alpha^3 - 3.4\gamma^3$$
$$- 1.1\delta^3$$

0.25 quantile

$$TCDA\ (cm^2) = -1208842 - 16507.1\alpha + 3.8\beta - 1672.5\gamma - 2019.9\delta \quad \textbf{4.8}$$
$$- 1932.6\varepsilon + 51844.4\zeta + 27540.1\eta + 64969.3\theta$$
$$+ 11738.2\kappa - 23653.5\lambda + 25203.2\mu - 18762.3\xi$$
$$- 17806.3\rho - 4701.5\sigma + 33625.2\varphi + 34433.8\chi$$
$$- 23579.8\psi - 63664.4\omega - 12889.4ü - 15665.3ï$$
$$- 177.6\alpha\gamma + 18.5\alpha\delta - 37.8\alpha\varepsilon + 93.9\alpha\zeta + 552.2\alpha\eta$$
$$- 103.3\gamma\delta + 40.3\gamma\varepsilon - 73.1\gamma\zeta + 315\gamma\eta - 35.3\delta\varepsilon$$
$$+ 86.8\delta\zeta + 46.6\delta\eta + 111.8\varepsilon\zeta - 2.2\varepsilon\eta - 38.4\zeta\eta$$
$$+ 278.9\alpha^2 - 66.7\gamma^2 + 239.7\delta^2 + 37.1\varepsilon^2 + 922.8\zeta^2$$
$$+ 393\eta^2 - 1.6\alpha^3 - 2.0\delta^3$$

0.5 quantile

$$TCDA\ (cm^2) = -3462643 + 7623.3\alpha + 150.7\beta + 9079.6\gamma - 1090.2\delta \quad \textbf{4.9}$$
$$- 6000.2\varepsilon + 104240.2\zeta + 109884.1\eta - 6940.3\theta$$
$$+ 66270.6\kappa - 32648.2\lambda + 33243.8\mu - 153616.4\xi$$
$$- 35568.9\rho - 18159.6\sigma + 13238.7\varphi + 40849.5\chi$$
$$- 12431.5\psi - 98712.7\omega - 9688.1ü - 25110.6ï$$
$$- 240.9\alpha\gamma + 74.1\alpha\delta - 47.8\alpha\varepsilon + 183.6\alpha\zeta - 351.5\alpha\eta$$
$$- 4.4\beta\zeta - 132.8\gamma\delta + 74.1\gamma\varepsilon - 136.6\gamma\zeta + 69.8\gamma\eta$$
$$- 30.1\delta\varepsilon + 203.7\delta\zeta - 135\delta\eta + 197.4\varepsilon\zeta + 15.6\varepsilon\eta$$
$$- 1526.2\zeta\eta + 81.9\alpha^2 - 653.9\gamma^2 + 155.2\delta^2 + 4.6\varepsilon^2$$
$$+ 1090.9\zeta^2 + 668.9\eta^2 - 7.1\gamma^3 - 1.1\delta^3$$

0.75 quantile

$$TCDA\ (cm^2) = -5770395 + 22187.7\alpha + 315.6\beta + 36639.4\gamma \quad \textbf{4.10}$$
$$+ 2012.2\delta - 19.4\varepsilon + 155239.9\zeta + 180311.4\eta$$
$$- 112808\theta + 74003.5\kappa - 33904.9\lambda + 36975.8\mu$$
$$- 200185\xi - 33887.8\rho - 22720.4\sigma - 15990.7\varphi$$
$$+ 43406.8\chi - 13422.8\psi - 128644.4\omega + 5275.4ü$$
$$- 27187.8ï - 208.4\alpha\gamma + 43.5\alpha\delta - 46.2\alpha\varepsilon + 464.1\alpha\zeta$$
$$- 878.4\alpha\eta - 6.6\beta\zeta - 144.4\gamma\delta + 157.2\gamma\varepsilon - 546.7\gamma\zeta$$
$$- 497.1\gamma\eta - 17.9\delta\varepsilon + 49.7\delta\zeta - 56.2\delta\eta + 63.1\varepsilon\zeta$$
$$- 20.3\varepsilon\eta - 2466.2\zeta\eta + 262.2\alpha^2 - 1300.5\gamma^2 + 71\delta^2$$
$$- 32.8\varepsilon^2 + 1416.9\zeta^2 + 1022.8\eta^2 + 1.5\alpha^3 - 12.5\gamma^3$$

0.95 quantile

$$TCDA\ (cm^2) = -5569179 + 47167.4\boldsymbol{\alpha} + 383.5\boldsymbol{\beta} + 39300.6\boldsymbol{\gamma} \quad \textbf{4.11}$$
$$- 509.9\boldsymbol{\delta} - 22718.4\boldsymbol{\varepsilon} + 126013.3\boldsymbol{\zeta} + 210415.6\boldsymbol{\eta}$$
$$- 28072\boldsymbol{\theta} + 84032.7\boldsymbol{\kappa} - 32262.8\boldsymbol{\lambda} + 48593.6\boldsymbol{\mu}$$
$$- 257274.8\boldsymbol{\xi} - 74521.5\boldsymbol{\rho} - 34265.9\boldsymbol{\sigma} - 22767.6\boldsymbol{\varphi}$$
$$+ 8982.2\boldsymbol{\chi} - 19926.9\boldsymbol{\psi} - 124534.5\boldsymbol{\omega} - 1808.8\ddot{\upsilon}$$
$$- 37898.6\ddot{\iota} - 189\boldsymbol{\alpha\gamma} + 21.4\boldsymbol{\alpha\delta} - 102.6\boldsymbol{\alpha\varepsilon} + 353.2\boldsymbol{\alpha\zeta}$$
$$- 1224.2\boldsymbol{\alpha\eta} - 7.1\boldsymbol{\beta\zeta} - 3.3\boldsymbol{\beta\eta} - 160.9\boldsymbol{\gamma\delta} + 308.9\boldsymbol{\gamma\varepsilon}$$
$$- 736.2\boldsymbol{\gamma\zeta} - 908.7\boldsymbol{\gamma\eta} - 60.2\boldsymbol{\delta\varepsilon} + 97.3\boldsymbol{\delta\zeta} + 126.5\boldsymbol{\delta\eta}$$
$$+ 739.9\boldsymbol{\varepsilon\zeta} - 2678.4\boldsymbol{\zeta\eta} + 668.8\boldsymbol{\alpha^2} - 1413.1\boldsymbol{\gamma^2} + 70.8\boldsymbol{\delta^2}$$
$$+ 88.9\boldsymbol{\varepsilon^2} + 1153.2\boldsymbol{\zeta^2} + 1335.3\boldsymbol{\eta^2} + 3.9\boldsymbol{\alpha^3} - 17.1\boldsymbol{\gamma^3}$$

From the Interaction models shown above, the %IR coefficients (at the linear part) indicate an uncertain trend as TCDA increases. The %IR values signify a negative downward trend up to the 0.25 quantile. Above the median, the coefficient trend is positive with an increase at every quantile considered. The interaction terms which includes the %IR is dominant across the quantiles which seems to suggest the interpretation of %IR is not a function of TCDA alone but of other factors as well.

For the variable POPD, the linear part of the median quantile highlights a coefficient value of 9079.6. If we look at the 0.95 quantile, the coefficient value is 39300.6. The increase in value from the median to the 0.95 quantile supports the idea that for larger TCDAs, corrosion is more likely and hence deeper pits (similar to Figure 4-4). This finding does contradict the linear correlation (Figure 4-2) done earlier which only considers the variable POPD

and the TCDA. The combination of non-linear, interaction and other contributing terms has “influenced” the outcome which is different from the linear regression analysis.

To investigate further on this apparent inconsistency, the derivatives with respect to POPD on the interaction models were examined. All the plots in Figure 4-7, Figure 4-8 and Figure 4-9 is a representation of the partially derived quantiles (based on equation 4.7 to 4.11) with respect to POPD, PS and TIS respectively. They use varying limits of corrosion depths, pipe sizes and pipeline age as inputs to investigate on the effect of varying limits on the size of coating defect (TCDA). Additionally, to make the equations more interpretable, the derived equation uses the values of 35.2%, 109.5 cm, 32.5 years, 35.3 inches and 39.2% which correspond to DUC, DOC, TIS, PS and %IR respectively for the calculation of the derived quantiles. These are the mean values of each corresponding variable.

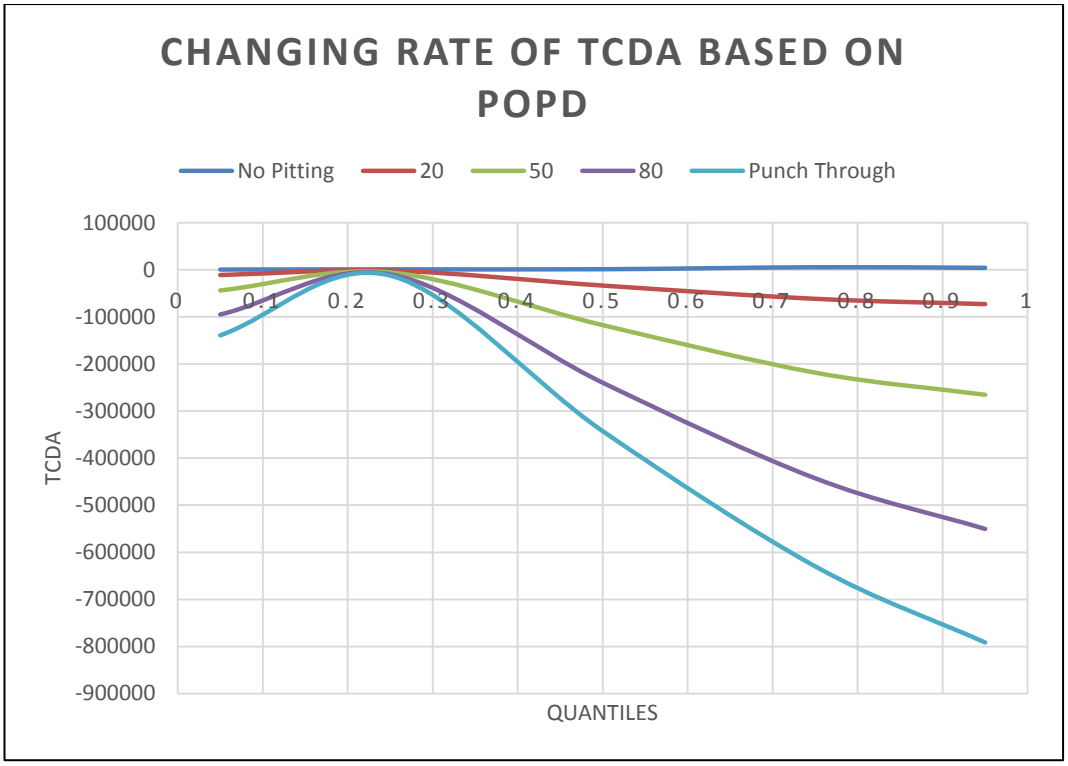


Figure 4-7: Changing Rate of TCDA, Computed as The Derivative of The Quantiles With Respect To POPD for the Interaction Model. The colored lines correspond to the depth of pits (in percentage)

Figure 4-7 highlights the trend of POPD with respect to the differing quantiles of the TCDA. The general trend shows that with increasing POPD, the inverse effect of TCDA increases. This can be seen clearly at larger TCDA where larger POPD has a larger effect compared with other pit depths. At smaller TCDA the effect of POPD is not as strong when compared to larger defects. Nevertheless, the negative values show that the effect is inversely proportional to the size of TCDA. When there is no pitting present on the pipeline, the effect is virtually zero for the different sizes of TCDA. There seems to be another area where the effect of POPD is zero which occurs at the 0.2 quantile region. This finding is interesting since we can assume that at smaller TCDA, the inverse effect of pit depth shows that deeper pits are

more likely and at larger TCDA a more general form of corrosion is highly probable. The transition point of these two opposing trends (with respect to size of TCDA) occurs at the 0.2 region. All of these findings are parallel to the result of the simple linear regression performed in Figure 4-2.

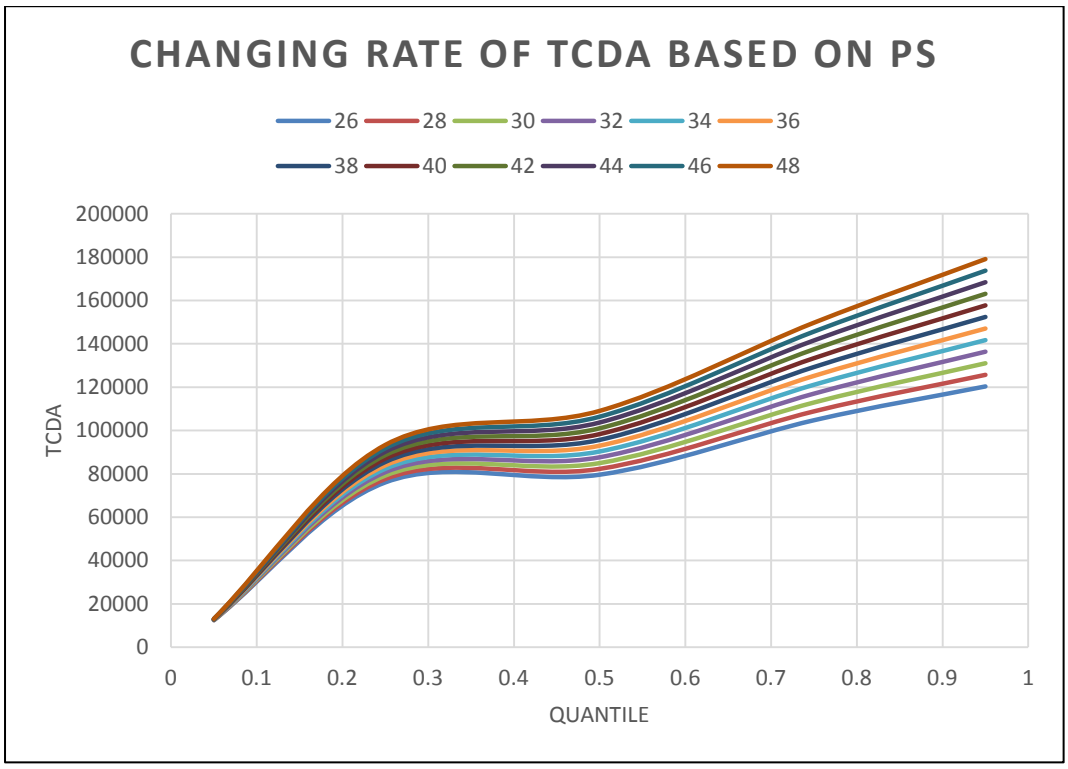


Figure 4-8: Changing Rate of TCDA, Computed as The Derivative of The Quantiles With Respect To PS for the Interaction Model. The colored lines correspond to pipeline sizes (in inches)

Figure 4-8 highlights that PS also plays a role in the determination of TCDA for the interaction model. The derivative of the quantiles with respect to PS shows an increase in effect with increasing coating size. The increase is similar to that of the Non-Interaction model. It can be generalised that the effect applies to all the pipe size considered. At lower quantiles of the coating defect, the estimated coating defect size for different size pipes shows

similar estimates. This trend is similar to the Non-Interaction model (Figure 4-5). Smaller coating defects will occur irrespective of pipe size. However, for larger coating defects, the effect of pipe size is more apparent. This can be seen by the more dispersed estimates of TCDA for the various pipe sizes. Again, this parallels the results obtained in Figure 4-5. With larger pipes, the likelihood of finding larger coating defects is more likely and this is what we have come to observe with the MEOC data. The two models in Figure 4-5 and Figure 4-8 support the validity of the models produced in this section where similar results are obtained by two different means. It also complies with industry experience that larger pipe dimensions will be more susceptible to coating failures (due to larger exposed area).

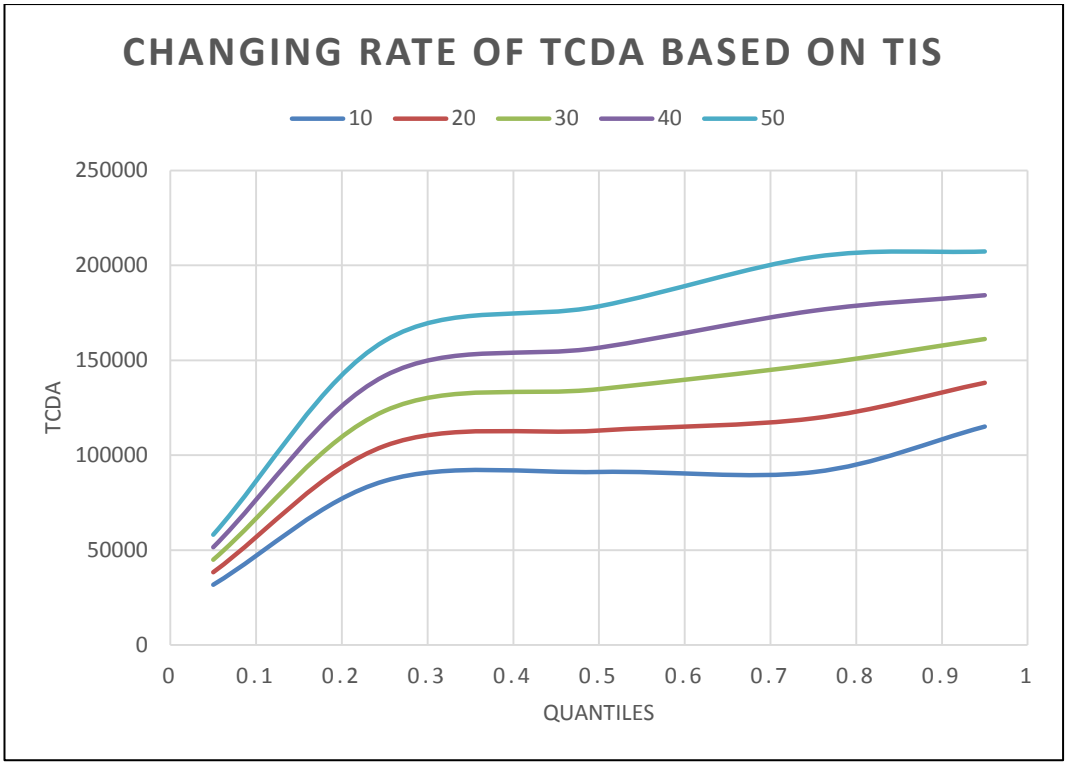


Figure 4-9: Changing Rate of TCDA, Computed as The Derivative of The Quantiles With Respect To TIS for the Interaction Model. The colored lines correspond to the age of pipelines (in years)

Similar to the previous assessments, the models were derived with respect to the age of the pipes (TIS) to investigate on the changing rate of TCDA. It can be seen in Figure 4-9 that the increasing age of the pipelines contribute to an increasing effect towards the determination of TCDA. Below the 0.2 quantile, the slope of the increment of the estimated TCDA is rather steep when compared to the estimated values beyond the 0.2 quantile. This demonstrates that at lower or smaller defect areas the effect of increasing age is quite obvious. Looking back at the Non-Interaction model where a similar assessment was done with respect to the pipeline's age (Figure 4-6), the inflection point where the effect of TIS is greatest occurs at the 0.75 quantile. In Figure 4-9, the greatest effect of increasing age occurs at the

0.95 quantile for the larger coating defects. This can be seen with the largest value coming from the oldest pipelines. The difference in quantile where the effect of age is greatest is probably due to the different factors and terms considered within the interaction. Nevertheless, the two assessments show that age has a positive effect to coating defect growth and the effect is felt most at higher quantiles. The results concur with past and current suspicions of industry, that age is one of the most important factors when considering the severity of coating defects on any given pipeline.

The interaction variable between %IR and the DOC provides an idea of the interaction effects between the two variables. The trend shows increasingly negative values which maximizes at the 0.95 quantile. This outcome is expected since the effect of increasing depth will weaken the %IR signal and thus its contribution to the prediction of TCDA is also weakened. Interaction effects can also be seen between %IR and the TIS of the pipes. The upward trend continues until the 0.75 quantile and drops slightly at the 0.95 quantile. The interaction here is in parallel with the aforementioned assessment where the age of the pipe has a positive effect on the size of TCDA. The slight dip at the 0.95 quantile also indicates the inconsistencies of the DCVG technique where lower %IR values are paired with larger TCDA and *vice versa*.

4.6 Discussion

4.6.1 Linear Correlation of Variables

The results and trends from the correlation section showed low correlation of variables relating to %IR and TCDA. As was stated earlier, PR1 showed some degree of association between %IR and TCDA where the data is 40% explained by the model. There seems to be some degree of association between the %IR and corrosion in terms of area (TCA). However, when considering POPD, which accounts for the depth of the pit, the correlation is not as high. This can also be seen in the results of the correlation between %IR and corrosion in terms of volume (TCV) which also is not significant. These findings seem to suggest that the DCVG technique is able to detect corrosion from an area perspective but not its protruding dimension (depth). This suspicion is further investigated in Chapter 6.

The low linear correlation between the variables can be explained by a few scenarios.

- The relationships between variables are non-linear.
- Not all the variables contribute to the %IR value *per se*. A collection of variables which interacts as one mechanism gives rise to the %IR reading. Taking the %IR as the dependent variable will be discussed in Chapters 5 and 6.
- Pipelines in the area of inspection run in parallel and sometimes cross other pipelines in the vicinity of the complex. Currents coming from other adjacent CP systems which are protecting other pipes will affect

the outcome of the DCVG readings hence unreliable %IR indications arise.

- Inappropriate isolation of the pipelines to be protected from other pipelines. If this should occur, the location where the current leaves the metal will accelerate corrosion.
- Figure 4-10 below, it can be seen that prior to the installation of the ICCP, a sacrificial anode system was installed. The metal wire attached to the pipe was not removed completely and was joined by a tack weld. This is another cause of current leaving the pipe in unintended directions. The currents will cause inaccurate readings to the DCVG.
- The presence of AC high voltage transmission above the pipelines under assessment gives off magnetic fields which interact with the pipeline buried beneath. The interaction induces voltages in the pipe thus affecting the DCVG readings.
- The value of %IR is taken as the DCVG voltage which gets its current from the CP system. DCVG and other ECDA assessment tools rely heavily on the full working capacity of this CP system. Interference in the form of telluric or stray currents can strongly influence the effectiveness of the CP and hence the DCVG measurement.


Figure 4-10: Part of the Metal Rod for the Sacrificial Anodes (no longer present) Attached to the Pipe via a Tack Weld

4.6.2 Summary of Results from the QR Models

4.6.2.1 Non-Interaction model

The uncertain trend of the %IR highlights the model's limitation in representing the size of the defect. A closer look at the relationship between %IR and the TCDA employing the Non-Interaction (for various quantiles) model is demonstrated in Figure 4-11 below.

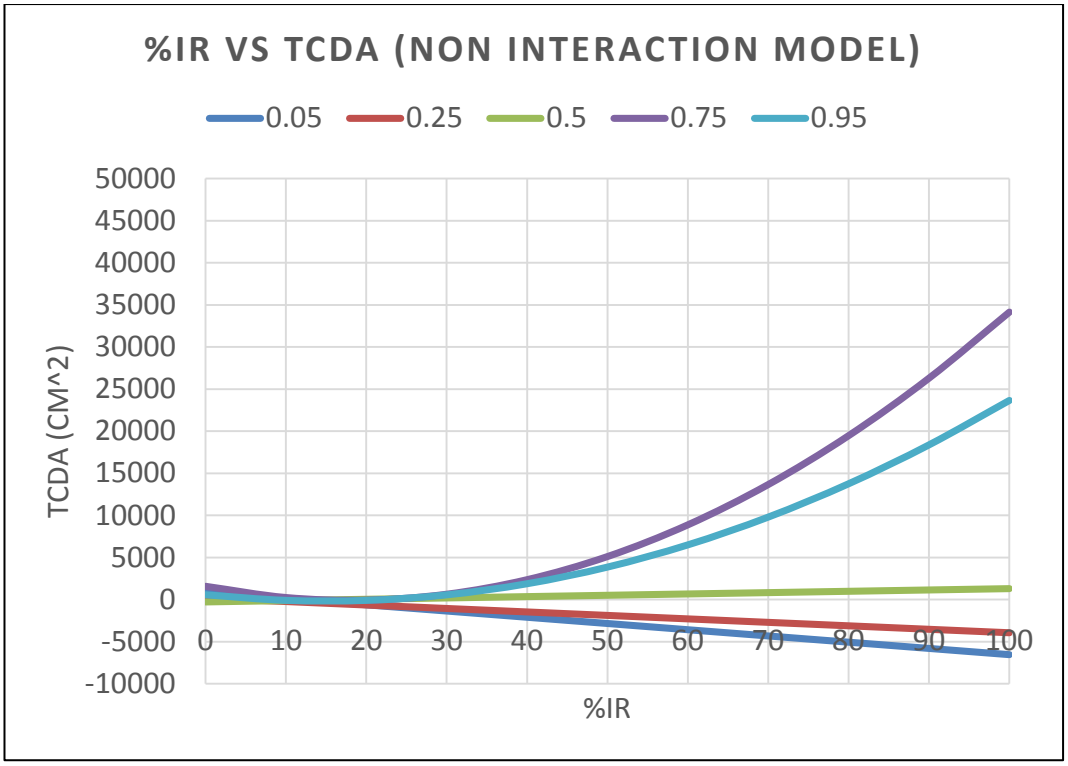


Figure 4-11: %IR vs TCDA for the Non-Interaction Model

Figure 4-11 shows the derivation of the Non-Interaction model with respect to its %IR. It can be observed that the models generated are acceptable but only at quantiles of 0.5 and above. Due to the quadratic nature of the Non-Interaction models, the origin of estimates does not begin at the absolute zero and parts of the estimation transcend the zero line. The apparent violation is possibly attributable to the nature of the mathematical method or the traits of the TCDA data where the DCVG technique is unable to detect (0% %IR reading) coating defects (where in fact there is one).

Negative predictions are seen at lower quantiles with increasing %IR readings translating into negative TCDAs. For lower quantiles (0.05 and 0.25) of the Non-Interaction model, the estimated TCDA is unreliable and is

due to the presence of many small coating defects relating to higher %IR. For the upper quantiles, the estimation seems satisfactory. However, when one considers the MEOC data, the maximum range of the TCDA is 270,000 cm^2. The model's estimation shown here indicates a much lower reading with a maximum predicted at approximately 35,000 cm^2 which occurs at the 0.75 quantile. The apparent underestimation is possibly due to the model's simplicity which does not consider the interaction of variables.

The variables, POPD, PS and TIS all showed positive effect with increasing TCDA. For the case of PS and TIS, this trend makes sense since the larger the diameter of a pipe or the longer the pipe is in operation the more likely it is that larger forms of coating defect will occur. As in the case of POPD, the general increase in trend illustrates the general understanding that more metal exposed to the environment will likely produce or enhance corrosion which could mean deeper pits. This goes against the results found earlier in Figure 4-2. However, the Interaction model highlights an opposite scenario (after further analysis) which seems to suggest that interaction terms are important for different variables.

4.6.2.2 Interaction model

The trend of %IR with respect to TCDA shows an increase with a maximum at the 0.95 quantile. This supports the underlying theory that larger %IR values correspond to larger TCDA. The estimated size of the TCDA is given

as the derivative of the TCDA with respect to %IR. The resulting estimates are given in Figure 4-12 below.

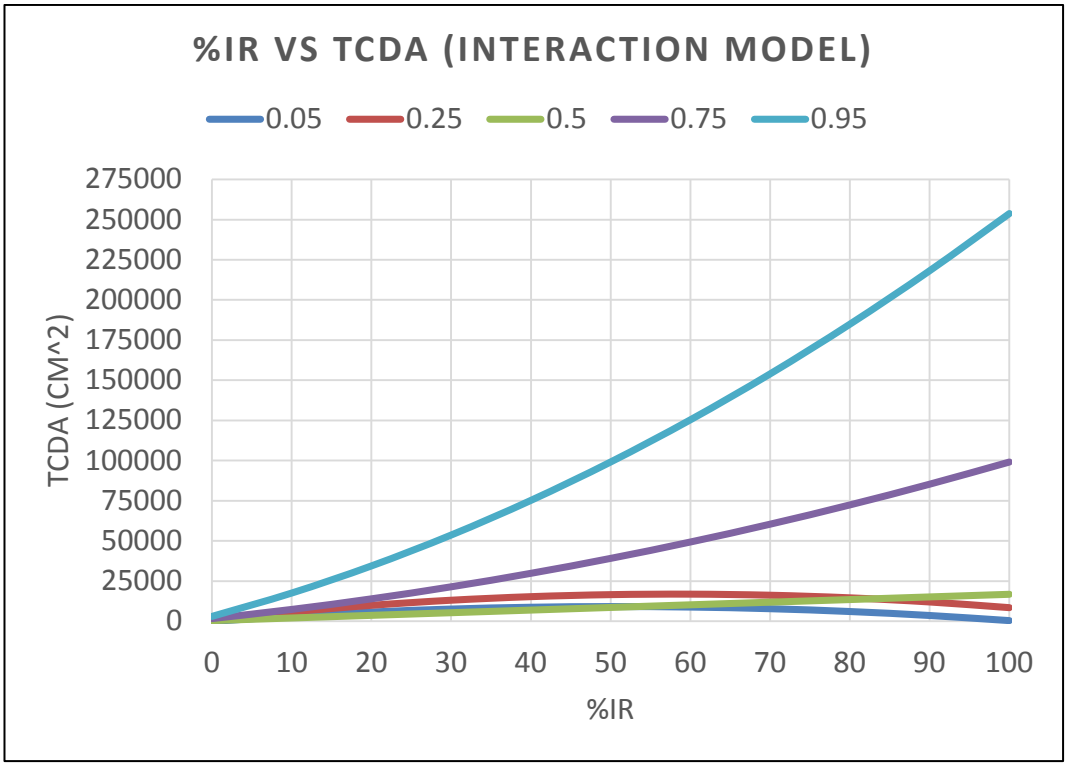


Figure 4-12: %IR vs TCDA for the Interaction Model

As demonstrated in Figure 4-12, there seems to be good predictive indication of the size of TCDA with respect to %IR. Models from the lowest to the highest quantiles cover the range of the TCDA sizes. The lower quantiles give an inverted parabola which illustrates the uncertainty of readings of lower TCDA associated with %IR. For the estimation of larger TCDAs, the models in Figure 4-12 seems to adequately represent the underlying philosophy of the DCVG technique. The inclusion of interaction variables and the non-linear approach (up to 5^{th} order) taken to model TCDA sizes has influenced the outcome of this analysis. For the case of the MEOC data,

these inclusions suggest their importance and not considering them would possibly lead to inconclusive results. As a word of caution, these models are applicable only to the MEOC data since this is where the models were generated from. Since the ECDA is an iterative process, these models could be useful when conducting future assessments on the same pipes. A Bayesian method could be an alternative approach to repetitive assessments where the results in this thesis could act as the prior information for future inspection processes. This approach is further elaborated in Chapter 5.

Results have shown that similar trend exists for the variable POPD when compared to the simple linear regression in Figure 4-2. This means that with smaller defect area, POPD has a much stronger presence. Larger defect areas will tend to have lower POPD values hinting that corrosion is much more generalised. Moreover, the results indicated that the growth rate of TCDA is not in parallel with the growth rate of pits. The growth of pits seems to be faster resulting in deeper pits occurring at smaller TCDA. This idea is further elaborated in Chapters 5 and 6.

As highlighted in the Non-Interaction model, the PS plays a vital role in determining the TCDA. This is also true for the Interaction model. A larger diameter pipe has a lot more surface area and hence is likely to demonstrate a larger TCDA. As with the variable TIS, the older or the longer a pipe is in service, the higher is the probability of developing coating defects.

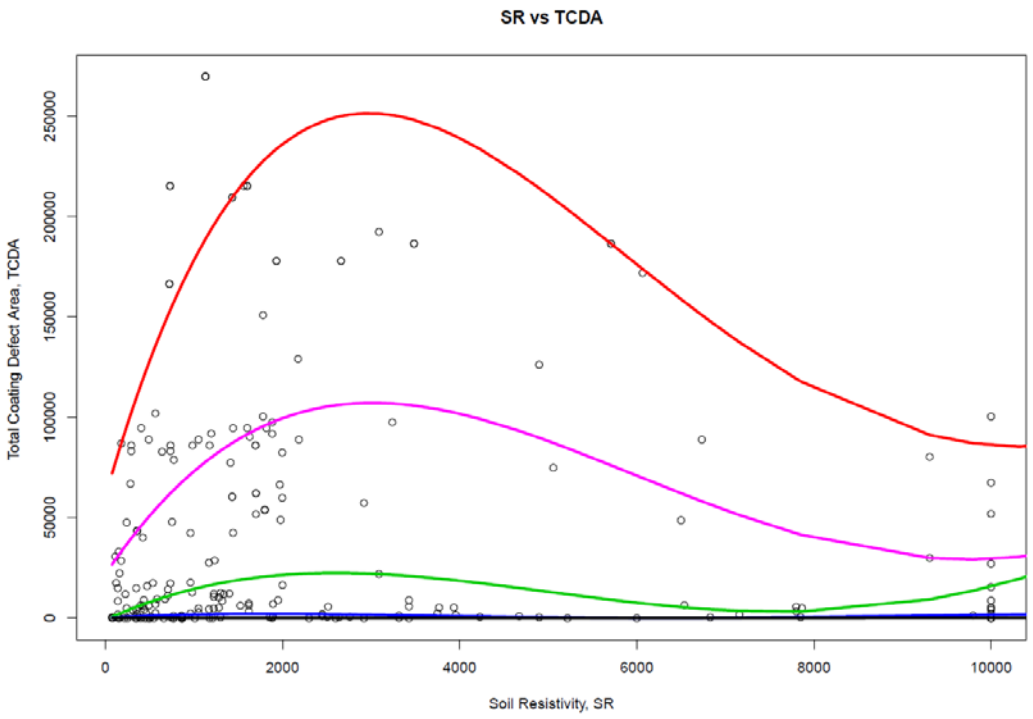
An interesting effect was found at the Interaction variable between the %IR and the DOC. A decreasing trend was found and was most negative at the 0.95 quantile. This can be explained by the effect of the DOC (soil) on top of the pipe on the %IR reading. As an inspector walks the path of the pipeline conducting a DCVG survey, indications from defects will not be detected if out of range. The signal from the defect can only be picked up by the DCVG probes if the ground surface cuts through the equipotential lines of the DCVG voltage gradient spheres. For underground pipes, this signal will not be apparent, especially for small defects or in deeper soil cover. This finding is also confirmed by findings in Chapter 5 and 6 using Bayesian QR and logistic QR.

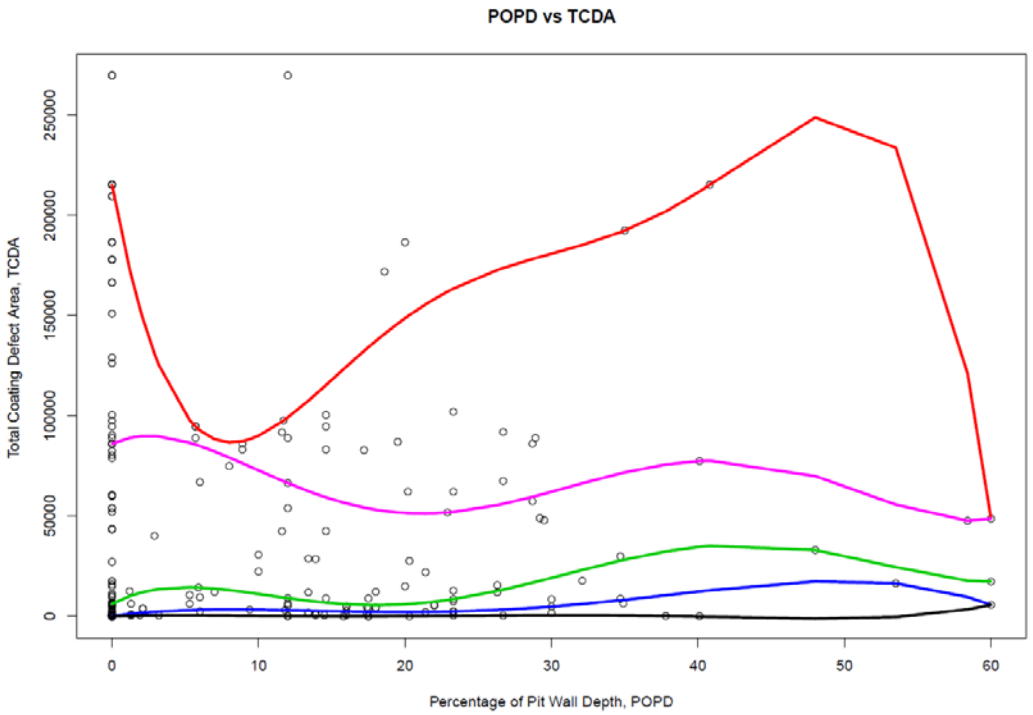
The Interaction effects can also be seen from the %IR interacting with the TIS variable. Unlike the previous example, the trend demonstrated here indicates an upward trend across the quantiles. This perhaps is due to ageing pipelines which contributes to larger TCDA (stated earlier) corresponding to a higher %IR value. The effect of a pipeline's age is greater for larger TCDA.

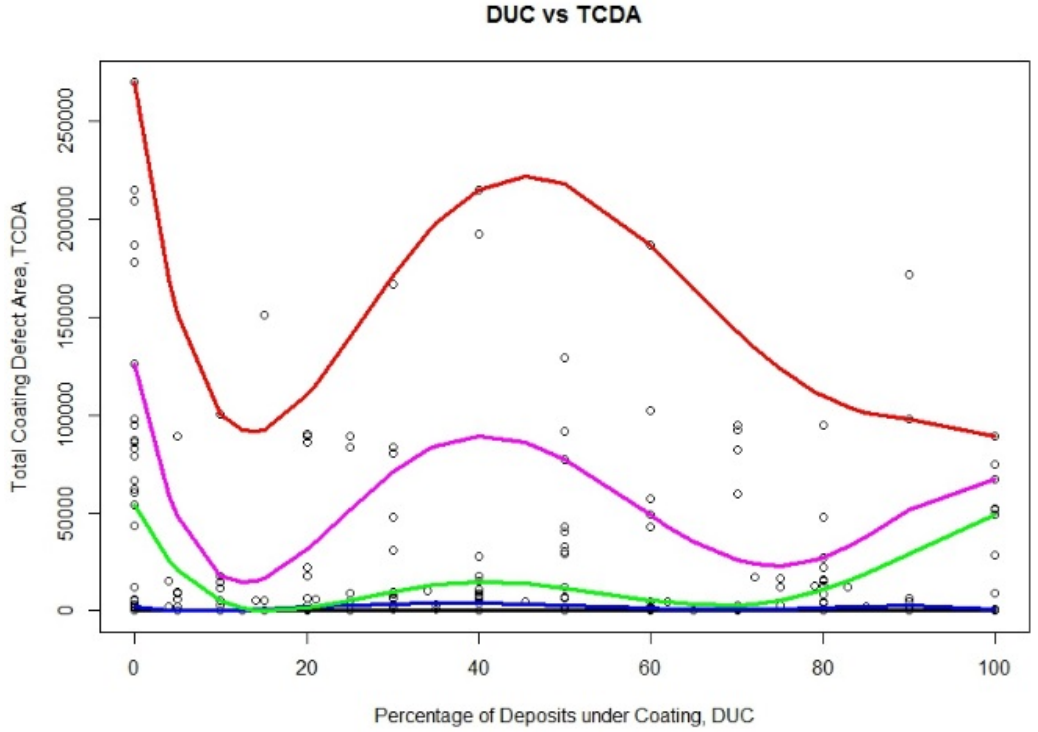
4.7 Least Absolute Shrinkage Selection Operator (LASSO)

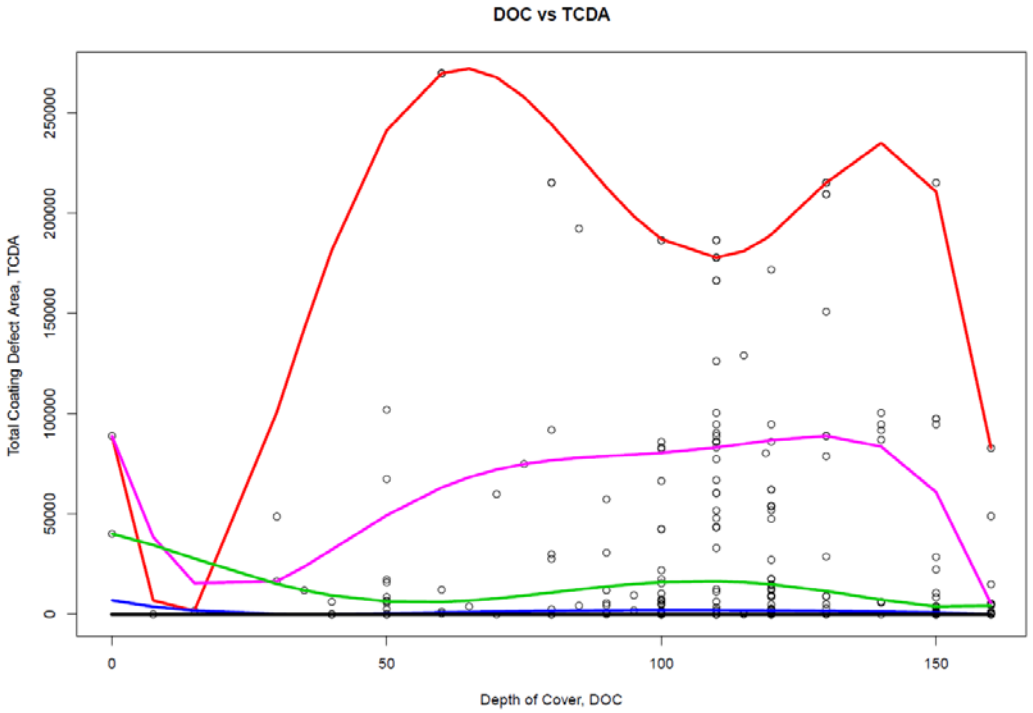
In the models previously constructed, it has been noted that the number of variables in the interaction and the Non-Interaction models is relatively large when compared to the number of data points in the MEOC data. The large number of variables generated was due to the higher order polynomials used

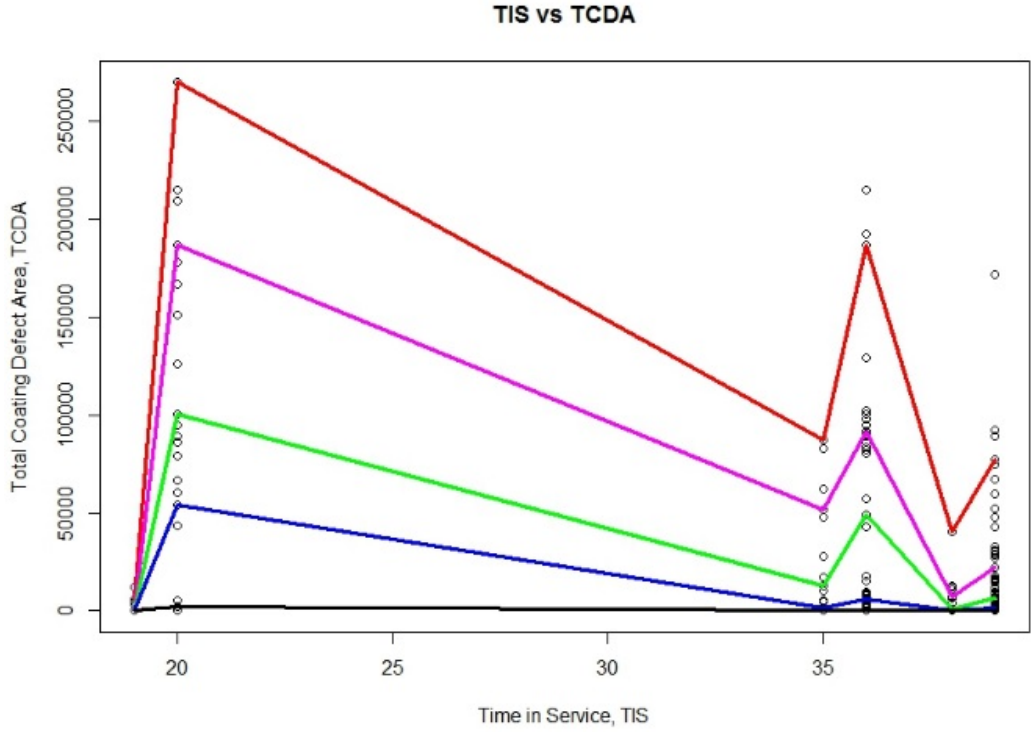
for model construction which may lead to overfitting of the current data set. Therefore, the constructed models might be of little contribution for the assessment of other pipelines beyond the scope of those highlighted in this research. One can argue that the models are suited to pipelines that reside in similar environments with identical operating conditions to those in this thesis. While this is true to a certain extent, the ideal way is to generalise the functions through a process called regularisation. The LASSO technique was used to perform variable selection and regularisation on the Interaction and Non-Interaction models. By doing so, a much more general model is produced and hence a better fit for upcoming datasets. Figure 4-13 illustrates the extent of the relationship between selected independent variables and the fixed dependent variable, TCDA. Each independent variable is modelled using non-linear QR separately to show how each variable relates to TCDA.




POPD vs TCDA
Total Coating Defect Area, TCDA
0
50000
100000
150000
200000
250000
0
10
20
30
40
50
60
Percentage of Pit Wall Depth, POPD


DUC vs TCDA
Total Coating Defect Area, TCDA
0
50000
100000
150000
200000
250000
0
20
40
60
80
100
Percentage of Deposits under Coating, DUC


DOC vs TCDA
Total Coating Defect Area, TCDA
0
50000
100000
150000
200000
250000
0
50
100
150
Depth of Cover, DOC


TIS vs TCDA
Total Coating Defect Area, TCDA
0
50000
100000
150000
200000
250000
20
25
30
35
Time in Service, TIS

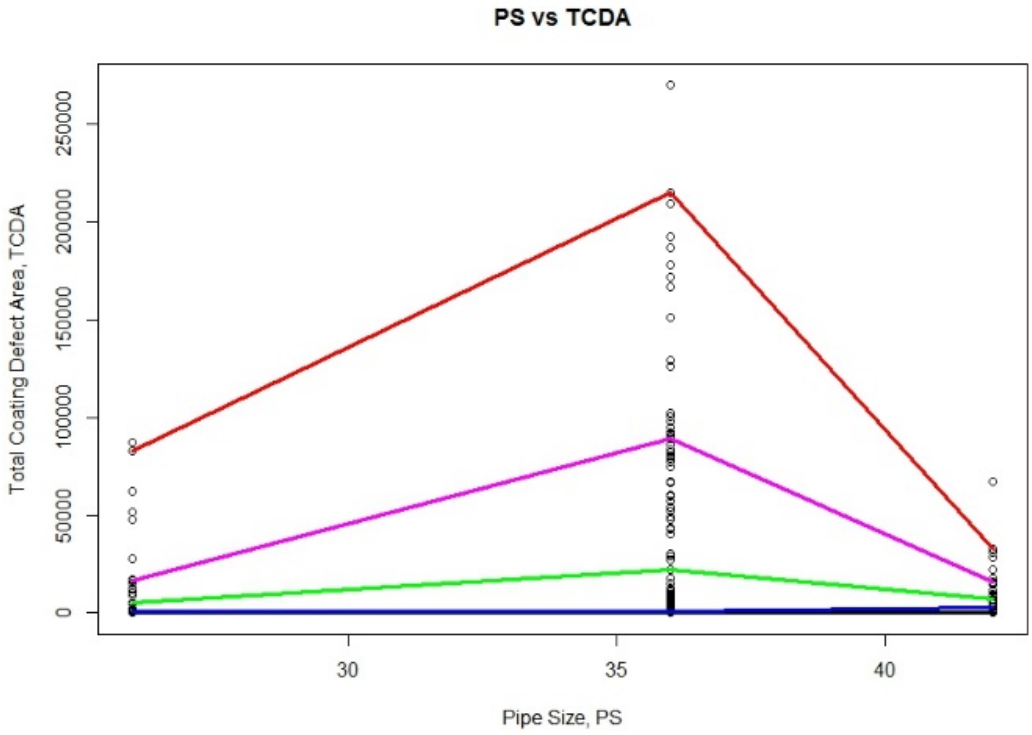


Figure 4-13: The Red, Magenta, Green, Blue and Black functions represent the 0.95, 0.75, 0.5, 0.25 and 0.05 Quantiles Respectively

As mentioned earlier, Figure 4-13 highlights the relationship of various variables with the dependent variable, TCDA. It is evident, from Figure 4-13, that the models are tightly knitted to the data set. The relationship of SR and TCDA further illustrates this with the models strictly obeying the polynomial order with which the models were constructed from.

Polynomials up to 5^{th} order were utilised in modelling the relationship of the dataset. For the interaction models, interaction terms were introduced to the polynomial structure. This further inflated the number of variables relative to the number of data points. The problem of overfitting may arise when there exists a non-unique solution to the relationship. When the number of

variables is quite large, as in the case of the polynomial solution proposed in this research, a number of models that fit the data equally well may exist. This potentially leads the model to overfit since the models have a large number of ways in making the error term zero (or close to zero) and may not represent the correct picture of the true underlying relationship.

For the case of the MEOC data, λ (the regularisation parameter), symbolizes the general intention of the DCVG technique which is to correlate the %IR value to the area of the coating defect. The selection of the regularisation parameter is chosen by trial and error, producing models which are simple and yet illustrate the relationship of %IR vs TCDA. The relationship must reflect the general understanding of the DCVG technique where increasing %IR is a signal of an increase in the area of coating defect. Although it is quite obvious that this relationship is non-linear (based on the MEOC data), a general working guideline (model) is essential for engineers and operators in deciding on where to conduct further direct assessment and on which defect requires immediate attention. This guideline can also be a quick reference tool for engineers at inspection sites (in the field) to interpret the %IR signals obtained during the DCVG inspection. With this intention in mind, the regularisation parameter was tuned to yield linear relationships (models) between %IR and TCDA and also to produce the simplest form possible that the model can take. Additionally, coefficients with values within the -1 and 1 range after they have gone through the LASSO process are discarded for further generalization.

The Non-Interaction and Interaction Models which have gone through the LASSO procedure are summarised below.

4.7.1 Non-Interaction (R) Model

0.05 Quantile (R)

$$TCDA\ (cm^2) = -24751.9 + 1.2\boldsymbol{\beta} + 1.6\boldsymbol{\alpha^2} - 1.8\boldsymbol{\gamma^2} + 2.3\boldsymbol{\delta^2} + 1.0\boldsymbol{\eta^2} + 6.0\boldsymbol{\zeta^3} \quad \textbf{4.12}$$

0.25 Quantile (R)

$$TCDA\ (cm^2) = -1986887.0 - 135.4\boldsymbol{\delta} + 6.5\boldsymbol{\alpha^2} - 35.9\boldsymbol{\gamma^2} + 14.1\boldsymbol{\delta^2} - 3.3\boldsymbol{\varepsilon^2} + 16352.1\boldsymbol{\zeta^2} + 6.6\boldsymbol{\eta^2} + 1.6\boldsymbol{\gamma^3} - 668.3\boldsymbol{\zeta^3} \quad \textbf{4.13}$$

0.5 Quantile (R)

$$TCDA\ (cm^2) = -2891689.0 - 1.2\boldsymbol{\beta} + 15.3\boldsymbol{\gamma} - 2288.6\boldsymbol{\varepsilon} + 2225.0\boldsymbol{\eta} + 11.9\boldsymbol{\alpha^2} - 75.2\boldsymbol{\gamma^2} - 5.1\boldsymbol{\delta^2} + 43.9\boldsymbol{\varepsilon^2} + 21757.6\boldsymbol{\zeta^2} - 17.8\boldsymbol{\eta^2} + 3.2\boldsymbol{\gamma^3} - 827.7\boldsymbol{\zeta^3} \quad \textbf{4.14}$$

0.75 Quantile (R)

$$TCDA\ (cm^2) = -3041816.0 + 9.2\boldsymbol{\beta} + 2045.7\boldsymbol{\gamma} - 852.7\boldsymbol{\delta} - 2139.1\boldsymbol{\varepsilon} + 7380.5\boldsymbol{\eta} + 17.9\boldsymbol{\alpha^2} - 343.7\boldsymbol{\gamma^2} + 25.6\boldsymbol{\delta^2} + 66.1\boldsymbol{\varepsilon^2} + 13896.4\boldsymbol{\zeta^2} - 88.5\boldsymbol{\eta^2} + 11.5\boldsymbol{\gamma^3} - 248.5\boldsymbol{\zeta^3} \quad \textbf{4.15}$$

0.95 Quantile (R)

$$TCDA\ (cm^2) = -1790099.0 - 6.0\boldsymbol{\beta} - 4226.0\boldsymbol{\delta} + 35.5\boldsymbol{\alpha^2} - 360.0\boldsymbol{\gamma^2} + 278.4\boldsymbol{\delta^2} + 56.4\boldsymbol{\varepsilon^2} + 28.0\boldsymbol{\eta^2} + 15.9\boldsymbol{\gamma^3} - 4.4\boldsymbol{\delta^3} + 471.1\boldsymbol{\zeta^3} \quad \textbf{4.16}$$

4.7.2 Interaction (R) Models

0.05 Quantile (R)

$$TCDA\ (cm^2) = 14609.4 + 57.4\boldsymbol{\alpha\gamma} + 7.1\boldsymbol{\alpha\delta} - 7.9\boldsymbol{\alpha\varepsilon} + 71.7\boldsymbol{\alpha\zeta} - 43.4\boldsymbol{\gamma\delta} + 41.0\boldsymbol{\gamma\varepsilon} + 37.2\boldsymbol{\gamma\zeta} - 34.2\boldsymbol{\gamma\eta} - 21.3\boldsymbol{\delta\varepsilon} + 67.6\boldsymbol{\delta\zeta} - 11.2\boldsymbol{\delta\eta} - 10.0\boldsymbol{\varepsilon\zeta} + 1.4\boldsymbol{\varepsilon\eta} + 124.0\boldsymbol{\zeta\eta} + 3.4\boldsymbol{\alpha^2} + 82.6\boldsymbol{\gamma^2} + 37.3\boldsymbol{\delta^2} + 3.8\boldsymbol{\varepsilon^2} - 105.7\boldsymbol{\zeta^2} - 58.2\boldsymbol{\eta^2} \quad \textbf{4.17}$$

0.25 Quantile (R)

$$TCDA\ (cm^2) = 147153.6 - 4.0\boldsymbol{\beta} - 6179.3\boldsymbol{\delta} - 2964.8\boldsymbol{\varepsilon} - 17542.9\boldsymbol{\lambda} + 5773.0\boldsymbol{\mu} + 167.0\boldsymbol{\sigma} - 1120.2\boldsymbol{\psi} - 12017.1\boldsymbol{\omega} - 5336.9ü - 2967.0ï + 77.5\boldsymbol{\alpha\gamma} - 14.5\boldsymbol{\alpha\delta} - 2.6\boldsymbol{\alpha\varepsilon} + 10.5\boldsymbol{\alpha\zeta} + 53.9\boldsymbol{\alpha\eta} + 1.1\boldsymbol{\beta\eta} - 69.8\boldsymbol{\gamma\delta} + 48.2\boldsymbol{\gamma\varepsilon} - 83.0\boldsymbol{\gamma\zeta} + 99.3\boldsymbol{\gamma\eta} - 14.9\boldsymbol{\delta\varepsilon} + 140.2\boldsymbol{\delta\zeta} + 78.7\boldsymbol{\delta\eta} + 87.1\boldsymbol{\varepsilon\zeta} - 5.1\boldsymbol{\varepsilon\eta} + 486.3\boldsymbol{\zeta\eta} + 4.2\boldsymbol{\alpha^2} - 156.2\boldsymbol{\gamma^2} + 199.4\boldsymbol{\delta^2} + 37.6\boldsymbol{\varepsilon^2} - 339.9\boldsymbol{\zeta^2} - 272.5\boldsymbol{\eta^2} + 1.8\boldsymbol{\gamma^3} - 1.4\boldsymbol{\delta^3} \quad \textbf{4.18}$$

0.5 Quantile (R)

$$TCDA\ (cm^2) = -63592.1 + 129.7\boldsymbol{\beta} + 3349.2\boldsymbol{\varepsilon} + 5805.5\boldsymbol{\zeta} + 8095.4\boldsymbol{\eta} \quad \textbf{4.19}$$
$$- 28247.7\boldsymbol{\lambda} - 18716.4\boldsymbol{\rho} - 4403.2\boldsymbol{\sigma} - 3965.9\boldsymbol{\psi}$$
$$- 15360.1ü + 89.9\boldsymbol{\alpha\gamma} + 31.7\boldsymbol{\alpha\delta} - 3.0\boldsymbol{\alpha\varepsilon} - 16.2\boldsymbol{\alpha\zeta}$$
$$+ 73.9\boldsymbol{\alpha\eta} - 5.2\boldsymbol{\beta\zeta} + 2.3\boldsymbol{\beta\eta} - 107.1\boldsymbol{\gamma\delta} + 73.9\boldsymbol{\gamma\varepsilon}$$
$$+ 42.1\boldsymbol{\gamma\zeta} + 8.7\boldsymbol{\gamma\eta} - 34.9\boldsymbol{\delta\varepsilon} + 81.4\boldsymbol{\delta\zeta} + 9.7\boldsymbol{\delta\eta} - 76.9\boldsymbol{\varepsilon\zeta}$$
$$+ 16.9\boldsymbol{\varepsilon\eta} + 886.8\boldsymbol{\zeta\eta} + 15.3\boldsymbol{\alpha^2} - 343.1\boldsymbol{\gamma^2} + 76.7\boldsymbol{\delta^2}$$
$$- 29.7\boldsymbol{\varepsilon^2} - 614.0\boldsymbol{\zeta^2} - 653.7\boldsymbol{\eta^2} + 4.3\boldsymbol{\gamma^3}$$

0.75 Quantile (R)

$$TCDA\ (cm^2) = 278481.8 + 235.4\boldsymbol{\beta} + 14150.3\boldsymbol{\gamma} - 1219.2\boldsymbol{\delta} + 11452.5\boldsymbol{\varepsilon} \quad \textbf{4.20}$$
$$- 30844.7\boldsymbol{\lambda} + 160.9\boldsymbol{\alpha\gamma} + 51.8\boldsymbol{\alpha\delta} - 35.5\boldsymbol{\alpha\varepsilon} + 513.5\boldsymbol{\alpha\zeta}$$
$$- 340.7\boldsymbol{\alpha\eta} + 1.3\boldsymbol{\beta\gamma} + 7.9\boldsymbol{\beta\zeta} + 1.9\boldsymbol{\beta\eta} - 119.9\boldsymbol{\gamma\delta}$$
$$+ 147.2\boldsymbol{\gamma\varepsilon} - 799.4\boldsymbol{\gamma\zeta} + 355.7\boldsymbol{\gamma\eta} - 40.0\delta\boldsymbol{\varepsilon} + 40.3\boldsymbol{\delta\zeta}$$
$$+ 69.6\boldsymbol{\delta\eta} - 284.6\boldsymbol{\varepsilon\zeta} + 43.6\boldsymbol{\varepsilon\eta} + 1031.5\boldsymbol{\zeta\eta} + 45.3\boldsymbol{\alpha^2}$$
$$- 817.2\boldsymbol{\gamma^2} + 233.1\boldsymbol{\delta^2} - 110.2\boldsymbol{\varepsilon^2} - 764.6\boldsymbol{\zeta^2} - 580.6\boldsymbol{\eta^2}$$
$$+ 1.1\gamma\boldsymbol{\zeta^2} + 7.2\boldsymbol{\gamma^3} - 1.8\boldsymbol{\delta^3}$$

0.95 Quantile (R)

$$TCDA\ (cm^2) = 596300.1 + 170.9\boldsymbol{\beta} + 163.7\boldsymbol{\alpha\gamma} + 72.2\boldsymbol{\alpha\delta} - 92.5\boldsymbol{\alpha\varepsilon} + 326.4\boldsymbol{\alpha\zeta} + 42.1\boldsymbol{\alpha\eta} + 1.4\boldsymbol{\beta\gamma} - 8.0\boldsymbol{\beta\zeta} + 2.8\boldsymbol{\beta\eta} - 44.3\boldsymbol{\gamma\delta} + 201.3\boldsymbol{\gamma\varepsilon} - 770.6\boldsymbol{\gamma\zeta} + 295.8\boldsymbol{\gamma\eta} - 72.8\boldsymbol{\delta\varepsilon} + 268.8\boldsymbol{\delta\zeta} - 119.7\boldsymbol{\delta\eta} + 20.1\boldsymbol{\varepsilon\zeta} + 88.3\boldsymbol{\varepsilon\eta} + 934.6\boldsymbol{\zeta\eta} + 61.6\boldsymbol{\alpha^2} - 501.0\boldsymbol{\gamma^2} + 260.9\boldsymbol{\delta^2} - 71.4\boldsymbol{\varepsilon^2} - 939.7\boldsymbol{\zeta^2} - 604.2\boldsymbol{\eta^2} + 1.2\boldsymbol{\gamma\zeta^2} + 8.8\boldsymbol{\gamma^3} - 2.2\boldsymbol{\delta^3} \quad \textbf{4.21}$$

The Non-Interaction (R) and Interaction (R) models were compared based on the Akaike Information Criterion (AIC) to identify the best model in terms of goodness of fit and number of parameters. Comparisons of the models were conducted based on their quantile. By doing so each model corresponding to a quantile will be analysed and the best ones elicited. The results of the comparison are tabled below in Table 4-5.

Model/Quantile	0.05	0.25	0.5	0.75	0.95
Non-Interaction (R) Model	5264.3	5297.4	5388.7	5430.2	5521.2
Interaction (R) Model	5553.7	5610.7	5658.8	5696.8	5672.7

Table 4-5: AIC Values for the Different Quantiles Corresponding to Each Non-Interaction (R) and Interaction (R) Model

As can be seen from the values obtained, the Non-Interaction (R) model gives a lower index across the quantiles whereas the Interaction (R) model shows high AIC values. This indicates that the model gives a better balance

between goodness of fit and model complexity. The Interaction (R) model however, showed high AIC values which signals a model with relatively high variance.

The derivative of equations 4.12 to 4.21 with respect to the %IR is summarised in the table below.

Model / Quantile	Regularization Parameter, λ	$\frac{\partial TCDA}{\partial \%IR}$
Non-Interaction 0.95 (R)		70.96α
Non-Interaction 0.75 (R)		35.82α
Non-Interaction 0.5 (R)	11.346	23.82α
Non-Interaction 0.25 (R)		12.96α
Non-Interaction 0.05 (R)		3.22α
Interaction 0.95 (R)		$163.6\gamma + 72.2\delta - 92.5\varepsilon + 326.3\zeta + 42.1\eta + 123.2\alpha$
Interaction 0.75 (R)		$160.9\,\gamma + 51.7\delta - 35.4\varepsilon + 513.4\zeta - 340.7\eta + 90.6\alpha$
Interaction 0.5 (R)	2.209	$89.8\gamma + 31.6\delta - 3\varepsilon - 16.2\zeta + 73.9\eta + 30.6\alpha$
Interaction 0.25 (R)		$77.4\gamma - 14.5\delta - 2.6\varepsilon + 10.5\zeta + 53.8\eta + 8.4\,\alpha$
Interaction 0.05 (R)		$57.3\gamma + 7.1\delta - 7.8\varepsilon + 71.7\zeta + 6.8\,\alpha$

Table 4-6: Derived Models for Various Quantiles and their Respective Regularisation Parameter Values

The Non-Interaction model proves much simpler compared to the earlier original quantile regression models. With the tuning out of the %IR variable and after it has gone through the same steps as the interaction model stated above, what's left of the models are coefficients of the linear function. It must be noted that the values of the coefficients in this thesis are subject to the

MEOC data. More comprehensive datasets in the future regarding ECDA and DCVG will further refine and generalise the models where they can be applied to different environments from what we have here.

For the Interaction model, the phasing out of the %IR variable resulted in models containing only higher order terms. Since the outcome (after the LASSO process) are non-linear in nature, partial derivation of the models was done with respect to %IR to obtain the values of the estimated TCDA for each increasing %IR. The derivatives use values of 10.5%, 35.2%, 109.5 cm, 32.5 years and 35.3 inch for the POPD, DUC, DOC, TIS and PS variables respectively for the TCDA estimation. These values are the mean of each variable.

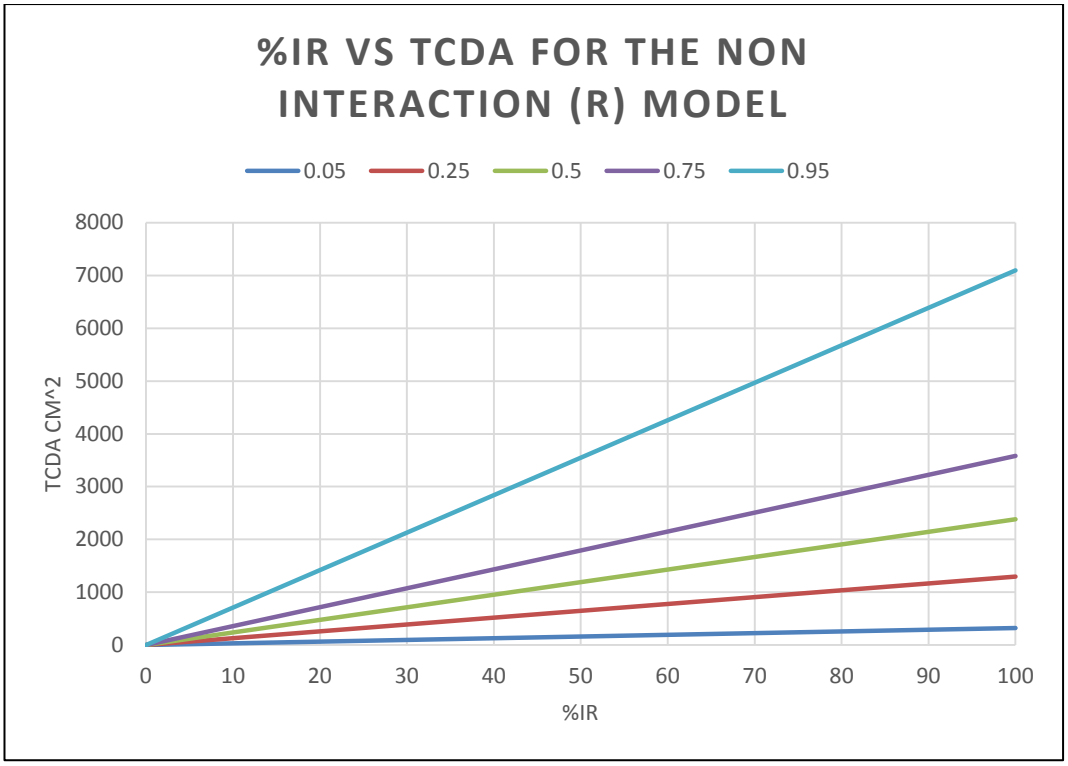


Figure 4-14: Estimated TCDA based on %IR for the Non-Interaction Model after LASSO

Figure 4-14 shows the estimated TCDA based on the %IR values for the Non-Interaction models. The figure provides a simple outlook on estimation of the size of coating defects by relating the slope of each model by a singular coefficient. For the 0.05 quantile, the estimated sizes do resemble those of Figure 4-3 where most of the estimated TCDA is zero or close to it. For the 0.95 quantile, the slope of the model seems to be disproportionate to other quantiles where it registers as having the biggest coefficient value. This can be explained by the way in which the TCDA data is spread with regards to %IR and also by referring to Figure 4-3 where the highest quantile highlights the largest coefficient values. The overall trend of the plot shows that increasing %IR is related to a bigger TCDA value which concurs with general understanding of the technique.

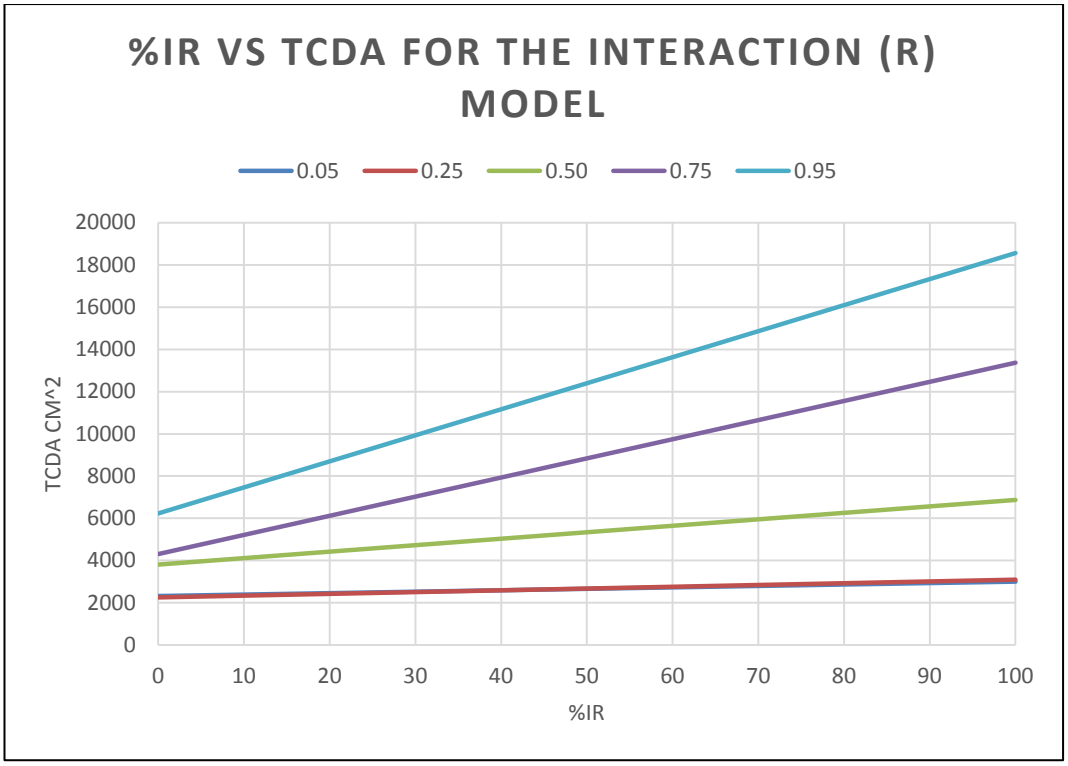


Figure 4-15: Estimated TCDA based on %IR for the Interaction Model after LASSO

In Figure 4-15, it is observed that the estimated TCDA follows the general trend found in Figure 4-14. However, the origin of the TCDA estimation does not start at zero due to the contributing factors of the interaction variables. The effect of the interaction variables can also be seen by the larger estimated values as compared to the Non-Interaction model. Additionally, the range of predictive values is closer to Figure 4-3 than to Figure 4-14. Interestingly, the 0.05 and 0.25 quantiles illustrate the same trend as that in Figure 4-3 where the two lowest quantiles seem to overlap due to the large number of zero TCDAs occurring across the spectrum of %IR. Overall, both models in Figure 4-14 and Figure 4-15 look concise and compact (due to linearity) with a more general approach to the MEOC data.

The generalised models with LASSO indicate a shrinkage in the coefficient values as expected. But choosing the appropriate model to represent the true underlying mechanism of coating defects is a challenge. The lower quantiles (0.05 and the 0.25) of the Interaction model (R), Figure 4-15, showed some similarity in trend when compared to the nonlinear models highlighted in Figure 4-3. However, it is assumed that most data with zero or close to zero readings for TCDA found in the MEOC dataset is attributable to error, i.e. stray or foreign currents have contributed to the %IR readings giving erroneous %IR values. Although this may be true, these errors resemble the true nature of how the MEOC data is spread out. The Non-Interaction models in Figure 4-14, however paint a picture of a more generalised model which is less affected by these “errors”.

The estimated quantiles for the Non-Interaction models (R) is an appropriate representation of the general relationship between %IR and TCDA. The model’s higher quantiles are more spread out, illustrating a more generalised estimation. This can also be said about the lower quantiles where the increasing trend is roughly proportional (evenly spaced) between one quantile to the next. As for the Interaction model’s (R) estimates, the quantile’s prediction is higher. The trend shows similarity when compared to Figure 4-3 where lower quantiles are packed closely together while the upper quantiles are more dispersed.

Both the models are subject to the same number of data points from the MEOC dataset. However, the Non-Interaction (R) and the Interaction (R)

models use different number of variables to predict TCDA. As the number of variables increases (in the case of the Interaction model (R)), the likelihood of the model achieving low errors by going through (fitting) all the data points are expected to increase. Such a model's application to other datasets is severely compromised since the model overfits the dataset in hand. As mentioned earlier, this can be observed by looking at the similarities between the Interaction model and Figure 4-3. Additionally, the results in Table 4-5 illustrate the AIC indices for the Non-Interaction (R) model are lower than the Interaction (R) model across the quantiles. This underpins the notion that the Non-Interaction (R) model is more general and balances the number of variables and the model fit. The AIC indices for the Interaction (R) model seems to suggest the model is overfitting the data where high variance is likely.

Due to this, it is safe to suggest that if one were to apply these models to other pipelines (which are not part of MEOC dataset) then the Non-Interaction (R) model is more suitable. As in the case of subsequent ECDA assessments on the MEOC pipelines, the Interaction (R) model is more suitable since it "follows" the structure of the MEOC dataset. However, application should be done with care since the estimated values of the TCDA is lower than the actual due to the shrinkage process of the coefficients.

Chapter 5

Analyses of DCVG Indications and Coating Defect Size Prediction with Bayesian Quantile Regression

5.0 Introduction

In the previous Chapter, the MEOC data were used to model the relationship between the variables by using classical quantile regression. The main objective of that Chapter is to produce a model to predict the size of TCDA with regard to its contributing factors. In this Chapter, however, a Bayesian approach is used. The Bayesian method provides an alternative to the classical method for parameter estimation.

Lately, Bayesian inferences combined with Markov Chain Monte Carlo (MCMC) algorithms have become increasingly popular. In the context of quantile regression, the Bayesian approach was first elaborated by Yu and Moyeed in [79]. The paper suggested that Bayesian inferences is more advantageous than the classical approach in mainly two instances; 1) Bayesian statistics does not rely on asymptotic variances of the estimators and 2) The estimated parameter includes the parameter uncertainty in the form of a posterior distribution.

The distribution of the response variable from the MEOC data, the TCDA and the %IR variable demonstrates a distribution which is neither normal nor symmetric. Distributions which are not symmetrical need more complex solutions to describe the entirety of the response variable's distribution [88]. The usage of quantile regression (QR) on distributions such as the TCDA and %IR are most effective due to these distributions being asymmetric (skewed) [89], [90].

The Bayesian approach to quantile regression is used to characterise the %IR and TCDA distributions. Benefits of using such an approach include easy interpretation of the outcome based on the posterior distribution. Additionally, the resulting posterior distribution has the potential to be used as the prior distribution for the next Bayesian assessment for the ECDA process. Implementing the Bayesian method for the ECDA process (through successive means) will ensure the parameter estimation becomes more accurate with each iteration step. This will provide operators with a better

understanding of the pipelines in terms of corrosion and coating defects and ultimately uphold its integrity.

5.1 Middle Eastern Oil Company (MEOC) Data

The data used for the BQR analysis are taken from the MEOC project. In brief, TWI Ltd., was appointed by the Middle Eastern Oil Company (MEOC) as the contractor for conducting integrity assessments on 9 of their pipelines. These pipelines are not piggable and hence the ECDA approach (as suggested by the NACE SP0502 [8]) was chosen as the best method for the assessment. The main idea of an ECDA is to assess and reduce the impact of external corrosion to the structural integrity of buried pipelines. The method consists of 4 phases/steps where each step should be completed in turn. The variables from the MEOC data used for the BQR analysis are given in Table 5-1 below. A more comprehensive description of the data and the ECDA process is given in Chapter 2 and Chapter 4 (Table 4-1).

Symbol	Variables Considered	Type of Variable
α	IR Drop (%IR)	Continuous
β	Soil Resistivity (SR)	Continuous
γ	Percentage of Pit Depth to Wall Thickness (POPD)	Continuous
δ	Deposits under Coatings (DUC)	Continuous
ε	Depth of Cover (DOC)	Continuous
ζ	Time in Service (TIS)	Continuous
η	Pipe Size (PS)	Continuous
C	Backfill Type (Clay)	Categorical
θ	Backfill Type (Rock)	Categorical
κ	Backfill Type (Sand + Clay)	Categorical
λ	Backfill Type (Stones + Clay)	Categorical
CW	Coating Type (PVC Cold Wrap)	Categorical
μ	Coating Type (Coal Tar)	Categorical
ξ	Coating Type (Polyethylene)	Categorical
R	Backfill Geometry (Round)	Categorical
ρ	Backfill Geometry (Angular)	Categorical
σ	Backfill geometry (Round + Angular)	Categorical
φ	pH Of Water in Soil (Acidic)	Categorical
χ	pH Of Water in Soil (Alkaline)	Categorical
ψ	pH Of Water in Soil (Neutral)	Categorical
ω	pH Of Water Underneath Coating (Acidic)	Categorical
$\ddot{u}$	pH Of Water Underneath Coating (Alkaline)	Categorical
$\ddot{\imath}$	pH Of Water Underneath Coating (Neutral)	Categorical

Table 5-1: Variables Considered for the Model

The objectives for this Chapter can be divided into two. The first is the construction of a model which summarizes the contributing factors towards the DCVG indication. A further refinement of the model is also constructed based on the general understanding on the system. Variables such as the Percentage of Pit Depth to Wall Thickness (POPD), Deposits under Coatings (DUC), Time in Service (TIS) and pH Of Water Underneath Coating were omitted from the original model. The purpose for this refinement is to have a model which is concise and simple. Therefore, application of the model in future ECDA projects will be straightforward and not computationally expensive. Additionally, the data of the omitted variables are obtained mainly through the excavation of the pipe (except for TIS) for further investigation. By not considering these data, one is able to make use of data collected before the excavation. Hence, the model can also be viewed as an input into consideration of where to excavate.

The second objective is to present a model which predicts the TCDA based on environmental and DCVG factors. The data used for this objective were also taken from the MEOC project. Two version of the dataset exists. The 1st version is the dataset that includes all the measurements. We shall name this “Oriset”. This is the original dataset received by the author. The 2nd version of the data is called “Filtset” which is “filtered” by the author’s expert judgment on what to expect from a DCVG indication. A total of 4 data points considered as outliers were taken out of the assessment. Apart from that, everything remained the same. The model and factors considered in this

Chapter are given below in Table 5-2. For ease of referencing the models are named as follows.

Description of Model	Dataset	Model Name
Contribution to %IR Model - Full	Oriset	**Model 1**
Contribution to %IR Model – Refined	Oriset	**Model 1a**
Contribution to %IR Model - Full	Filtset	**Model 2**
Contribution to %IR Model - Refined	Filtset	**Model 2a**
TCDA Model	Oriset	**Model 3**
TCDA Model	Filtset	**Model 4**

Table 5-2: Names of the Various Models Corresponding to Each Dataset

Two techniques were applied to the data. The first approach is by the usage of the BQR to obtain the model estimates. All discussion in the next section will refer to this approach. The second is the classical frequentist approach which employs classical quantile regression. The classical approach is used only to identify differences in the two techniques when compared to each other.

5.2 Model Estimation and Result Analysis

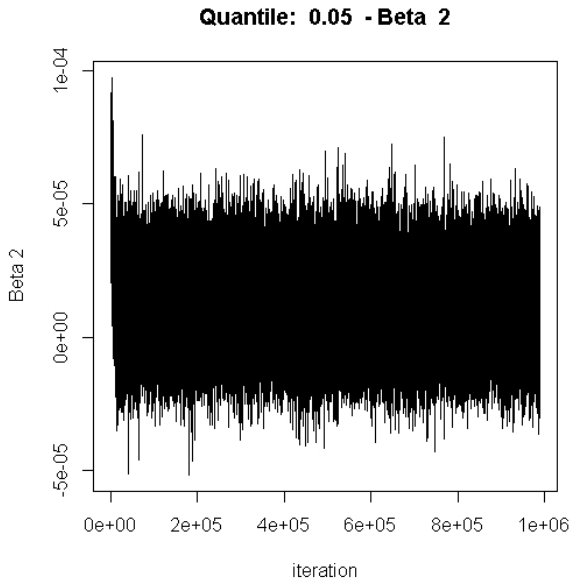
This section presents the results of the BQR technique applied to the MEOC data. The objective of this Chapter (as with other Chapters in this thesis) is to come up with an initial working model capable of incorporating (provision)

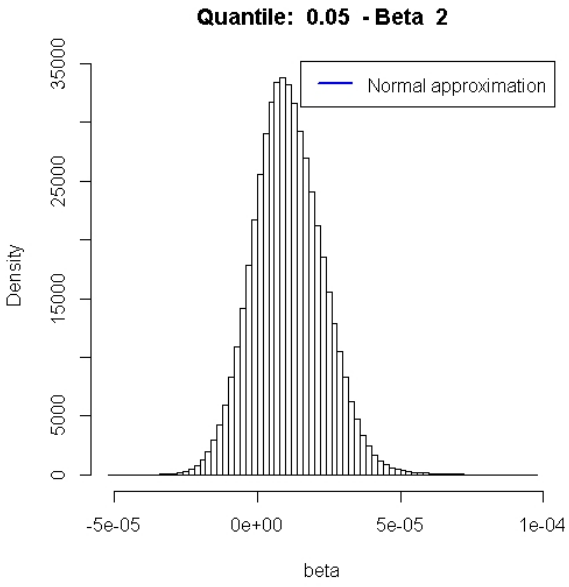
future inspection results for pipeline operators and regulators to make informed decisions on the prioritisation of coating defects as well as determining excavation locations. Before we can do that, it is important to understand the DCVG indications and what are the contributing factors that have given rise to such measurements. Based on common industry understanding, a DCVG indication arises from the equipotential lines as a result of the cathodic protection current flowing to a coating defect on the pipeline. It was previously thought that the equipotential lines are only subject to coating defect size with larger defects emitting larger potentials. Work by McKinney [64] shows factors such as the SR contribute positively to the generated %IR signal. However, the data used for that assessment uses simulated data and is anticipated to be different from the assessment conducted here. All the analyses were done in the statistical software R by using the BQR package. All of the analyses in this Chapter use the same 95% value for their respective credible intervals.

5.2.1 Contributing Factors to %IR (Model 1)

The estimates of variables from Table 8-1 show interesting results particularly for the TCDA variable. Iterations up to 1 million were conducted to achieve convergence. This can be seen in Figure 5-1 where the trace plot and posterior histogram of various quantiles are presented. The BQR technique estimated very low values across the whole of the %IR distribution. The maximum estimated coefficient value occurs at the 0.5 quantile which is a 1 cm^2 increase in coating defect size reflected in an

increase of 0.0000687% IR indication. If we increase the percentage values to 100% (maximum reading of a DCVG indication), the maximum coating defect size that the DCVG technique is able to detect is 1,455,604 cm^2. The lowest estimated value for the TCDA occurs at the 0.05 quantile. The estimated coefficient reveals a 1 cm^2 increase in TCDA will increase the %IR value by 0.0000022%. This shows that medium size coating defects give the largest signal on the DCVG indication whereas small defects contribute the least. Also, at the 0.5 quantile, the credible interval is much narrower compared to those at the two opposite ends indicating lower uncertainty.





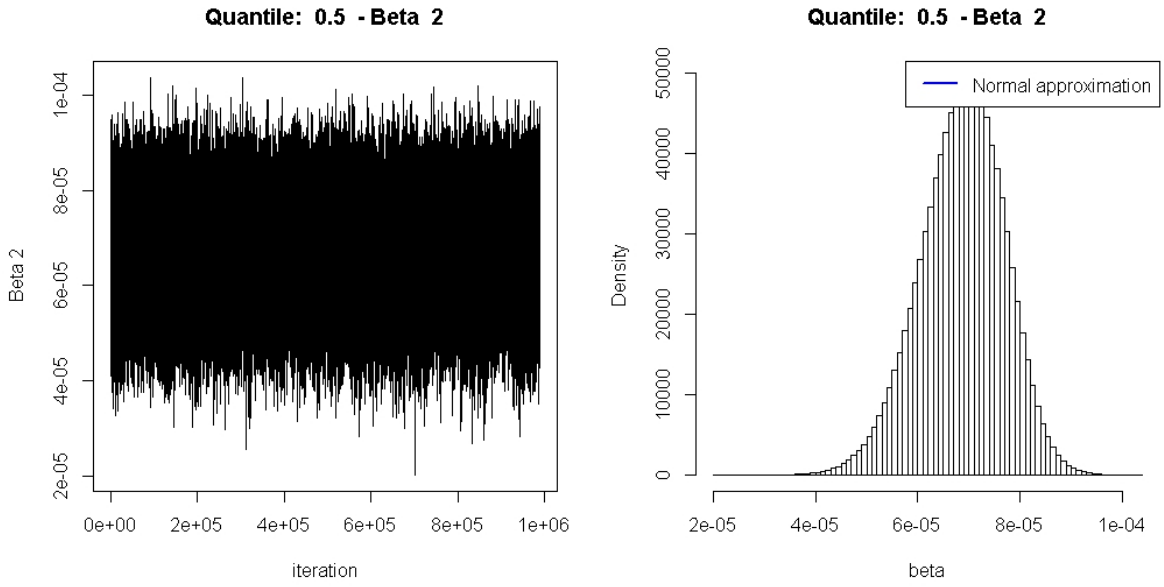


Figure 5-1: Trace Plots and Posterior Histogram of the 0.05 and 0.5 Quantile for the Estimated Coefficient, TCDA for Model 1

Equations of various quantiles are presented in the following.

$$\%IR_{0.05} = 14.2 + 0.0000022TCDA - 0.0000235\beta + 0.00611\gamma + 0.0079\delta + 0.0549\varepsilon - 0.336\zeta - 0.0818\eta + 5.2\theta - 1.03\kappa + 1.72\lambda - 3.26\mu - 6.28\xi + 0.754\rho - 2.64\sigma + 1.17\varphi + 8.41\chi + 7.24\psi - 0.943\omega - 2\ddot{\upsilon} + 2.56\ddot{\imath} \quad \textbf{5.1}$$

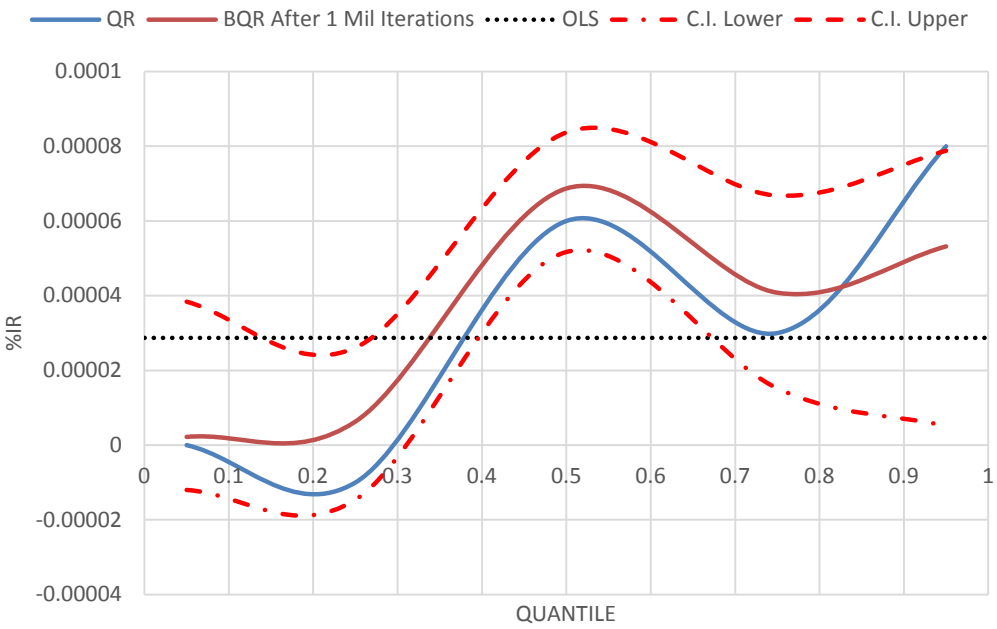
$$\%IR_{0.5} = 86.1 + 0.0000687TCDA - 0.000567\beta + 0.0439\gamma - 0.0372\delta + 0.0933\varepsilon - 0.374\zeta - 1.31\eta + 50.8\theta + 16.3\kappa + 0.562\lambda - 0.215\mu + 0.368\xi - 19.9\rho - 0.835\sigma - 8.1\varphi + 0.753\chi + 7.03\psi - 3.24\omega - 7.78\ddot{\upsilon} - 0.125\ddot{\imath} \quad \textbf{5.2}$$

$$\%IR_{0.95} = 23.6 + 0.0000532TCDA - 0.000346\beta + 0.108\gamma - 0.0704\delta \quad \textbf{5.3}$$
$$+ 0.0364\varepsilon + 1.19\zeta + 0.285\eta + 10.7\theta - 11.6\kappa + 0.411\lambda$$
$$+ 11\mu + 2.44\xi - 8.69\rho - 0.446\sigma - 4.8\varphi - 0.804\chi$$
$$- 14\psi - 11\omega + 0.991ü - 1.99ï$$

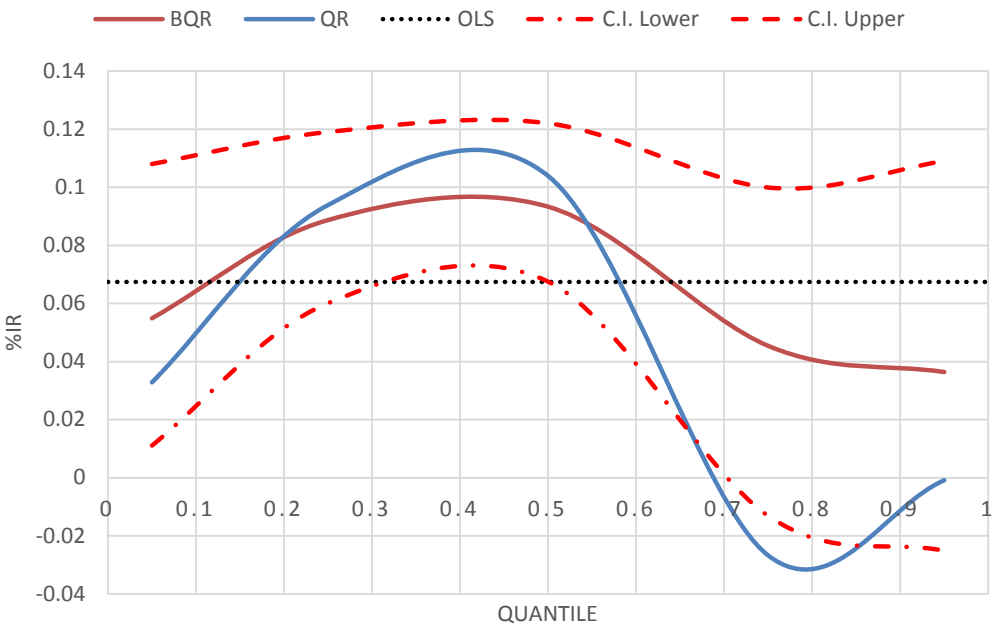
SR also plays a role in the contribution to %IR. Across the quantiles the Bayesian estimates show low negative values. The maximum (lowest) estimated value for SR occurs at the 0.5 quantile with a value of -0.000567. This can be interpreted as a 1 unit increase of SR leading to a decrease of 0.000567% with respect to %IR. However, the variable Backfill Type – Rock which is related to the resistant nature of the soil, shows an inverse effect. Across the quantiles, the estimated coefficients point to meaningful contributions to the %IR readings especially within the quantile range 0.25 to 0.75.

Another interesting variable is the DOC, i.e. the depth of the buried pipe. The estimated coefficient increases in value as one approaches the 0.5 quantile where it maximizes with an estimated value of 0.0933. This means that for 1 unit increase in the depth of pipe, the value of the %IR will also increase by 0.0933%. After this it decreases from the 0.75 quantile to the 0.95 quantile.

TCDA



DEPTH OF COVER



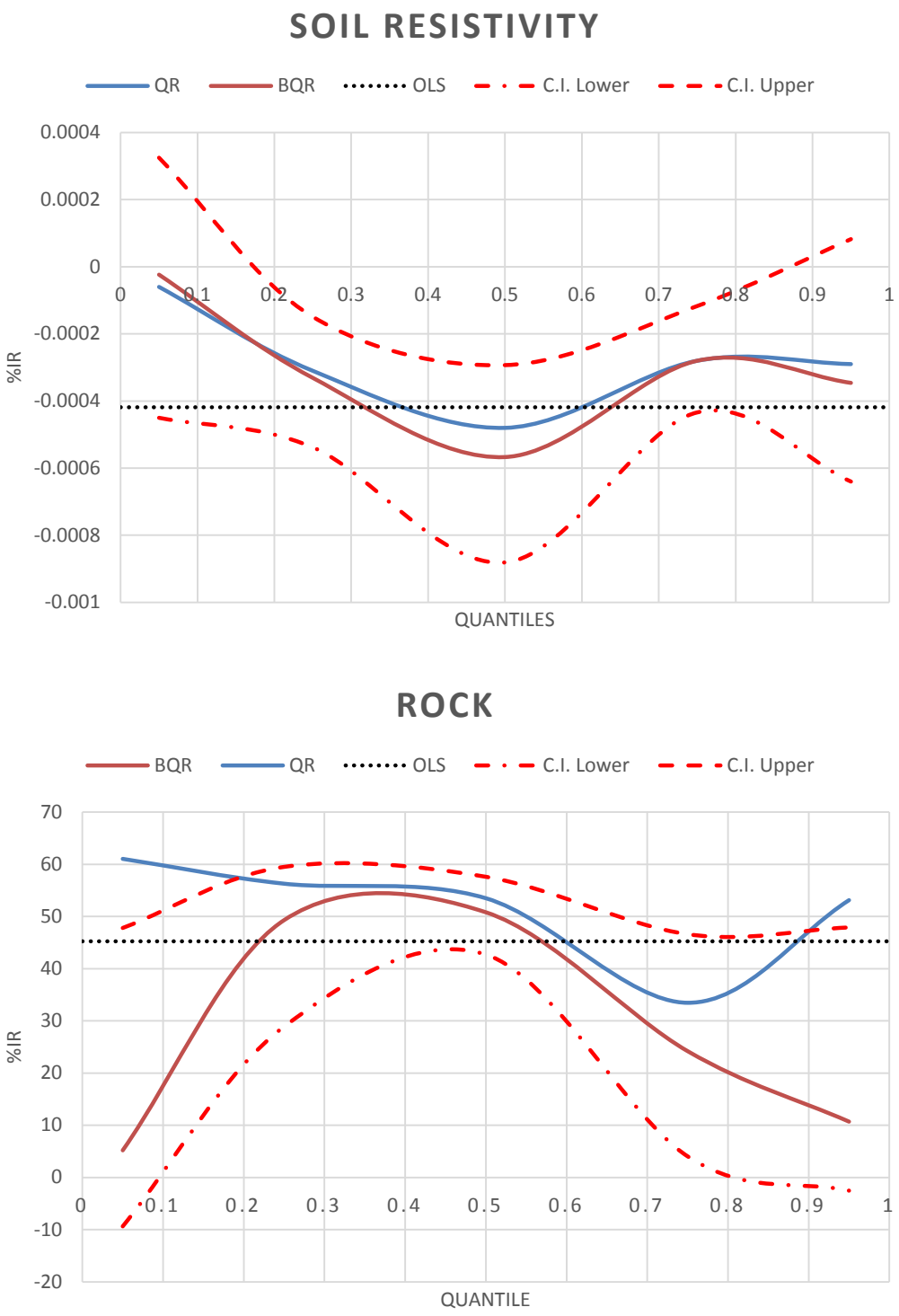


Figure 5-2: Quantile Plots of various Variables of Interest for Model 1

5.2.2 Refined %IR (Model 1a)

The results of the estimated coefficients by BQR for Model 1a are presented in Table 8-3 below. In achieving convergence for all the variables, iterations of up to 300,000 were determined with the initial 5,000 steps regarded as burn-in. For the variable of interest, TCDA, the maximum estimated value occurs at the 0.5 quantile. This prediction is similar to that predicted by Model 1. The overall estimated trend also follows the same pattern as Model 1 with Model 1 being more pronounced. The value of the coefficient at the maximum is 0.0000828. 1 cm^2 of TCDA will have an effect on the %IR of 0.0000828%. At the 0.05 quantile, the coefficient value is at its lowest of -0.0000353. The negative value signifies that a 1 cm^2 increase in TCDA will yield a 0.0000353% decrease in %IR. The trace plots and corresponding posterior histogram are shown below.

The trend of the estimated coefficients for the variable SR is also similar to Model 1. From 0.25 quantile upwards, the trend is negative with its most negative at the 0.5 quantile. The reason for this can be considered consistent with the assessment for Model 1 when one looks at the Rock variable with most of the estimates showing high positive values. The peak is also found at the 0.5 quantile suggesting that the effect of having coarse-grained soil affects %IR values at its median quantile. There is also the factor of heterogeneity of the soil itself which also contributes to the non-linearity effect on certain quantiles of the %IR distribution.

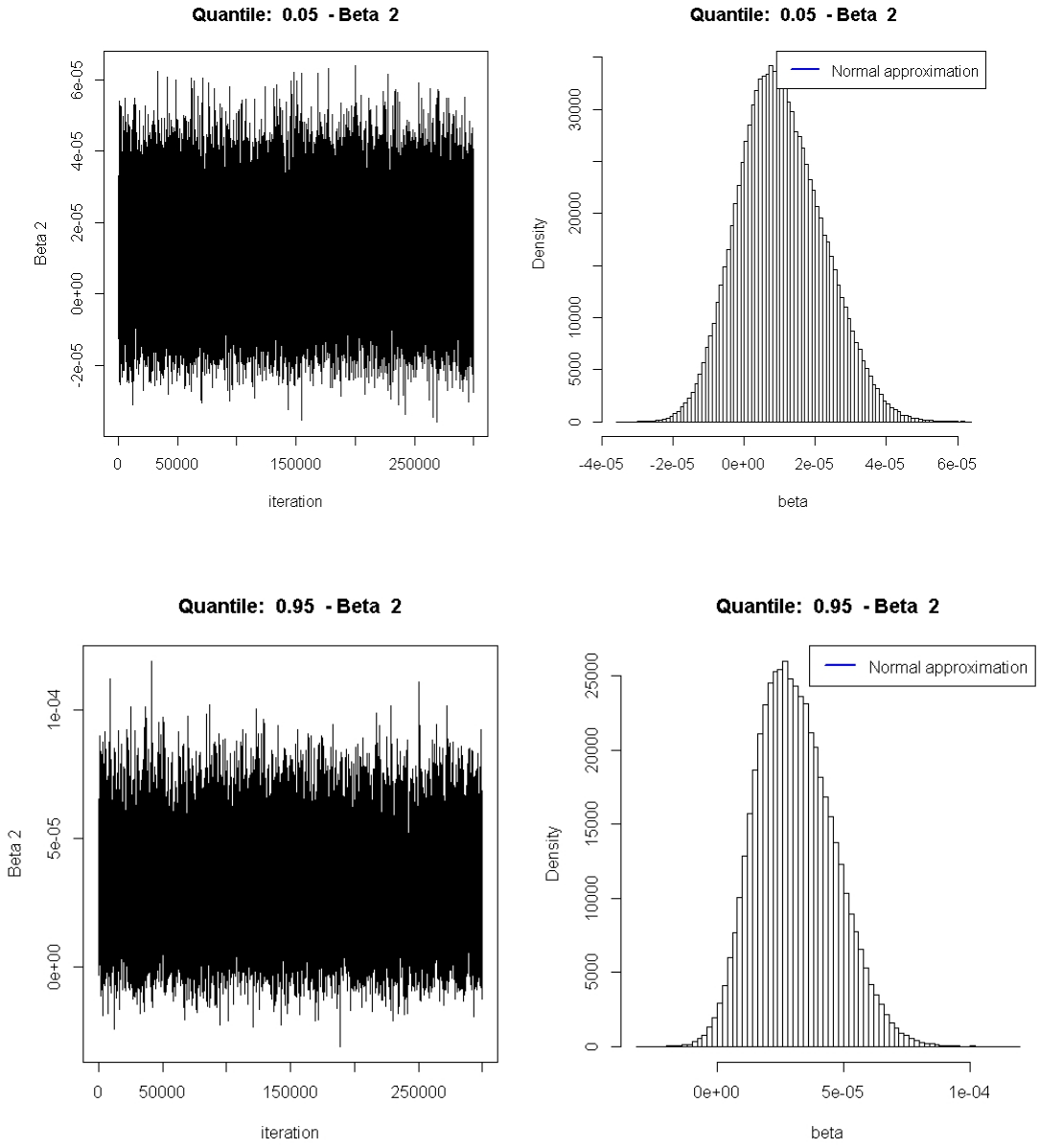


Figure 5-3: Trace Plots and the Posterior Histogram of the 0.05 and 0.95 Quantile for the Estimated Coefficient, TCDA for Model 1a

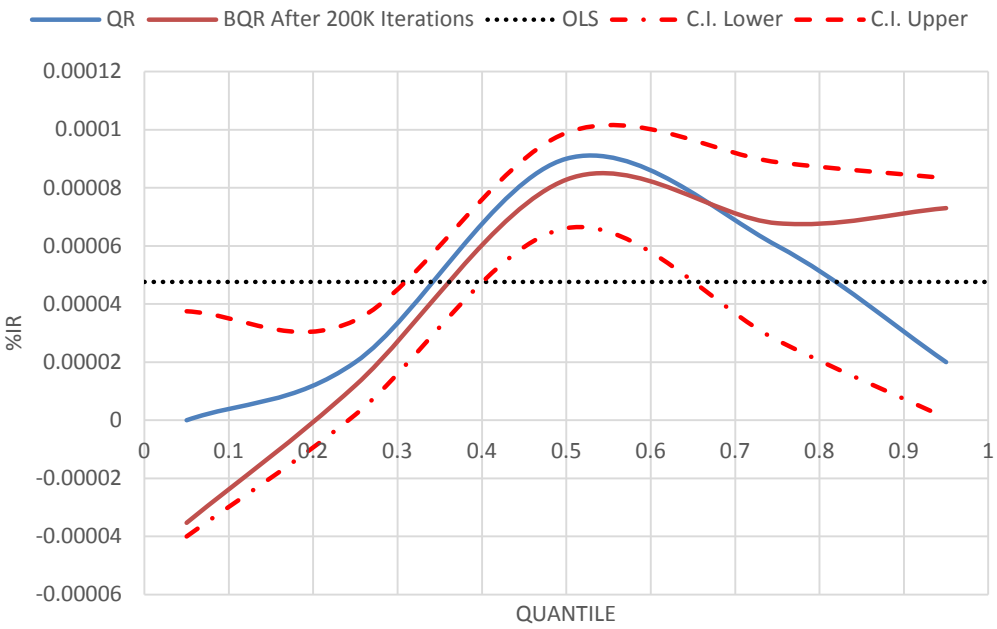
Equations of various quantiles are presented below:

$$\%IR_{0.05} = 4.74 - 0.0000353TCDA + 0.000000565\beta + 0.0508\varepsilon - 0.158\eta + 5.23\theta - 0.939\kappa + 1.56\lambda + 2\mu - 0.113\xi + 0.434\rho - 3.65\sigma + 1.2\varphi + 8.11\chi + 5.06\psi \quad \textbf{5.4}$$

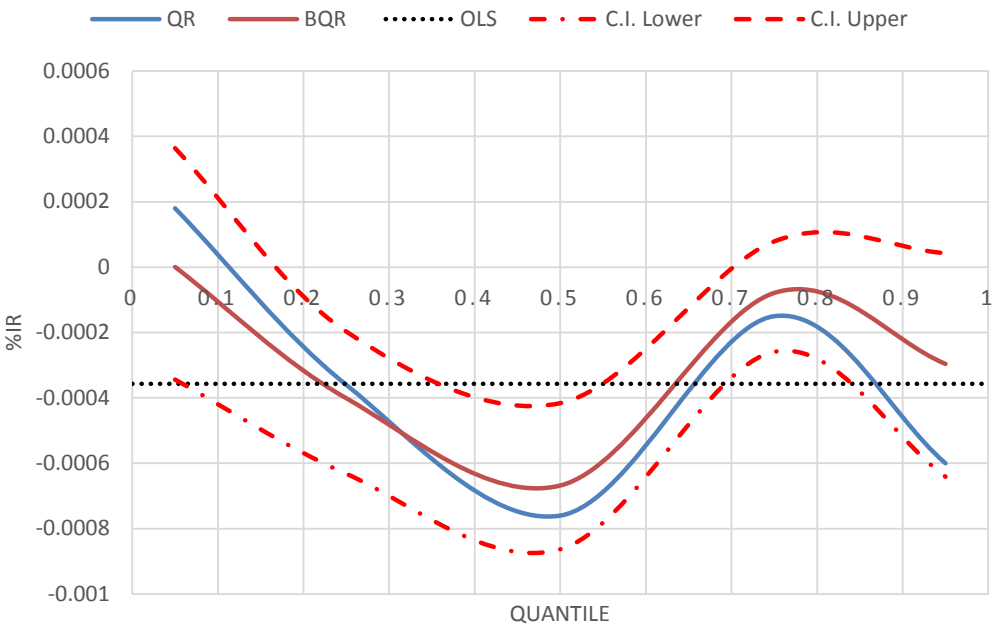
$$\%IR_{0.5} = 87.5 + 0.0000828TCDA - 0.000668\beta + 0.0722\varepsilon - 1.77\eta + 53.4\theta + 25.4\kappa + 0.619\lambda + 5.54\mu + 6.77\xi - 18.2\rho + 0.251\sigma - 6.07\varphi + 1.76\chi + 1.14\psi \quad \textbf{5.5}$$

$$\%IR_{0.95} = 64.9 + 0.000073TCDA - 0.000296\beta + 0.0228\varepsilon + 0.432\eta + 6.45\theta - 8.21\kappa - 1.46\lambda - 6.78\mu - 15.3\xi - 6.73\rho - 1.24\sigma - 4.67\varphi - 0.575\chi - 12.6\psi \quad \textbf{5.6}$$

TCDA (RED.)



SOIL RESISTIVITY (RED.)



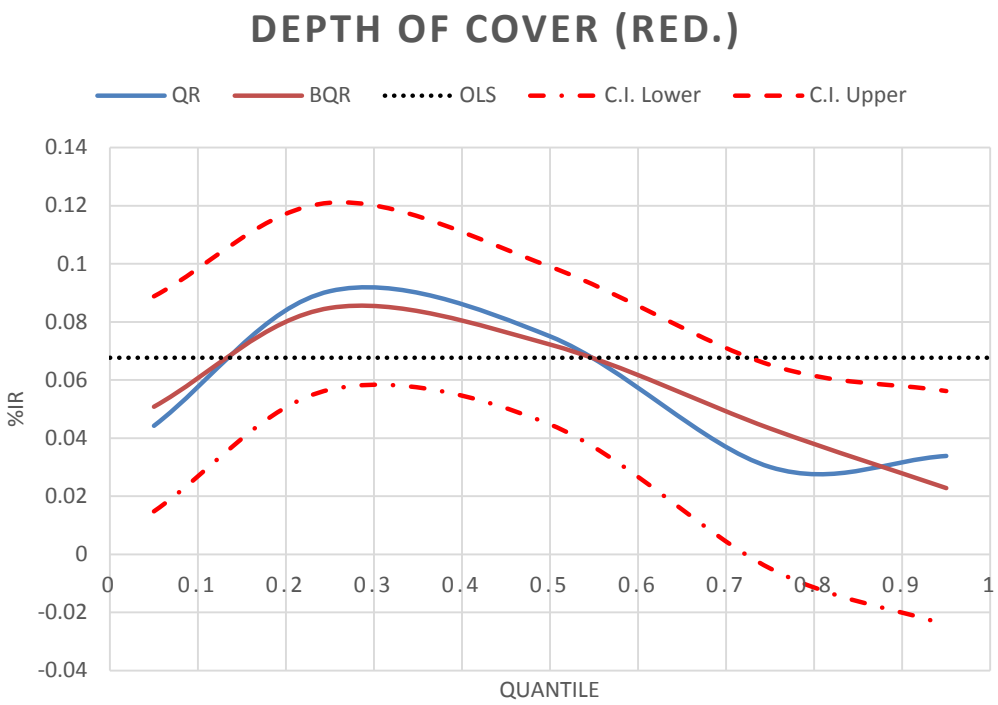


Figure 5-4: Quantile Plots of various Variables of Interest for Model 1a

The DOC seems to show the same behaviour one would expect from a DCVG inspection. Starting from the 0.25 quantile, the decreasing trend indicates that by increasing the depth of the pipe, the amount of signal that would be picked up by the DCVG voltmeter would be less. At the 0.95 quantile, a one unit increase in depth would have an effect of an increase to the %IR of 0.0228%. Compare this to the 0.25 quantile where registering a one unit increase in depth of the pipeline would increase the %IR by 0.084797% (stronger effect at this quantile).

5.2.3 Contributing Factors to %IR (Model 2)

The estimated coefficients for Model 2 are given in Table 8-2. However, the reference variable for the categorical variable is substituted to be Backfill Type – Rock, Coating Type – Polyethylene and Backfill Geometry – Angular. This is due for investigation amongst the factors concerning soft soils which include clay with rounded grain structure. A total of 400,000 iterations were made to get to the point of convergence with the initial 5,000 readings as burn-in. Table 8-2 shows the TCDA variable coefficient has an upward trend with a slight dip at the 0.25 quantile. The highest value is reached at the 0.95 quantile with a value of 0.000229%. For a 1 cm^2 increase in the coating defect area, a 0.000229% increase in %IR is expected. This is higher than the maximum obtained by the Model 1. Additionally, this happens at the 0.95 quantile which fits well with common understanding of the technique, in contrast to Model 1 where the maximum occurred at the 0.5 quantile. This is mainly due to the contribution of the expert intervention of the author which obliterated 4 points from the original set.

Estimated coefficients for the SR variable show an increasing trend from the 0.25 quantile up to the maximum at the 0.95 quantile. The maximum Bayes estimate is 0.000373. Therefore, for a 1 unit increase in SR, an increase of 0.000373% of %IR is expected. Moreover, large uncertainties were observed at the upper and lower ends of the quantiles as compared to the median region.

The variable clay shows an increasing trend across the %IR distribution with a dip at the 0.95 quantile. The maximum estimated coefficient was noted at the 0.75 quantile with a value of 60.8. This can be translated as the effect of clay on %IR will be greatest at the 0.75 quantile of the %IR distribution.

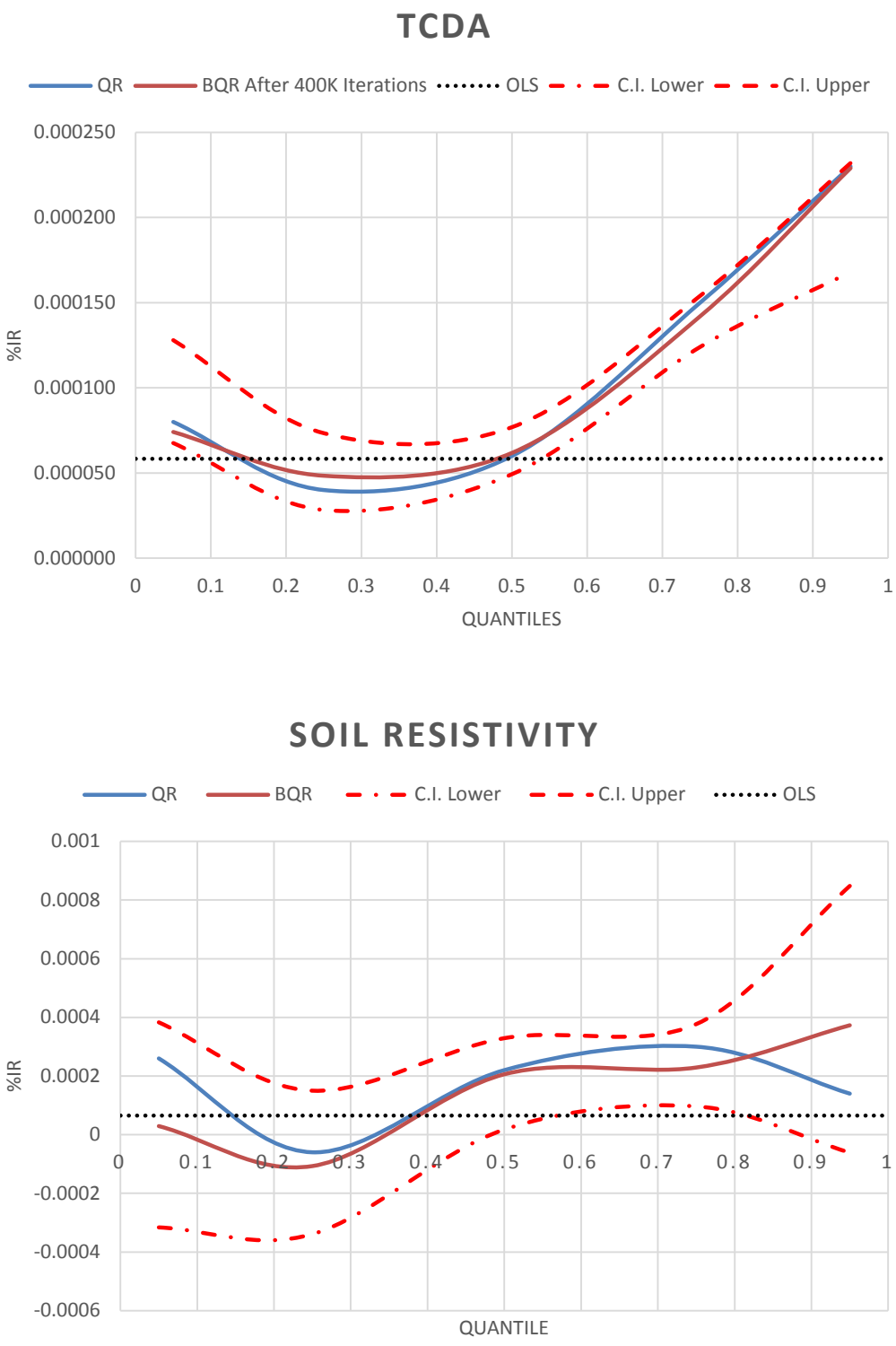


Figure 5-5: Quantile Plots of various Variables of Interest for Model 2

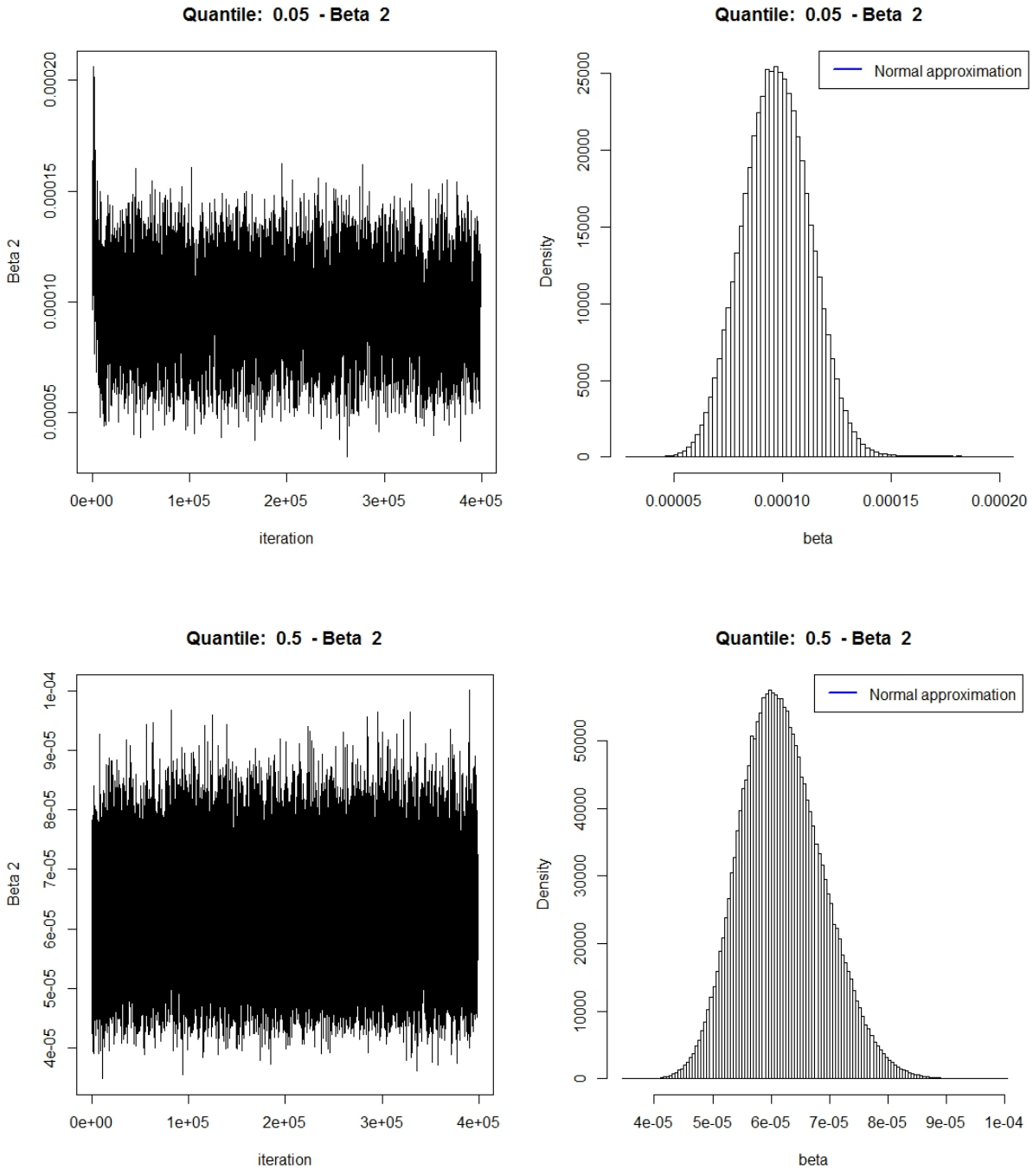


Figure 5-6: Trace Plots and the Posterior histogram of the 0.05 and 0.5 quantile for the Estimated Coefficient, TCDA for Model 2

The following are selected models (Model 2) for the contribution of %IR based on various quantiles.

$$\%IR_{0.05} = 14.7 + 0.0000741TCDA + 0.0000293\beta + 0.0334\gamma - 0.0209\delta + 0.0668\varepsilon - 0.116\zeta - 0.126\eta - 2.23C - 1.63\kappa + 11.4\lambda - 9.26CW + 7.57\mu + 1.73R - 0.246\sigma + 3.98\varphi - 1.84\chi - 11.4\psi + 2.8\omega - 1.62\ddot{\upsilon} - 4.96\ddot{\iota} \quad \textbf{5.7}$$

$$\%IR_{0.5} = 79.4 + 0.0000618TCDA + 0.000206\beta + 0.161\gamma - 0.0373\delta - 0.00696\varepsilon - 0.234\zeta - 0.3\eta + 3.35C - 16.7\kappa - 3.29\lambda - 31.4CW - 1.45\mu + 1.02R - 0.156\sigma + 7.4\varphi - 21.2\chi - 11.2\psi + 1.02\omega - 8.36\ddot{\upsilon} - 6.67\ddot{\iota} \quad \textbf{5.8}$$

$$\%IR_{0.95} = 22 + 0.000229TCDA + 0.000373\beta + 0.0558\gamma - 0.05\delta + 0.0682\varepsilon + 0.179\zeta + 0.186\eta + 16.3C + 4.9\kappa + 25.4\lambda - 5.69CW + 29.7\mu - 1.58R + 1.58\sigma + 9.2\varphi + 2.62\chi - 16.8\psi - 4.98\omega - 3.43\ddot{\upsilon} - 13.1\ddot{\iota} \quad \textbf{5.9}$$

5.2.4 Refined %IR (Model 2a)

Table 8-4 shows the estimated coefficients predicted by the BQR method with the Filtset data for Model 2a. 400,000 iterations were made with the initial 5,000 recordings regarded as burn-ins. Keeping with the theme of the thesis, the variable TCDA is the prime interest since it represents the bedrock on which the DCVG technique was built. At the 0.05 quantile, the predicted coefficient shows similar behaviour to that obtained for Model 2. The coefficient value drops at the 0.25 quantile and rises steadily after this up to the 0.95 quantile where it reaches its maximum. The maximum

predicted value stands at 0.000221 which means a 1 cm^2 increase in TCDA will give an increase of 0.000221% in %IR. Similar characteristics were observed previously for Model 2, with only slight difference in the predicted values.

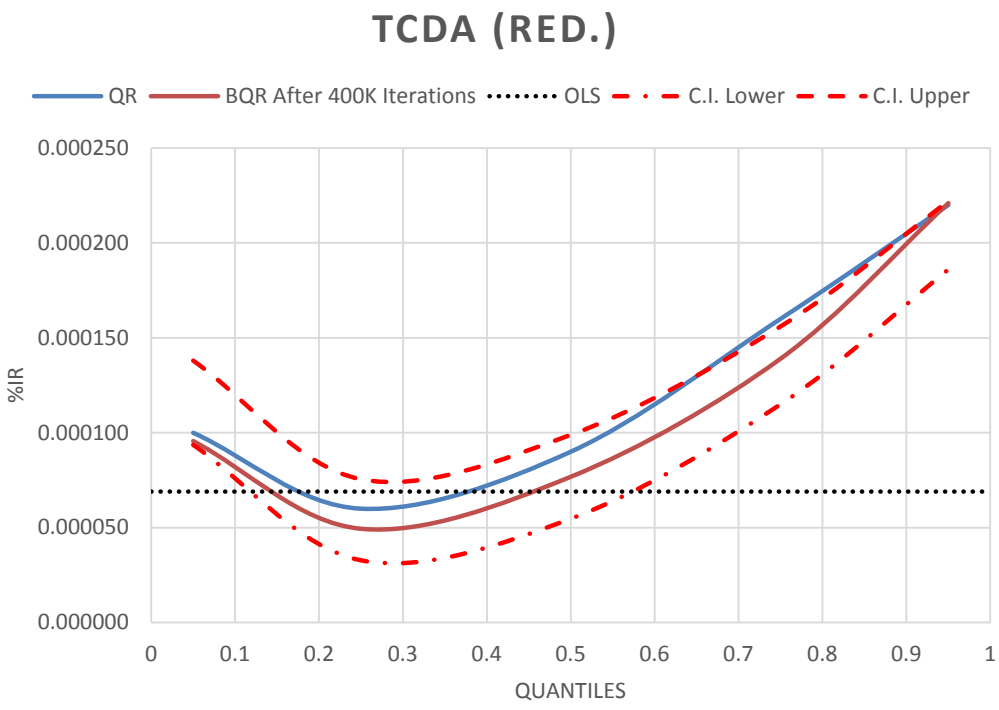


Figure 5-7: TCDA Quantile Plot for Model 2a

SR plays a role in Model 2a where an increasing trend is observed starting from the 0.25 quantile all the way to the 0.95 quantile. The highest predicted value is at the 0.95 quantile with a Bayes estimate of 0.000482. At the 0.95 quantile, a one unit increase in the value of SR will mean a 0.000482% increase in %IR.

The presence of clay as the backfill material will affect the %IR differently across the percentile of the %IR distribution when compared to the SR

variable. Clay affects the 0.75 quantile the most with the 0.05 quantile the least affected. The value of the maximum estimate coefficient is 57. This is not far off the estimated value at the same quantile for Model 2. The upward trend to the 0.75 quantile reflects the positive effect of clay on the contribution to the %IR reading.

Models of various quantiles are presented in the following equations.

$$\%IR_{0.05} = 30.3 + 0.0000956TCDA - 0.000132\beta + 0.0561\varepsilon - 0.301\eta - 18.9C - 19.1\kappa - 6.25\lambda - 6.48CW + 8.51\mu + 1.22R - 0.0602\sigma + 4.13\varphi - 1.26\chi - 11.3\psi \quad \textbf{5.10}$$

$$\%IR_{0.5} = 86.2 + 0.0000768TCDA - 0.000178\beta - 0.0665\varepsilon - 0.452\eta + 0.785C - 20.3\kappa - 3.21\lambda - 32.1CW - 0.72\mu - 0.279R - 0.243\sigma + 6.63\varphi - 18.9\chi - 11.6\psi \quad \textbf{5.11}$$

$$\%IR_{0.95} = 31.3 + 0.000221TCDA + 0.000482\beta - 0.0939\varepsilon - 0.00579\eta + 18.6C + 10.9\kappa + 25.9\lambda - 8.59CW + 29.7\mu - 0.829R + 1.22\sigma + 11.9\varphi + 4\chi - 16.8\psi \quad \textbf{5.12}$$

Trace plots and the posterior histograms for various quantiles are presented below.

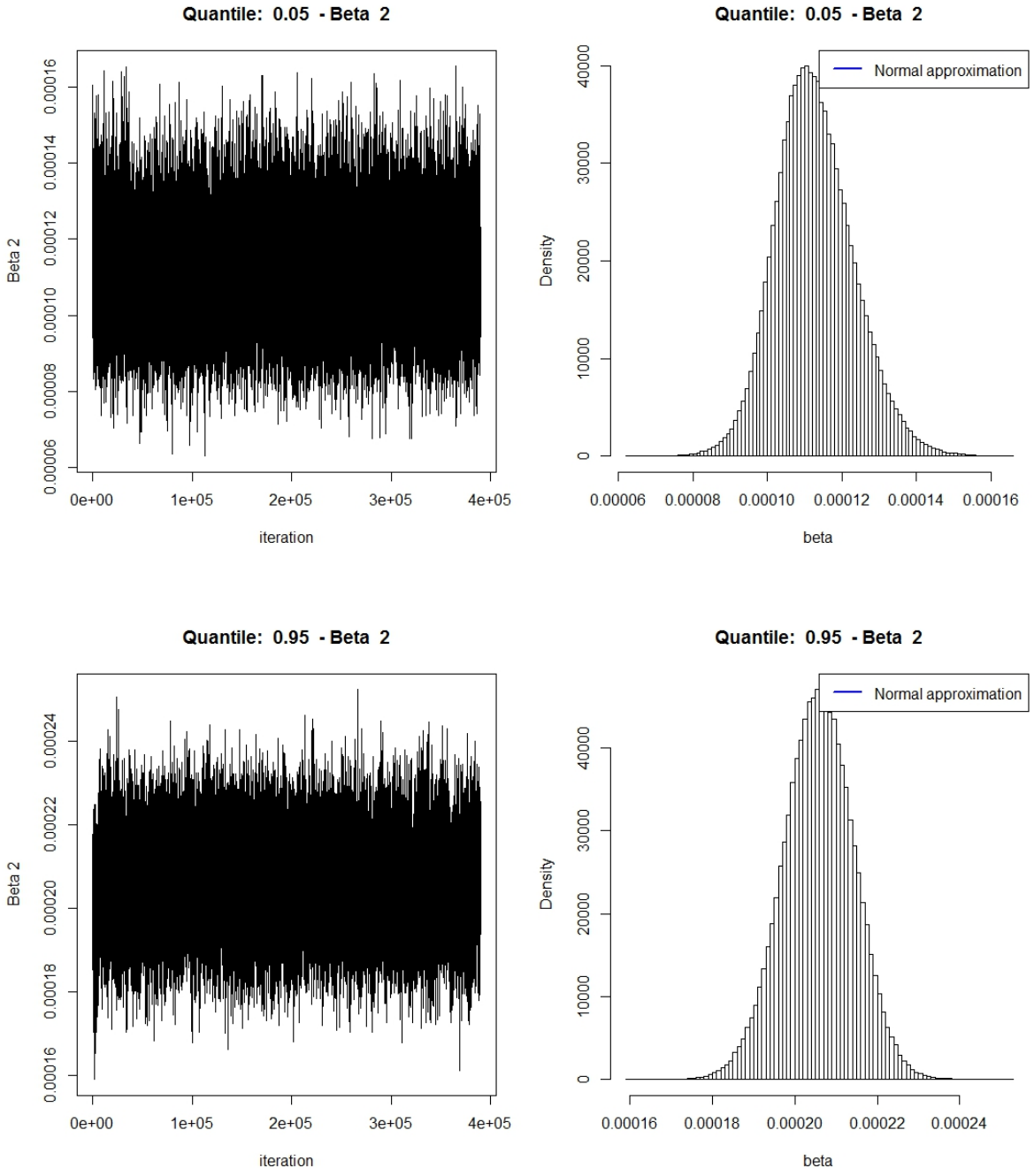


Figure 5-8: Trace Plots and the Posterior histogram of the 0.05 and 0.95 Quantile for the Estimated Coefficient, TCDA for Model 2a

5.2.5 Total Coating Defect Area (TCDA) Models

With the establishment of the %IR model with both the Oriset and the Filtset data (Model 1, 1a, 2, and 2a), the construction of the TCDA model will further increase the capability (by incorporating future inspection results) of operators and decision makers in prioritising coating defects based on severity. To add to this enhancement, the TCDA models (Model 3 and 4) can predict the coating defect area based on predictor variables. The variables chosen in this model are limited to continuous values due to the large amount of subjective interpretation of the categorical variables. Another reason for this is to avoid higher computational cost as Bayesian inference with the Metropolis – Hastings Algorithm (MCMC – MH) is known to take large amounts of computational memory when dealing with large quantities of data. An alternative to this restriction is the usage of the Gibbs sampling method which takes less time and where all proposed samples are accepted without rejection. However, as the posterior form of our parameter of interest is unknown, a general MCMC – MH is required.

As from the previous section, the variable %IR is of prime interest as this is one of the first measurements obtained when conducting a DCVG assessment. By correctly interpreting what the signal means, one is able to make a sound judgment on the state of the coating under inspection. The models developed in this Chapter are aimed at making those interpretations more accurate.

5.2.5.1 TCDA Model 3

The coefficients estimated by the BQR for Model 3 are presented in Table 8-5. The convergence process took 11 million iterations to achieve. Primary interest for the model is the %IR variable, which shows close to zero estimates for the 0.05 and 0.25 quantiles. Beginning at the 0.5 quantile, we can see the trend increasing up to the 0.75 quantile and down again at the 0.95 quantile. The maximum estimated coefficient is at the 0.75 quantile with a value of 849. This means a 1 unit increase in %IR represents an increase of 849 cm^2 in terms of TCDA. Therefore a 100% reading of the %IR translates into 84,900 cm^2. Although this estimation is promising in determining the size of coating defects, the 0.95 quantile illustrates a different picture. The estimated coefficient for this quantile is -93.1. The negative values signify that a one unit increase in %IR equals to a decrease of 93.1 cm^2 in TCDA. Credible intervals also show very narrow predictions indicating greater confidence. Equations below are selected models for the 0.05, 0.5 and 0.95 quantile for Model 3.

$$TCDA\ (cm^2)_{0.05} = -465 - 0.0178\alpha - 0.0034\beta + 6.04\gamma - 0.901\delta - 0.00321\varepsilon + 4.92\zeta + 10.4\eta \quad \textbf{5.13}$$

$$TCDA\ (cm^2)_{0.5} \quad \textbf{5.14}$$

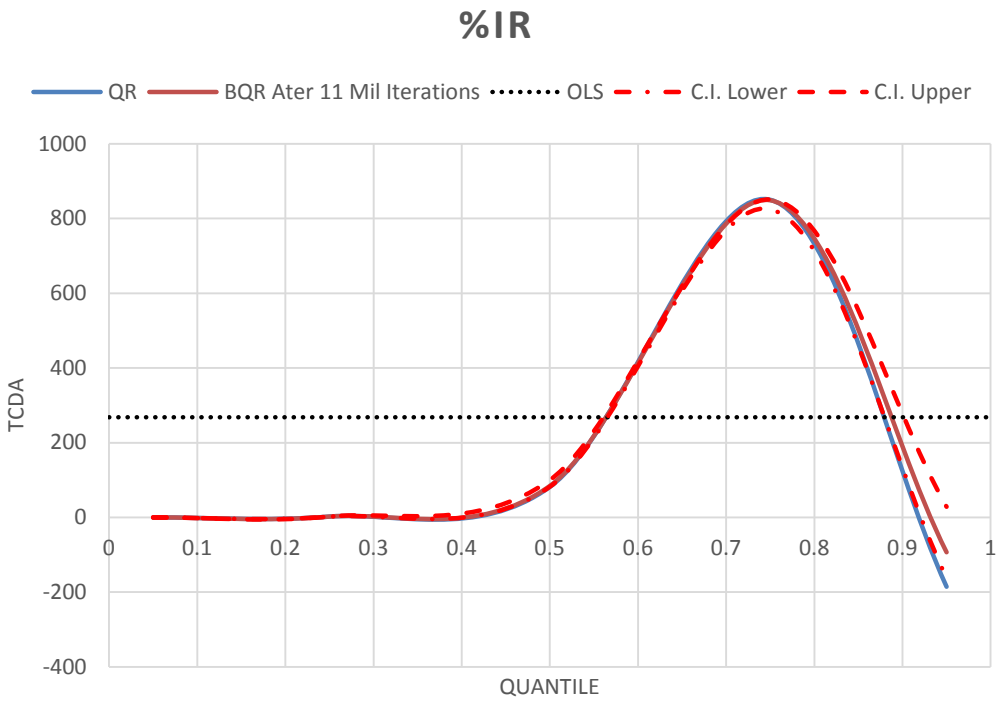
$$= 78687.177 + 84.428\alpha - 0.524\beta + 232.204\gamma$$

$$+ 19.543\delta - 69.776\varepsilon - 2351.485\zeta + 707.098\eta$$

$$TCDA\ (cm^2)_{0.95} \quad \textbf{5.15}$$

$$= 189000 - 93.1\alpha + 0.151\beta + 2740\gamma - 257\delta - 111\varepsilon$$

$$- 8030\zeta + 6040\eta$$



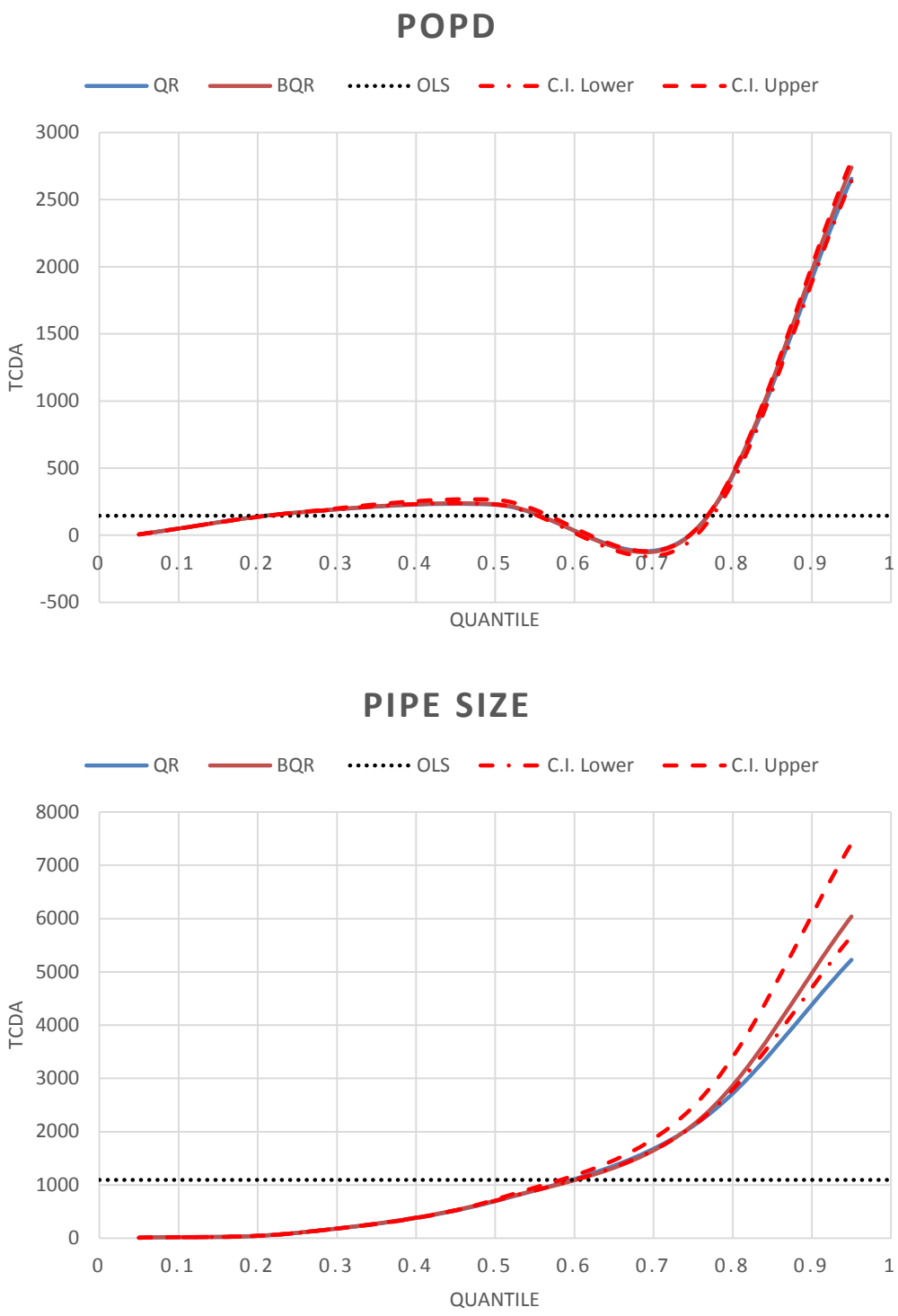


Figure 5-9: Quantile Plots of various Variables of Interest for Model 3

The variable POPD shows useful insights into the correlation between TCDA and corrosion. Based on the trend shown, as the quantile increases, so do

the estimated coefficient values. However there appears to be a sudden dip at quantile 0.75 and a picking up again at quantile 0.95. The maximum value estimated by the BQR is 2,740 which equates to a 1 unit increase in depth of the corrosion pit corresponds to a 2,740 cm^2 in TCDA. This occurs at the 0.95 quantile. Also, the widths of the credible intervals across the quantiles are narrow as compared to previous predicted models.

PS shows positive trends across the whole percentile of the TCDA. The increasing trend starts at the 0.05 quantile through to the 0.95 quantile. The maximum estimated coefficient occurs at the 0.95 quantile with a value of 6,040. This means, for this quantile, a 1 unit increase in PS will translate into 6,040 cm^2 of TCDA.

5.2.5.2 TCDA Model 4

The data considered for this assessment included the removal of 4 data points. The data points removed were in the form of the outliers present in the distribution of TCDA where larger TCDA is associated with lower values of %IR. The selection was done with the judgment of the author. As expected, the %IR variable shows a positive consistent increasing trend across the percentile. This fits well with common understanding and resonates with the whole idea of categorizing defect size based on %IR values. Starting at the 0.05 and 0.25 quantiles, the increase of the estimated coefficients is subtle but for the 0.5 quantile the changes are much more abrupt with the values tapering at the 0.75 and 0.95 quantile. The maximum

value occurs at the 0.95 quantile with an estimated coefficient of 1,481.9. In other words, an increase in 1 unit of %IR will reflect an increase in the TCDA of 1,481.9 cm^2. Therefore, for larger defects (0.95 quantile) a reading of 100% in the %IR value corresponds to a 148,190 cm^2 in TCDA which is the maximum size the model is able to predict. For the lowest percentile, the maximum predicted size is 2.21 cm^2. The maximum predicted defect sizes for all the quantiles are shown in Figure 5-11. Looking at the estimated values at the lower quantiles (0.05 and 0.25), the estimated TCDA size is considered low. This is due to the large amounts of close to zero readings of TCDA for increasing %IR. At the 0.5 quantile and above the distinction is more apparent because close to zero readings of the TCDA are sparse.

The POPD variable represents the amount of corrosion present on the pipelines under consideration. For corrosion to happen, favourable conditions are required to initiate the oxidation process. The pipelines considered in this thesis are all protected with an organic coating and an Impressed Current Cathodic Protection (ICCP) system as backup. Referring to Table 8-6, the estimated coefficients show an increasing trend. From quantile 0.25 up to 0.75 the predicted values show little difference. Abrupt changes can be seen only at the tails of the TCDA distribution, i.e. the 0.05 and the 0.95 quantiles.

The trend of the estimated regression coefficients for the SR variable illustrates the effects of SR on coating defect size. Please refer to Table 8-6 for these estimates. The low values of the initial quantiles signal the

insignificance of the resistance effect of the electrolyte compared to coating defects. Smaller-sized coating defects are not affected by SR. However, at the 0.75 and especially at the 0.95 quantile, SR does play a role in the size of coating defects. The highest estimated value for this variable occurs at the 0.95 quantile where a 1 unit increase in SR will increase the TCDA by 12.2 cm^2.

Increasing trends can also be seen for the variable PS. The estimated values are not much different from the values in Model 3. The maximum predicted value is at the 0.95 quantile with a 1 unit increase in PS translating into 4,610.6 cm^2 of TCDA. At the lowest quantile, a 1 unit increase in PS will expect an increase of 10.3 cm^2 of TCDA. For the median quantile, a 1 unit increase in PS will translate into 963.07 cm^2 of TCDA.

The following equations are the models for predicting TCDA based on various quantiles.

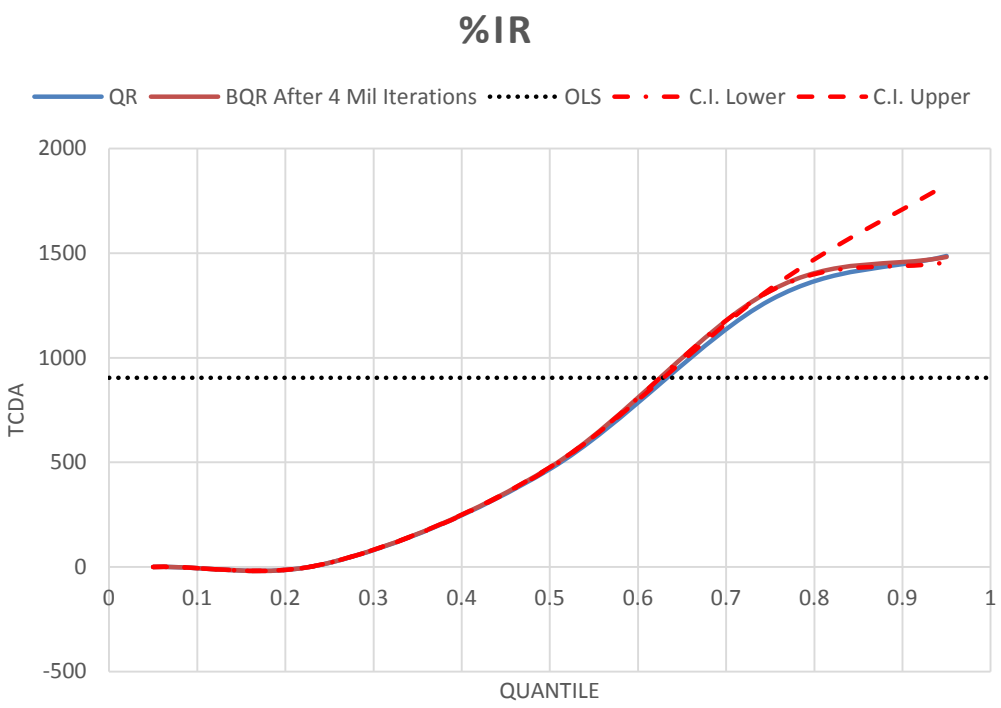
$$TCDA\ (cm^2)_{0.05} = -464 + 0.0221\alpha - 0.00336\beta + 6.06\gamma - 0.916\delta - 0.00871\varepsilon + 4.95\zeta + 10.3\eta \quad \textbf{5.16}$$

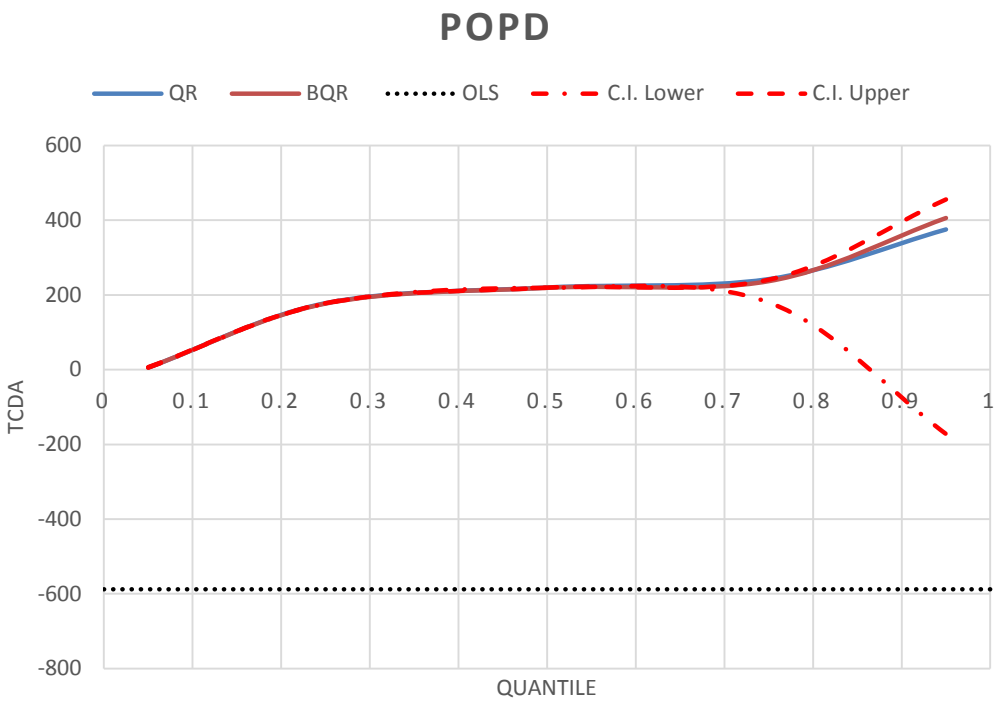
$$TCDA\ (cm^2)_{0.5} = 28978.772 + 475.876\alpha - 0.453\beta + 219.149\gamma + 34.219\delta - 20.033\varepsilon - 1681.945\zeta + 963.069\eta \quad \textbf{5.17}$$

$$TCDA\ (cm^2)_{0.95} \qquad \textbf{5.18}$$

$$= 77655.6 + 1481.9\alpha + 12.2\beta + 406\gamma - 218.8\delta - 97.8\varepsilon$$

$$- 5066.9\zeta + 4610.6\eta$$





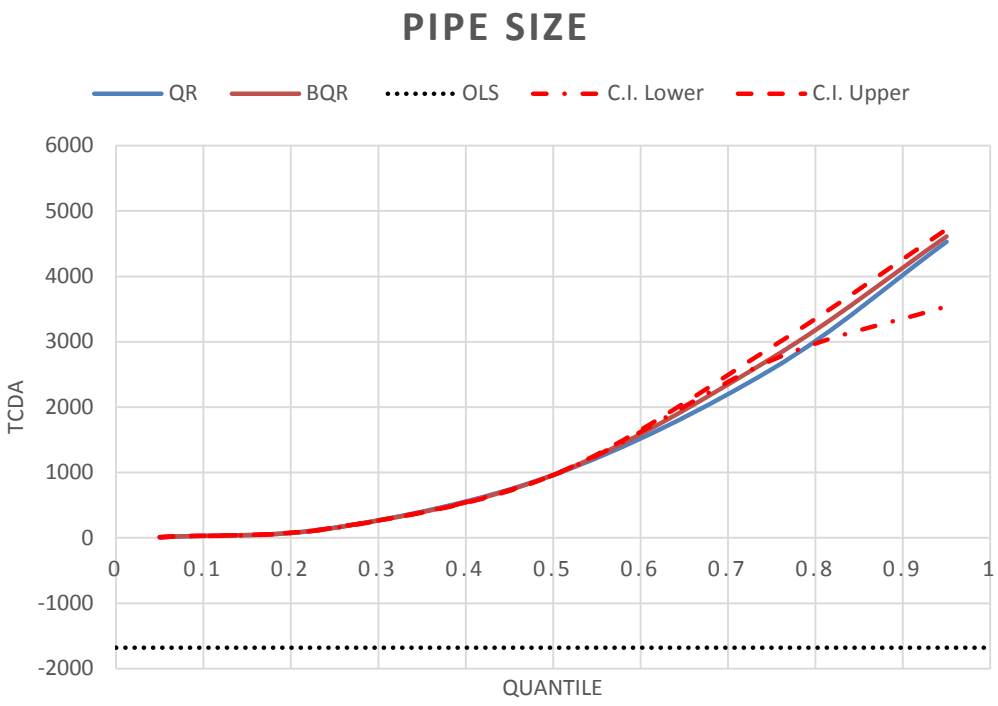


Figure 5-10: Quantile Plots of various Variables of Interest for Model 4

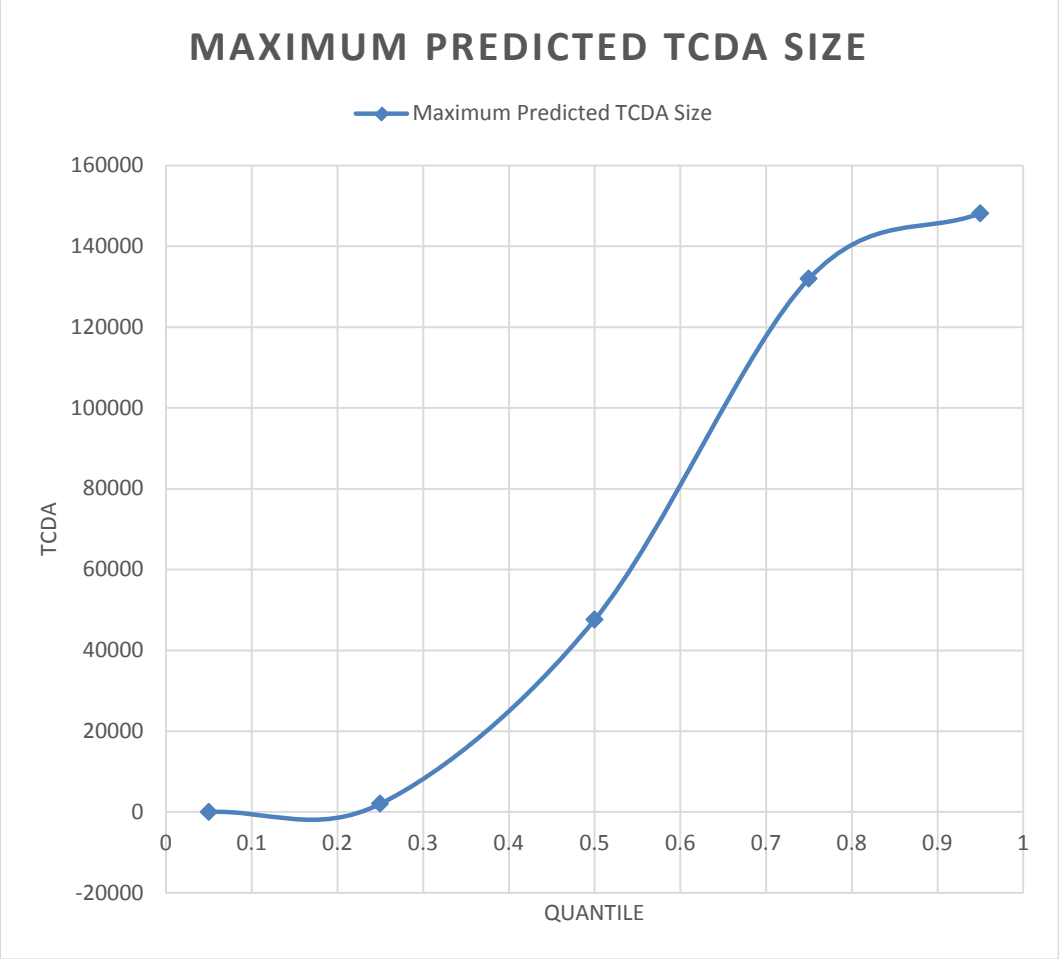


Figure 5-11: Maximum Predicted TCDA Size Based on BQR for Different Quantiles of the TCDA Model

5.3 Discussion

5.3.1 Contributing factors to %IR – (Model 1, 1a, 2 and 2a)

5.3.1.1 TCDA variable

The low coefficient values estimated for the TCDA variable (Model 1, 1a, 2 and 2a) were unexpected since the DCVG technique relies primarily upon coating defects to generate voltage drops. The results show coating defects in general have a mild effect (in terms of coefficient values) on the %IR reading. Other known and unknown factors might also be in play for the contribution to %IR. One of these factors could be SR and the nature of the backfill geometry. This will be discussed further in the following sections. Other factors could include the presence of interference in the form of stray currents especially if the pipeline is situated adjacent to other pipelines or is located near overhanging power cables. Although an interruption technique is used to eliminate foreign currents contributing to %IR indication, large structures such as buried pipelines need longer periods for it to depolarise and be considered IR free [91]. To picture this more clearly, the following figures show the relationship between TCDA and %IR. As was previously mentioned, other factors giving rise to the %IR readings such as the POPD, DUC, DOC, TIS, PS and SR were used to generate the models. These variables take values of 10.5%, 35.2%, 109.5cm, 32.5 years, 35.3 inches and 2,722.1 Ω-cm respectively, representing the mean value of each variable.

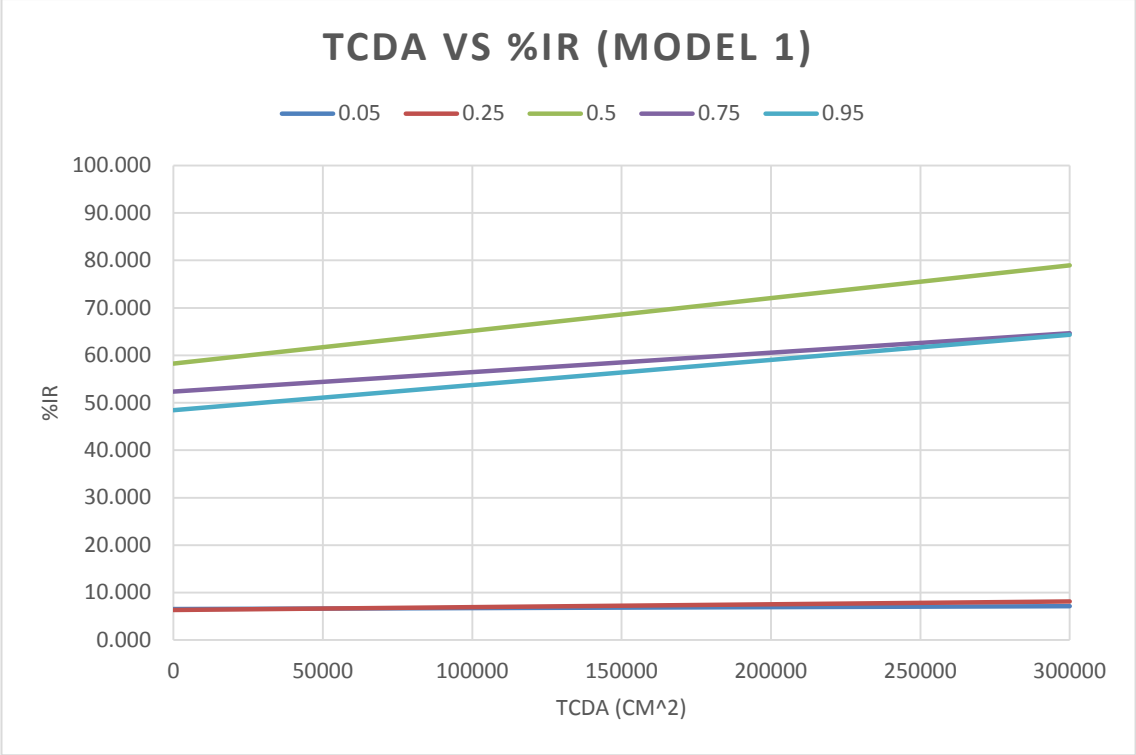


Figure 5-12: TCDA vs %IR for Model 1

Figure 5-12 shows the predictions made by Model 1 of the %IR with increasing TCDA. Generally, the models generated highlight an upward trend which parallels the current understanding of the system. However, the slope of the models indicates a small effect of TCDA on %IR. This can clearly be seen at the lower quantiles (0.05 and 0.25) where the line is almost flat. Also, the median quantile has the highest prediction value and the steepest slope which corresponds to the estimated coefficient values in Figure 5-2. A refined version of Model 1 is given by Model 1a presented below. Similarly, the models take on the mean values of each contributing variable.

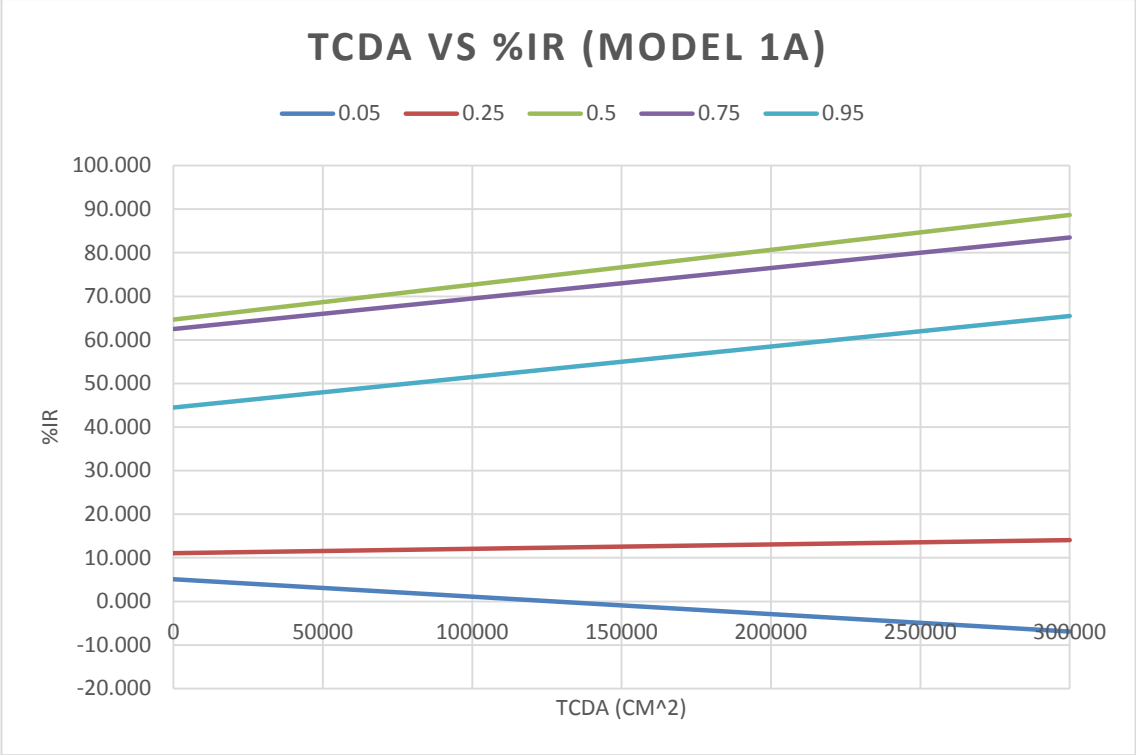


Figure 5-13: TCDA vs %IR for Model 1a

The prediction of the resulting %IR in Figure 5-13 shows an improvement in terms of the effect of TCDA on %IR with steeper slopes being observed. Similar to Model 1, the median of %IR shows the largest effect from the TCDA. The estimated %IR values based on the median are also higher with Model 1a as compared to Model 1. The removal of certain variables which do not contribute to the %IR has improved the %IR estimation for the top three quantiles. For the 0.25 quantile, small effects of the TCDA on %IR is seen which is similar to Model 1. However, the estimated values here are higher. The 0.05 quantile shows a decreasing trend where increasing TCDA relates to a decreasing %IR.

The inconsistency (higher TCDA does not reflect a higher %IR values) for Models 1 and 1a with respect to the 0.05 quantile could possibly be attributed to the outliers present at higher and lower quantiles of the TCDA distribution – large defect areas are paired with low readings (indication) of the %IR and *vice versa*. Additionally, credible intervals at higher and lower quantiles for Models 1 and 1a are much wider, indicating higher uncertainty compared to the median quantile where the maximum estimated value have occurred. Inconsistent results are also being summarised in the following bullet points.

- Interference in the form of stray or telluric currents will interfere with the voltage drop picked up by the DCVG instrument. Some section of the pipes under assessment is located within a network of pipelines with each of these pipelines having its own cathodic protection system. Currents from an adjacent ICCP system, electrified railway tracks (DC traction system), overhead power cables etc. have the potential to compromise the %IR signal. A more severe effect of this interference is corrosion.
- Power cables or transmission lines have the ability to compromise the DCVG signal in the form of AC currents. AC currents can also lead to accelerated corrosion of the pipelines running below [92]. In the case of MEOC pipelines, power cables can be seen running closely along and perpendicular to the direction of the buried pipelines which can be the cause of irregularities in the %IR readings.
- The heterogeneous nature of soils compromise or alter the measured voltage signals. The calculation of the %IR value requires input in the

form of the pipeline-to-electrolyte interface resistance. The resistance value is related to the SR value measured at test posts. However, DCVG readings are conducted away from test posts where the magnitude of SR is different. The differences will contribute to the inconsistencies of the %IR measurements where the heterogeneity of the soil is not considered in the %IR formula. Although, SR measurements were taken for every excavated area, this was not included into the %IR calculation.

- Defects occurring at the 6 o' clock position will tend to attenuate the voltage signal which will not correspond to the true size of a defect [63].
- Based on the report provided by TWI Ltd., there is a possibility that some of the coating defects were caused by the excavator during excavation of bell holes for the direct examination process. These defects were not present during the indirect assessment (DCVG measurements).
- Deposits of scale due to the cathodic protection current on the metal substrate will mask the true size of a coating defect. Measurements are perceived to be small based on the %IR reading. This is an erroneous representation of the true size of the defect.

The assessment on Models 2 and 2a which utilises the Filtset data, considers rock as the backfill material with an angular geometry and polyethylene as the type of coating as the reference variable. Due to the inconsistencies found in the Oriset data, 4 outlier points were taken out

based on the author's expert opinion, mainly in the region of low %IR readings. The estimated %IR readings based on Models 2 and 2a are given as follows. Similar to the previous Models 1 and 1a assessments, the mean values of POPD, DUC, DOC, TIS, PS and SR were used to generate these models.

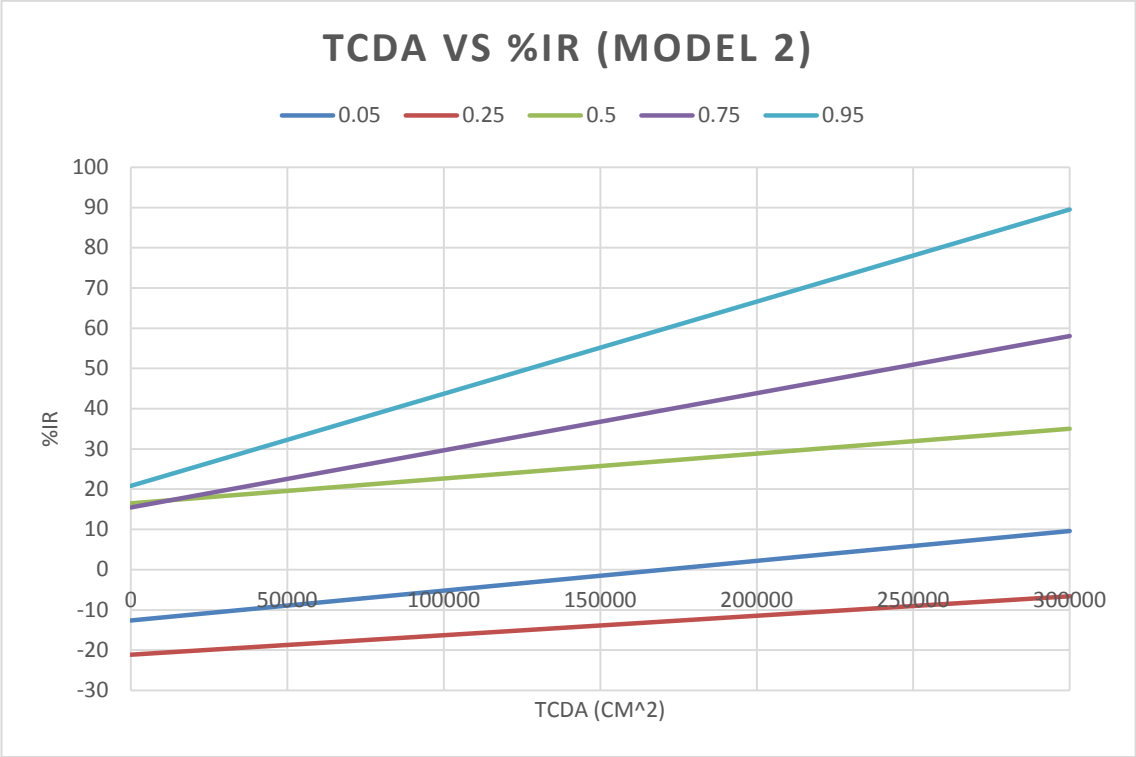


Figure 5-14: TCDA vs %IR for Model 2

Immediately, it can be seen that the estimated values of the %IR is much improved over Models 1 and 1a. The effect of TCDA on %IR is also greater which reflects the underlying intention of a DCVG assessment. The highest predicted value of the %IR is at the 0.95 quantile which indicates TCDA has the highest effect on higher readings of the %IR. Additionally, narrower credible intervals were obtained highlighting lesser uncertainty of the

estimated coefficients. Therefore, the removal of 4 excavation points by expert opinion improves the overall estimation of the role of TCDA on %IR. However, the 0.05 and 0.25 quantiles show an apparent effect of TCDA on %IR but the resulting estimates are below the zero line. For the 0.25 quantile, all the predicted readings of %IR are negative and sit lower than in the 0.05 quantile. Although the apparent outliers were removed for this assessment, there are other factors such as non-contributing variables that might have an overall effect on the %IR predictions. Model 2a tries to answer this by further refining the model through the omission of variables which in theory should not contribute to the generation of %IR. Model 2a prediction of %IR is given as follows.

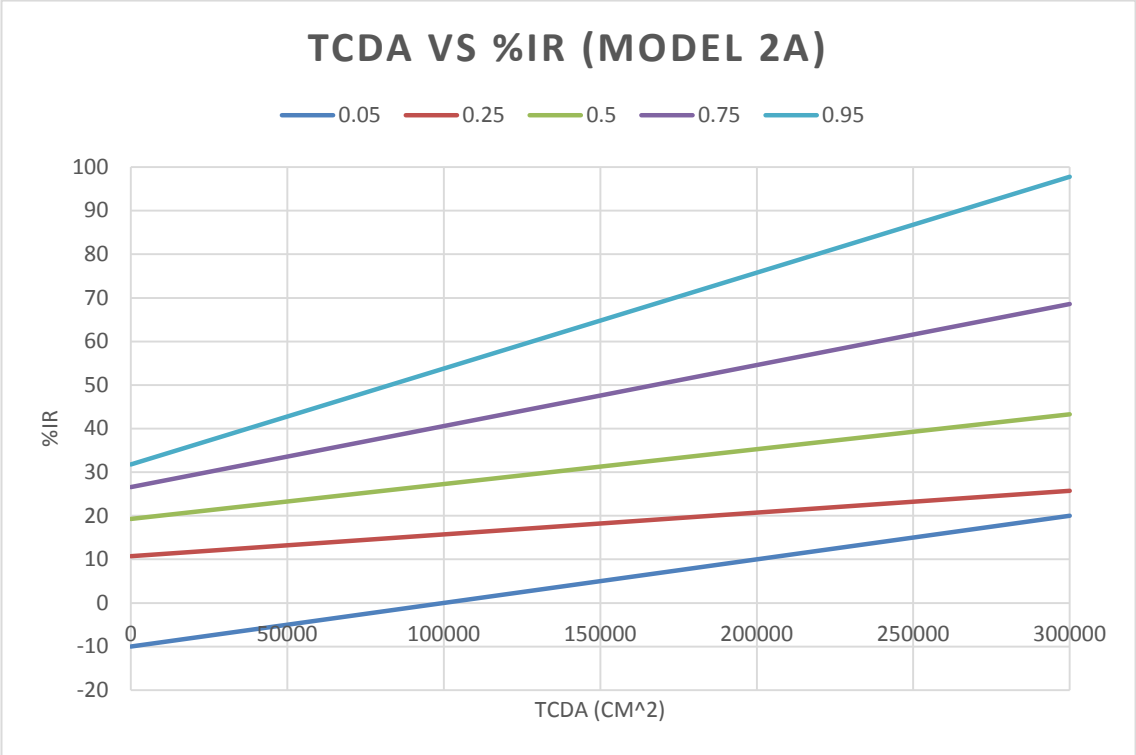


Figure 5-15: TCDA vs %IR for Model 2a

From Figure 5-15 the omission of certain variables has improved the overall prediction of the %IR based on TCDA. Significant effects of the TCDA on %IR are seen across all the quantiles. The effect of higher TCDA on higher readings of %IR is seen with the highest predicted values of %IR occurring at the 0.95 quantile. This can also be said of other quantiles where lower values of TCDA affect the lowest part of the %IR readings. However, the 0.05 quantile shows predicted values of %IR of less than zero. This small inconvenience can be stipulated as the cause of the linear approach taken by the author when modelling the relationship. To address this, Chapter 6 models the relationship between TCDA and %IR (based on similar variables but with the exclusion of categorical variables) using Logistic Quantile Regression which should constrain the resulting output within a specified range. Overall, Model 2a is an acceptable model in the prediction of %IR (based on established literature on the DCVG technique) with the added bonus of simplicity and brevity due to its utilisation of fewer variables.

5.3.1.2 SR and Backfill Type Variable

The SR estimated coefficients for Models 1 and 1a show a decreasing trend with lowest value occurring at the 0.5 quantile. However, the estimated effects of the rock variable on the contribution of %IR indicate a reverse trend with the maximum estimated coefficients occurring also within the region of 0.25 to 0.5 quantiles. Since these two variables are somewhat related, the opposite predictions seem to complement each other and highlights the heterogeneous nature of soil. Highly resistive electrolyte which

contains materials such as rocks will produce large voltage drops as current passes through it. These voltage drops will likely be picked up by the DCVG instrument indicating a defect more severe than it actually is. This is confirmed by the work of J.P.Mckinney [64] in his thesis which states that prioritization of DCVG indication will be more accurate if SR is taken into account. The higher quantiles highlight a relatively weak effect of the rock variable to %IR. However, this can be understood by also observing the value estimated for the general SR variable which highlights a stronger effect.

In Models 2 and 2a, the reference variable for the models was changed and the variable Backfill Type – Clay, shows an increasing trend until it reduces at the 0.95 quantile. Clay is considered as soil with a high degree of compaction thus possessing low resistance to current flow. The low resistance should not produce large voltage drops and hence one would not expect the rising trend in the estimated coefficients. However, if we were to look at the Backfill Geometry – Round variable, the estimates are much more aligned with common understanding. The presence of rounded soil grains creates an environment which is less resistant to electrical currents (similar to clay). Across the percentile, the estimated coefficient values show a downward trend with a slight increase at the highest quantile. This is the inverse of the clay variable's trend. Similar to Models 1 and 1a, the two variables seem to complement each other and can only be understood when both are looked at together. The decrease in the estimated value at the 0.95 quantile for the clay variable and the increase of the predicted value at the

0.95 quantile for the Backfill Geometry – Round variable is the cause of a possible mixture of fine to coarse grain soils in the backfill. Moreover, there is also the possibility of foreign currents interfering with the measured signal as mentioned above. Coupled with the heterogeneous nature of soils, unexpected outcomes like this are not unusual to find.

5.3.1.3 Depth of Cover Variable

A paper by Moghissi and the work from J.P Mckinney [64], [65], have argued that depth of cover plays an important role in the detection of potential gradients generated by the cathodic current. Equipotential lines are generated from coating defects which depends on the SR and also the level of the protective current provided to the pipeline. The size of these potential lines is largely based on the size of defects and the level of current supplied. From this it is obvious that the depth of the buried pipeline is a factor to consider in the detection of potential gradient signal. If the pipe is buried too deep and small defects are generating voltage drops, chances are that only the outer part of the spherical gradients will be detected or will be missed entirely. Potential gradients on the outermost surface of a voltage sphere represents only a small part of the whole voltage spectrum. The trend of the estimated coefficients for the reduced version of the %IR model (Model 2a) resonates with this theory. Apart from the 0.05 quantile, the downward trend continues until the 0.95 quantile. What this means is that as the depth of the buried pipeline increases the amount of potential gradient signal being picked up will decrease. A statement posed by NACE in their standard,

TM0109-2009 [93] and by Ukiwe et.al. [94] states that “the standard’s guideline on the usage of indirect inspection tools may be less sensitive to pipes buried at excessive depths”, echoing the results obtained in this research.

5.3.2 TCDA Model – (Model 3 and 4)

5.3.2.1 %IR Variable

The estimated coefficients for Model 3 have shown that the trend does not sit well with current industry understanding of DCVG. The trend can be attributed to a number of outlying data points present at higher quantiles of the TCDA distribution. A better way of visualising this is by plotting the predicted TCDA based on increasing %IR using Model 3 with respect to the different quantiles. Other variables in the model were kept constant where the mean of POPD, DUC, DOC, TIS, PS and SR similar to previous assessments in this Chapter were used as the contributing factors. Figure 5-16 shows this plot.

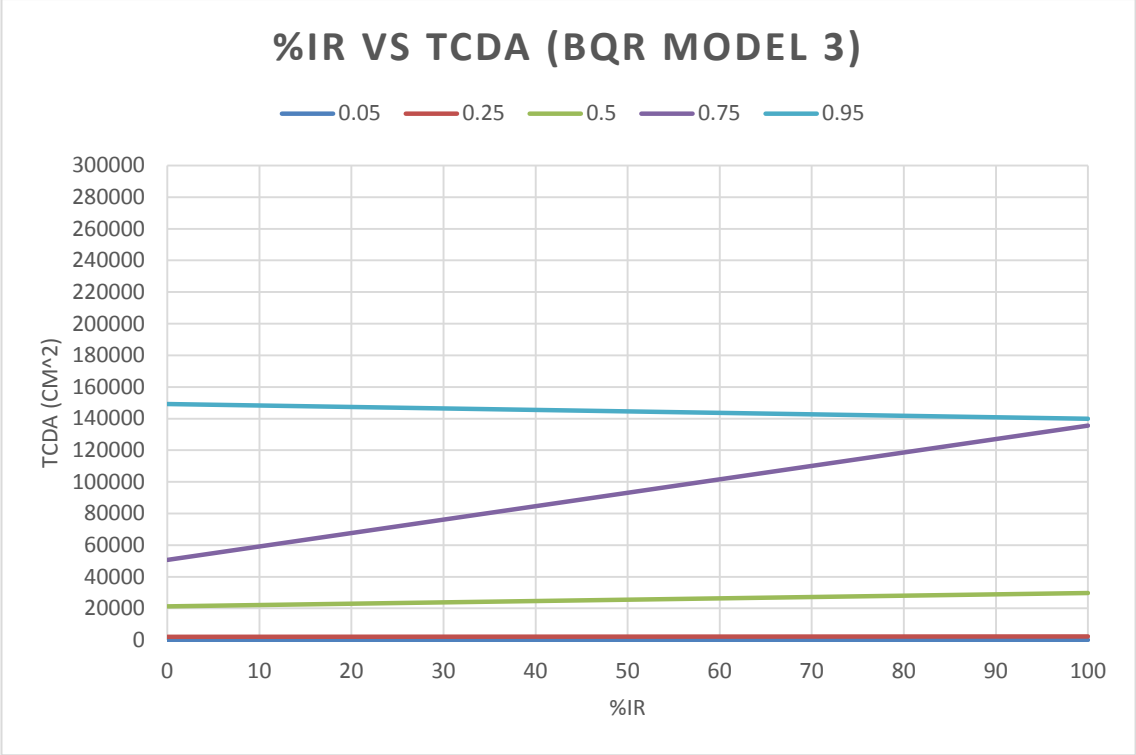


Figure 5-16: %IR vs TCDA for Model 3

Figure 5-16 shows the linear effect of %IR on the resulting TCDA estimation. The effect is small across the quantiles except at the 0.75 quantile. At the lowest quantiles (0.05 and 0.25) the effect is almost zero which is represented by the flat line. The trend of Figure 5-16 is not surprising if one looks at the Oriset data where many of the readings are inconsistent with common industry DCVG theories. Small indications in the form of %IR measurements have been paired with very large coating defects and large readings of %IR yielded very small coating defects. The same scenario is encountered during the construction of Models 1 and 1a. These irregularities can also be explained by the bullet points given in section 5.3.1.1 above.

Of the possible reasons given in the bullet points, the most probable explanation for this phenomenon is due to the disturbance coming from stray and telluric currents. Most of the pipes under assessment were situated within a network of pipelines which runs in parallel and perpendicular with the one under investigation. Currents from adjacent CP systems which are protecting other pipelines have the potential of leaving their intended path and being picked up by the DCVG instrument. This can produce incorrect readings for the DCVG assessment being based on the stray current and not the actual defect themselves. Kutz [91] has explained this problem in greater detail.

Another interesting finding was that the pipes were originally protected by a sacrificial anode system. The anodes were attached to the pipe via tack welds. Based on the pre-assessments photographs, tack welds were still visible and not insulated. Since these tack welds and their connecting rods are exposed to the environment, they provide an exit point for currents to leave the surface of the pipeline. The exiting currents can also interfere with the voltage gradient generated by the coating defects which in turns produces misleading information for the interpretation of %IR. Apart from disturbing the potential gradient signal, the exposed tack welds and associated rods could also lead to accelerated corrosion. However, corrosion was not observed at these points.

The relationship of %IR and TCDA based on Model 4 is illustrated in Figure 5-17. Similar to previous assessments, other contributing variables such as POPD, DUC, DOC, TIS, PS and SR were kept constant.

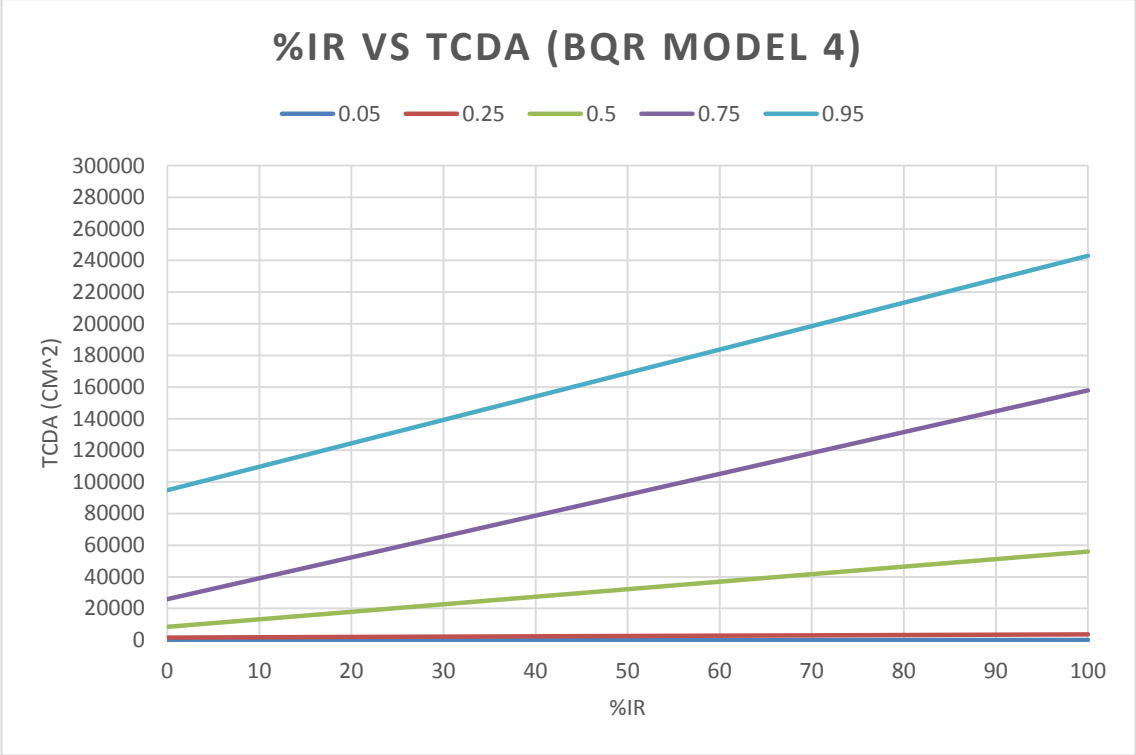


Figure 5-17: %IR vs TCDA for Model 4

The trend in Figure 5-17 illustrates the general industrial understanding of the relationship between %IR and TCDA. As the quantiles increases, so does the effect of %IR on TCDA which leads to the conclusion of higher %IR affecting larger coating defect areas in a positive way. It can also be said that the sensitivity of the DCVG technique relies on the size of the coating defect. Medium to large defects give a reasonable approximation of the defect size. However, the interpretation based on the %IR on smaller defects should be treated with caution due to multiple zero readings present at lower quantiles.

As mentioned earlier, outliers were omitted based on the author's expert opinion. Due to this, Model 4 does not suffer from the problems faced by Model 3, Model 1 and Model 1a where outliers play a role in the estimation of coefficients. As such, the models are more general and are sufficient in the case of subsequent inspection of the MEOC pipelines. Applying these models to other pipelines is not advisable since the filtered data used are not free from other forms of "noise", e.g. lower TCDAs appearing on larger readings of %IR (depicted by the 0.05 and the 0.25 quantile). More data is needed if the developed models are to be utilised for the general use by the pipeline industry. As in the case of subsequent iteration of the ECDA for the MEOC pipelines where prediction of TCDA with regards to the %IR readings is required, the estimated coefficients and their associated credible intervals found here can be used as the prior distribution for the estimation of coefficients for subsequent inspections. It is expected that this will produce better posterior densities with narrower credible intervals indicating less uncertainty in future estimated coefficients.

5.3.2.2 POPD Variable

Findings from Model 3 indicate that at large coating defect area the possibility of finding deeper corrosion pits is more likely. With larger TCDA, the amount of current provided by the cathodic protection system also should be large. When the level of protection current is inadequate or obstruction of the current's path in the form of a shielding electrolyte is present, corrosion activity is to be expected. However, a dip at quantile 0.75 on pipelines with

medium to large TCDA corresponds to corrosion pits with shallower depths than corrosion pits at the largest and the smallest of TCDAs, which goes against the normal assumption that a pit's depth is directly proportional to the size of TCDA. Model 4 does not exhibit such issues. At the same quantile, the coefficient predicted shows a smooth increase from the median quantile to the largest quantile. Moreover, for Model 4, a consistent upward trend can be seen across the TCDA quantiles. Between the 0.4 and the 0.6 quantile the estimates plateau suggesting that, for these defect sizes, the effect of an increasing POPD is minimal. The increase in values from the 0.05 quantile to the 0.25 quantile can be judged an initial step in the corrosion process which also contributes to coating defects. At this step, corrosion is initiated, and coating defects grow in tandem. The aforementioned plateau is an indication of the pit growth rate being faster than the growth of TCDA. This will produce deeper pits at smaller TCDA which solidifies the finding in Chapter 4. Deeper pits at smaller coating defect should be treated with caution as defects of such characteristics will normally go unnoticed with the consequences of failure being severe. The same effect can also be seen between the 0.75 and 0.95 quantiles. However, here the credible interval increases in width indicating a less certain prediction. The OLS prediction is also located in the negative region which means that all the above observations would be missed with the averaging approach.

5.3.2.3 SR Variable

Model 4's predicted coefficient quantile trend can be interpreted as highly resistive soil having a large effect on the size of coating defects. Coarse grained soil is known to be highly resistant to electrical current flow hence soils such as sand, silt or even rock pose high units of SR. These types of soil with angular particle geometry have the possibility of damaging pipe coatings through the process of abrasion. Pipe or soil movement also has the possibility of creating abrasion between the coating interface and the electrolyte. Another factor to consider is the stresses created by the self-weight of the backfill [38], [95]. The backfill weight applies stresses on the pipe's coating creating a wrinkling effect normally found at the 8 and 4 o'clock positions of the pipe. The wrinkling of the coating combined with the abrasion effects of the angular particle size (high SR) will sometimes result in coating defects.

5.3.2.4 Pipe Size Variable

Both Models 3 and 4 show similar upward increasing trends for the estimated coefficients across the quantiles. The larger pipes will likely to exhibit a larger TCDA. This is because as pipes increase in size, so does the surface area. Therefore, the possibility of having larger TCDA on larger pipes is high especially if the situation is not favourable (conducive environment for coating defects to occur). The estimated values of the two models are different but not substantial. Differing coefficient estimates are due to the previously mentioned effect (outliers were taken out). In Model 3, the predicted OLS would have missed the positive and negative effects of

increasing PS on TCDA while for Model 4, the OLS would have completely missed the estimates since it is calculated to be negative. The credible interval estimates for both models are very narrow with wider intervals occurring at the 0.95 quantile. However, this interpretation has to be treated with caution as the pipe sizes considered are within the range of 26 inch to 42 inches only – which contributes to the narrow credible intervals. To extrapolate the interpretation to other pipe sizes, a larger, more comprehensive dataset is needed to accommodate more variation in pipe sizes.

5.4 Bayesian Quantile Regression

Assessments were done on the MEOC data to obtain the predicted coefficient value by means of both a Bayesian approach and the classical approach. The coefficient estimates illustrated by both the Bayesian and classical methods are somewhat similar. Both approaches consider parameter uncertainty with the Bayesian approach being more reliable as it does not rely on asymptotic approximation of the variances. Classical approaches such as bootstrapping in the construction of confidence intervals use estimation of the asymptotic variances and depend on the model error density which is difficult to estimate reliably. Hence, the coverage probabilities of the true parameter of these methods are at best sufficient but not necessarily 100% reliable. This is supported by a paper from the Plant Physiology field by Devore [96] which shows the classical approach estimated a lower probability of the confidence interval as compared to the

Bayesian approach. This seems to suggest that a Bayesian method is better in terms of coverage and thus includes all parameter uncertainty. Other advantages of the Bayesian method are that it provides a simple explanation based on the credible interval. For this Chapter, the credible intervals are set to be 95% and thus the true value of the coefficients can be explained as "having a probability of 0.95 of falling within the credible intervals".

The prior used for all the estimation of the regression coefficients is a non-informative prior. Yu and Moyeed [79], have proved in their research that the use of an improper prior will lead to a proper posterior through the Bayesian Quantile process. This is especially useful in the case of the MEOC data. The research into estimating coating defect size by statistical means is new and therefore lacks expert judgment (belief) on what a prior distribution should be. In the case of choosing an incorrect prior, there is the possibility of having the resulting posterior heavily influenced by the prior (especially when data samples are small – as in the case of the MEOC data). This is often misleading and illustrates a different picture on the true parameter estimates. Freedom from this pitfall is advantageous in the context of this thesis. Additionally, the BQR method uses the ALD as the likelihood function. Since the likelihood function (ALD) disregards the original distribution of the data, specifying a specific distribution is not needed. The paper [79] goes on to say that the use of the ALD is a "very natural and effective way for modelling Bayesian quantile regression". After the Bayesian process, the resulting posterior statistics such as the mean estimates of the quantiles and the calculated credible intervals can be used as the new prior distribution

when new data are made available (further iteration of the ECDA process). The construction of the prior distribution will have to be done by eliciting the estimated statistics obtained within this thesis. An appropriate distribution will have to be chosen which will incorporate the Bayesian estimates i.e. the coefficient values and it's corresponding credible intervals. This process is often referred to as Bayesian updating. Bayesian updating will better predict the new regression coefficients with better reliability and reduction in uncertainty as more and more iterations of the Bayesian process are made. The Bayes estimates of the coefficients along with their corresponding credible interval are shown in the Appendix section, Table 8-7 to Table 8-12.

One of the biggest hurdles of the Bayesian method is the problem of convergence. As was seen in the results of Model 4, up to 11 million iterations were needed to achieve convergence. This is due to the algorithm (Metropolis-Hastings) used for the acceptance and rejection of samples. This will make the method computationally expensive which can be a hindrance to future researchers employing the technique.

Chapter 6

Logistic Regression and Logistic Quantile Regression for Analyses of DCVG Indications and Corrosion Depth

6.0 Introduction

The previous chapters have highlighted methods in approximating the size of the TCDA given its contributing factors. By doing so, the decision-making process of where to excavate for further assessments (direct assessments) can be made with greater accuracy. This will also be useful for pipeline operators managing the pipeline integrity while keeping costs low.

This Chapter however will shift its focus to the variable %IR (also looked at in Chapter 5) and POPD as the dependent variable. The two variables are important since examining both of them is crucial for the understanding of the DCVG technique and for maintaining the pipeline integrity.

As was also stated in previous chapters, the United States [10] has imposed a regulation on operators and managers to produce criteria for the identification and documentation of indications from an indirect assessment technique which will be considered for further assessment (direct assessment through excavation). The criteria also serve to define the urgency of the subsequent inspection based on documented and identified indications. Responding to this regulation, pipeline operators have come up with simple rules (criteria) for managing the integrity of pipelines. These rules cover all of the pipelines under their watch, irrespective of the pipe's specification and environment. The rule looks at the Close Interval Potential Survey (CIPS) data to identify defects. Moreover, the identification process is highly subjective since it requires the input of an expert to judge the location of defects and where to excavate. According to [65], this is necessary since the environment of each location constantly changes. However, the combination of this general rule with the input of expert opinion will yield a decision which is highly conservative [65].

In the context of the MEOC ECDA data, TWI Ltd., also provided a method in classifying defects based on their DCVG assessments. In accordance with [97], the pipeline was segmented into areas where the risk of corrosion is

expected. The risk profile of each segment was identified subjectively by observing the CIPS data and the environment in which the pipe resides. This is similar in method to that highlighted above by [65]. The DCVG data were later combined with this risk profile to identify potential corrosion hotspots for decisions on where most effort should be directed for further assessment. Unfortunately, according to the MEOC data which follows the steps above, a significant amount of excavations resulted in no defects albeit the %IR readings were relatively high. The same can be said about the low readings of %IR which yields a larger coating defect area.

Thus far, we have conducted regression analyses to determine the relationship of TCDA and %IR to their respective contributing factors. However, these analyses considered the dependent variable to be unbounded. The dependent variable is treated as having a non-constraining value where any value calculated from the regression process is possible. For bounded dependent variables, the values are restricted to some interval. The best-known outcome is binary where the intervals are 0 and 1. Other examples of bounded dependent variable are visual analogue scales between 0 and 10 cm, school grades of between 0 and a 100 and quality of life index between 1 and 100 [98]. Bounded dependent variables typically produce frequency distributions of shapes unimodal, J shaped and U shaped [80]. It has been shown that using traditional methods such as least squares regression on bounded dependent variable is inadequate [99].

Logistic Regression provides a solution for bounded dependent variables as it limits the interpretation of the dependent variable within a prespecified range. In our pursuit of predicting the size of coating defects through the interpretation of %IR and corrosion through modelling of variables, the logistic regression provides a useful tool for assessment since some of the variables under investigation possess a pre-specified range. This includes %IR, POPD and DUC, all ranging between 0 and 100%.

Quantile regression (QR) is a technique in modelling conditional quantiles of the dependent variable. Its main advantage is that it gives a fuller picture of the dependent variable for every value of the independent variable (see comprehensive description is given in Chapter 3). This is especially useful when applied to skewed or long-tailed distributions. QR makes it possible to extract information from the extreme ends of the dependent variable's distribution which is useful when studying the failure of pipelines (normally failure occur at the tail of a distribution). Also, when compared to traditional least squares regression, QR makes no assumptions on the distribution of the error term, is robust to outliers and is not affected by monotone transformation of the outcome variable in terms of its inferences. It is thus understood that the technique is becoming ever more popular in the field of economics, ecology, meteorology and biomedical sciences. Bottai et.al [80] has investigated on the use of quantile regression with bounded dependent variable by employing logistic transformation to the quantiles. With this combination, the analysis benefits from the effectiveness of the QR while limiting it to a certain range of the dependent variable. For this chapter, the

logistic and the logistic quantile regression are used for the analyses of variables found in the indirect and direct assessment of the MEOC data.

6.1 Middle Eastern Oil Company (MEOC) Data

Data are again taken from the MEOC ECDA project conducted by TWI Ltd. Extensive description on the data can be found in Chapter 2. In summary, the Middle Eastern Oil Company (MEOC) has appointed TWI Ltd., as their contractor to conduct integrity assessments on nine of their buried pipelines. Upon studying these pipes, it was concluded that the pipes are unpiggable. Therefore, an External Corrosion Direct Assessment or ECDA for short, based on the NACE SP0502-2010 [97] was identified as the best method to assess the integrity of the pipes. An ECDA comprises 4 major steps which include pre-assessment, indirect examination, direct examination and the post assessment. Details of all these steps and the data obtained from these assessments are discussed in detail in Chapter 2.

For the purpose of formulating a tool for pipeline operators and integrity personnel to make sound and informed decision on where to excavate, data considered for analyses are taken only from the indirect and direct assessment steps of an ECDA. One can also include the pre-assessment step as part of the analyses. However, as was stated earlier, the data from such steps involves highly subjective inputs (which can also be erroneous) from the so-called experts in the field. This is also one of the reasons for not explicitly stating in Chapter 5 an informative prior as a wrongly specified prior

will lead to erroneous results (inference). Data collected from the indirect and direct assessment steps are tabulated below.

Symbol	Variables Considered	Type of Variable	Bounded / Binary
α	IR Drop (%IR)	Continuous	Bounded
$TCDA$	Total Coating Defect Area	Continuous	Unbounded
β	Soil Resistivity (SR)	Continuous	Unbounded
γ	Percentage of Pit Depth to Wall Thickness (POPD)	Continuous	Bounded
δ	Deposits under Coatings (DUC)	Continuous	Bounded
ε	Depth of Cover (DOC)	Continuous	Unbounded
TEx	Excavation (1=Yes, 0=No)	Categorical	Binary
$TExx$	Excavation (1=Yes, 0=No) (Relaxed)	Categorical	Binary

Table 6-1: Lists of the Variables obtained from the Indirect and Direct Assessment Used for the LQR Assessment

In applying the logistic and the LQR to predict %IR and POPD, only the continuous variables are taken into consideration. This is done to keep subjectivity to a minimum within the assessment and to simplify the problem.

One of the advantages of logistic regression is the ability to transform the dependent variable, which is bounded, by a specific interval and infer it as a probability statement. This property can be exploited in the case of the MEOC data where the probability of excavation can be calculated and be of use for future ECDA projects. A new categorical variable, Excavation (TEx)

is introduced for the assessment of the probability of excavation. TEx is derived from the analyses of the direct assessment stage. A value of 1 is given where excavation should have occurred based on large coating defects and deep corrosion pits and 0 otherwise. The excavation sites are divided into two. An analysis was done to determine excavation sites which are justified, and the ones which are not. This can be thought of as a hit and miss approach. The condition for justification of excavation comes from looking at the %IR readings, the size of coating defect and the depth of corrosion pits. By combining these factors, a comprehensive judgement is achieved by considering the most detrimental factors to a pipeline's integrity. The limits of these factors as to render the excavation justifiable is set to values which consider engineering standards such as the NACE SP0502:2010 and API 579 [97], [100]. The factor TCDA is a problem since there are no definitive critical sizes for when the coating should be replaced (and hence excavate). Different companies and operators have different criteria for the critical size of TCDA before replacement. Hence, the author chose an arbitrary value as a limiting factor. Although this seems to be subjective, the determination of the value is necessary for the logistical analysis. This value is changeable in the future where more or less conservatism is needed for the assessment.

6.2 Probability of Excavation

The probability of excavation based on the TEx variable against other variables are conducted using logistic regression. The TEx variable's limit is

set at 60% for the %IR since the standard NACE SP0502:2010 [97] specified that for every indication of 60% and above, also called category 4, should be treated with high priority. A detailed description of this definition is given in Chapter 2.

For the POPD variable, the threshold for when repairs are required is determined to be 50%. The minimum required thickness, t_{min}, represents the required thickness of the pipe for safe continual operation. Going beyond t_{min} doesn't mean the pipe will fail but additional integrity assessments are needed in ensuring continual safe operation. There is a variety of formulas used in calculating t_{min}. However, we will not go into the detail here and as a rule of thumb, half of the thickness of the pipeline wall corresponds to pressure and structural t_{min}. Of course, if one were to refer to the API 579-1/ASME FFS-1- Fitness-for Service [100] standard, the value 50% for POPD is before the calculated minimum required thickness, the FFS t_{min}. Due to the determination of the TEx variable being based on three different factors (%IR, POPD and TCDA), choosing a higher threshold of say 70% of POPD will result in a lot of excavation in the MEOC project being unjustified. The determination of 50% as the POPD limit also has the added advantage of being a conservative estimate. As was previously mentioned, this value can be in the future, replaced with a higher threshold when more data come in.

The TCDA variable presents a subjective approximation by the Author in terms of its limit for justified excavation. To the Author's knowledge, there is so single value or percentage of TCDA for which excavation is required. The

closest guideline we have is the NACE SP0502:2010 [97]. In the standard, the indication that should be the main trigger for excavation is %IR. This is understandable since, during an ECDA, personnel inspecting the pipe using the DCVG technique are only provided with the inspection values of the voltage drop (%IR). Based on this, the inspector has a rough idea in terms of what to expect from the %IR reading. However, if one were to look at the readings obtained in the MEOC data, the idea that higher DCVG readings are proportional to larger TCDA is not entirely true. Thus, based on the author's experience, a limit of 10,000 cm^2 of TCDA and above was chosen for justified excavation.

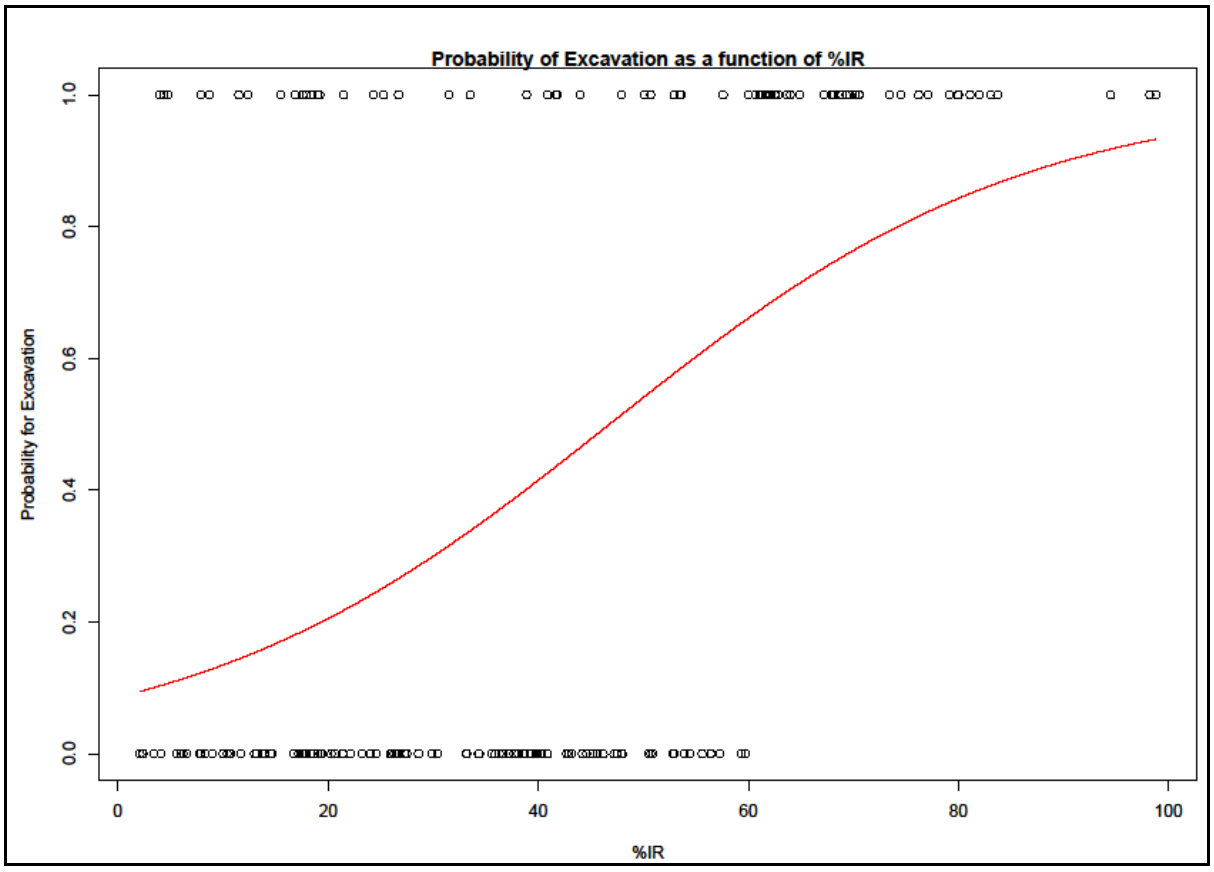


Figure 6-1: Probability of Excavation Based on %IR Considering the TEx Variable

Having made all the necessary considerations of the TEx variable, the resulting variable adopts the form of a binary structure having a value of zero for no excavation and 1 for excavation. Analysis was done using the R software. The following is the resulting model from the analysis.

$$\boldsymbol{logit(TEx) = \log\left(\frac{TEx + \Delta}{1 - TEx + \Delta}\right)} \qquad \textbf{6.1}$$

The symbol Δ represents 0.001, which is a small value so as to render the answer meaningful for all values of TEx. The model for the function is thus,

$$\boldsymbol{Q_{logit(TEx)} = -2.37 + 0.05\alpha} \qquad \textbf{6.2}$$

where α represents the %IR. The coefficient value for α has a p-value of 1.43e-09 which is statistically significant (at significance level of 0.05). The result in Figure 6-1 shows an increasing trend with the greatest probability of excavation occurring at higher %IR. This is expected since the TEx variable is considered to have taken into account the %IR factor. At the absolute lower end of the %IR readings, the probability of excavation is not zero. This is due to the way the data are spread, and to other factors such as the POPD and the TCDA which necessitate excavation even though the readings obtained during the DCVG assessment are zero. This can also be seen at the higher end of the %IR reading. Although the reading is 100% in terms of %IR, the probability of excavation is not 100% certain. This too is the result

of other factors which seems to say a 100% reading of DCVG does not mean a certain excavation. It can also be seen that the amount of missed judgment by the assessors of a project would be quite severe if it is based on the criteria set out in this research. The number of false calls for excavation highlights the inaccuracy of the input of expert judgment which can impact on the project cost (through excavation of unnecessary bell holes). A more "relaxed" approach to this analysis is to not consider the %IR reading to determine the potential excavation but rather to look at the hard evidence which is the data from the direct assessment stage. After all, the main goal of the ECDA is to manage the integrity of pipelines by assessing and reducing the impact of external corrosion activity. Thus, the main driver of maintaining integrity is through the assessment of TCDA and corrosion which is represented by the POPD variable. These two variables shall be our determining factor for the following assessment. The limits for these variables to justify excavation remain the same as before with a TCDA of area of more than 10,000 cm^2 and the pits having 50% depth from the pipe wall thickness. This relaxed version of the TEx variable shall be identified as the $TExx$. The following, Figure 6-2 highlights the outcome of the analysis. The analysis was done using the R software.

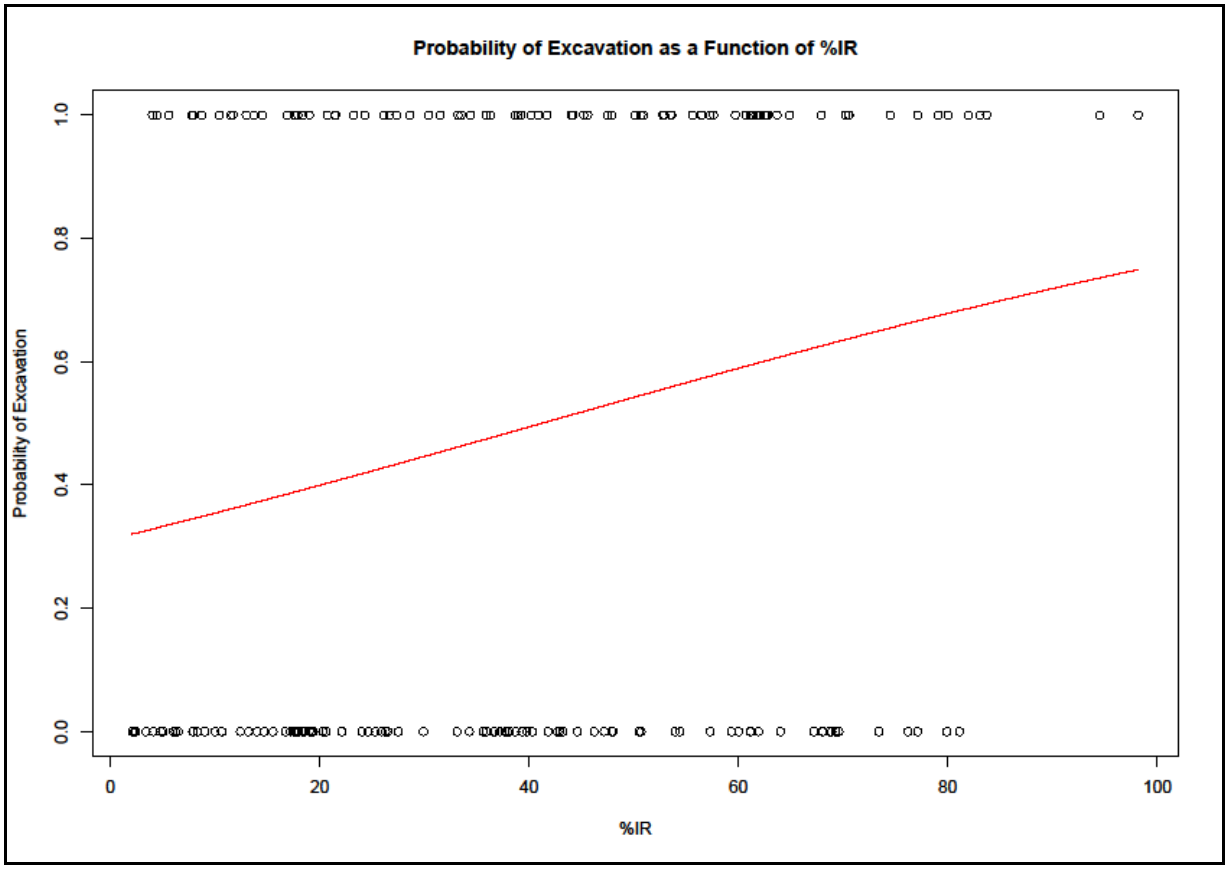


Figure 6-2: The Probability of Excavation based on %IR with considering the $TExx$ Variable

Figure 6-2 shows the effect of the decision-making outcome by not considering the %IR variable as an indicator for excavation. The following model describes the logistic function for $TExx$.

$$\boldsymbol{logit(TExx) = \log\left(\frac{TExx + \Delta}{1 - TExx + \Delta}\right)} \qquad \textbf{6.3}$$

The symbol Δ is a small value that ensures all values of $TExx$ are defined. The model for the probability of excavation is thus,

$$Q_{logit(TExx)} = -0.79 + 0.02\alpha \qquad 6.4$$

The term α is the %IR. Significance is highlighted by a p-value of 0.004. In comparison with the previous analysis, by using $TExx$, the gaps at the lower and higher end of the %IR extremes are larger. The findings here suggest that the contribution of POPD and TCDA to the probability of excavation is quite high regardless of the low values of %IR. The assessment in previous Chapters highlights this scenario where low values of %IR are related to high values of TCDA. This can be seen also at the higher end of the %IR values where a 100% reading of the DCVG assessment still does not implicate a definite need for excavation. The reason for this can be attributed to the error made by the assessor when specifying an excavation for a reading of 100% of the DCVG, but only to find that there are no coating defects present.

The limiting value for POPD to justify for excavation is half the wall thickness. This is somewhat conservative when related to the FFS t_{min} (since the t_{min} for FFS is well below the half way point of the wall thickness – subject to pressure, loading, etc.). If this assessment were to consider a less conservative approach, many more excavations would be deemed unjustified since many fewer locations are needed for inspection. The lesser conservatism would likely "pull" the model downwards in Figure 6-2 which seems to suggest most of the excavations for this project are meaningless. The same can be said for the TCDA factor. A less conservative approach would have made the limiting value higher than 10,000 cm^2 of TCDA for excavation. However, this would result in more unjustified excavations and

having more data points on the no excavation category. As mentioned in previous statements, the value for each limit of the TCDA and POPD variable can be moderated depending on the degree of conservatism that is needed for a particular assessment.

Figure 6-2 also demonstrates the errors committed by the assessors of the project in deciding the location of excavation based on %IR and the input of subjective expert opinion. Although this is the reality, and since an ECDA is an iterative process whereby subsequent assessment are needed for continued safe operation, the findings here may be used as an indicator of the probability of excavation based on the obtained %IR values during a DCVG assessment for future inspection. However, the limitation is that it is only applicable to the MEOC pipelines since the assessments were done on them. More data are needed to make the model more general.

6.3 Logistic Quantile Regression on %IR

LQR has the benefit of accommodating bounded outcomes such as the %IR values where the range is bounded between 0 to a 100%. This ensures that the resulting model will not be outside these bounds (see Model 2a in Chapter 5) making them attractive in situations where a simple linear approach is insufficient to accommodate the non-linear nature of the relationship between an independent variable and its dependent variable. The LQR also has the ability seamlessly to handle data which are skewed in nature, have a non-linear relationship amongst variables and possess non-

constant variances [80] which makes the MEOC data a perfect fit for LQR assessments.

In the previous section, we observe the impact of %IR on the dependent variable TEx. In this section, %IR is determined to be the dependent variable where contributing factors in Table 6-1 are the independent variables. The variables were selected by investigating literature where only the variables that give rise to %IR are considered. Furthermore, only continuous variables are selected for simplicity (also based on results obtained in Chapter 5 on refined versions of the same model) and we filter out the subjectivity in the categorical variables. This is mostly aimed at the soil variable where interpretation of the soil grains and structure is highly subjective. Initial assessments by using LQR were done by experimenting with different continuous variables and observing each effect separately on the dependent variable %IR. Subsequently, the coefficients which show a good relationship (in terms of their coefficient magnitude) were chosen as the variables to be considered for the final combination model. The logit transform of the dependent variable, in this case is %IR, is given as follows,

$$\boldsymbol{logit}(\%\boldsymbol{IR}) = \mathbf{log}\left(\frac{\%\boldsymbol{IR} + \boldsymbol{\Delta}}{\mathbf{100} - \%\boldsymbol{IR} + \boldsymbol{\Delta}}\right) \qquad \textbf{6.5}$$

where Δ is a value to define all possible values of %IR. Instead of 1 in the log transform for the previous analyses for TEx and $TExx$, we have 100 to represent the 100% maximum value of the %IR range.

6.3.1 TCDA vs %IR

An assessment was done to see how the TCDA would contribute to the %IR value by employing the LQR method. A common understanding within the industry states that the DCVG assessment is done solely to identify and assess the severity of pipeline coating defects. Therefore, it is expected that these two variables are highly correlated. However, after looking at the data from the MEOC project, the case turns out to be the opposite. The linear correlation value is poor (based on Chapter 4) which suggests little to no effect between the two. Therefore, it is only wise to assess the contribution of TCDA to %IR in this section to be nonlinear. Figure 6-3 below highlights the relationship between the TCDA and %IR.

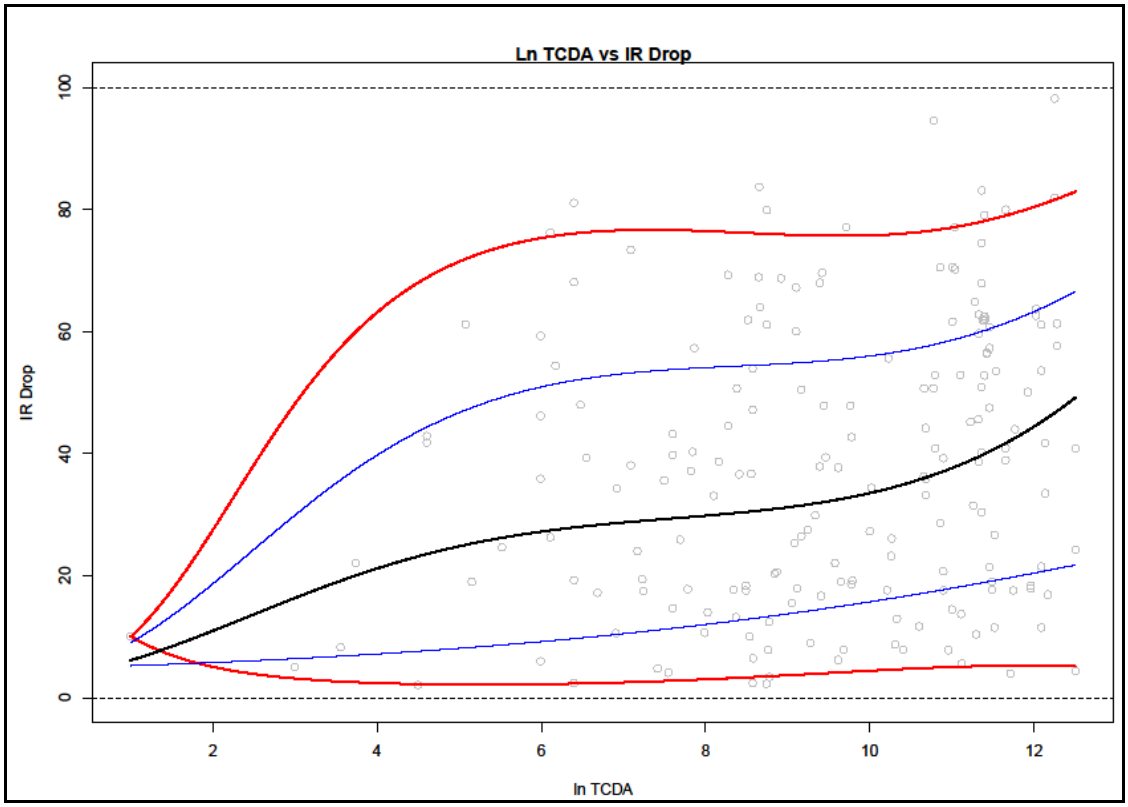


Figure 6-3: LQR on ln(TCDA) vs %IR

Figure 6-3 shows the relationship between ln(TCDA) and its response %IR. Prior to the assessment, the TCDA variable was transformed by applying the natural log to obtain a much more meaningful relationship. If the transformation was not done, the effect of TCDA on %IR would not have been substantial. The resulting model is as follows.

$$\boldsymbol{Q_{logit(\%IR)}(\tau) = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3} \qquad \textbf{6.6}$$

β_1 is the intercept and β_2, β_3 and β_4 all represent coefficients for X where X is ln(TCDA). τ is the quantile of the various models. The coefficient values for each quantile are given in the table below. The polynomial order used in this Chapter was determined based on the general structure of how the data is spread and also from trial and error. As a general rule, the lowest order is preferred for model's brevity and simplicity.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-1.16389	-2.9779	-3.56713	-3.40549	-3.82327
$\boldsymbol{\beta_2}$	-1.18056	0.07914	0.95152	1.23253	1.83168
$\boldsymbol{\beta_3}$	0.16192	0.00726	-0.11755	-0.14431	-0.21987
$\boldsymbol{\beta_4}$	-0.00629	-0.00022	0.00512	0.00575	0.00863

Table 6-2: Estimated LQR Coefficient Values for ln(TCDA) vs %IR

From the LQR estimated coefficients, it can be seen that the value of β_2, which is the ln(TCDA) is increasing from the lowest quantile to the highest quantile. The max value of β_2 occurs at the 0.95 quantile with a coefficient value of 1.83. The increase is expected since a larger TCDA will have the biggest effect on %IR. Similarly, we can see for the lowest quantile, 0.05, the coefficient value is -1.18. For the non-linear terms, β_3 and β_4 the two coefficients show opposite trends across the quantiles. The coefficient values for β_3 show positive coefficient for the first two quantiles while the rest are all negative. β_4 illustrates a picture which is quite the opposite. Figure 6-4 shows the quantile plot for the LQR estimated coefficients, β_2, along with its 70% confidence interval. All the confidence intervals (70%) referred to for the remaining analyses in this Chapter shall be according to this. It can be seen that most of the estimate's confidence intervals do not include zero which seems significant.

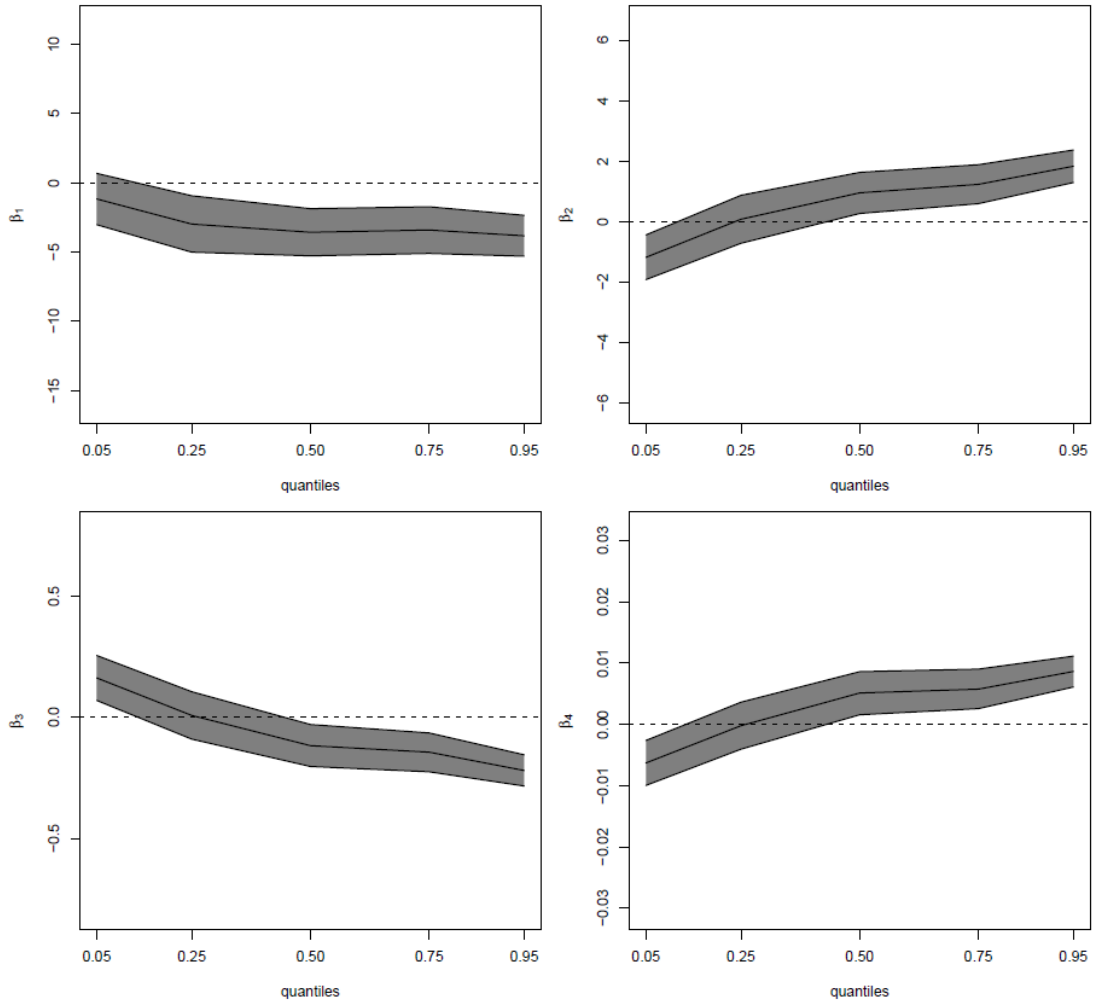


Figure 6-4: Estimated Coefficients across the Quantiles by LQR for β_1, β_2, β_3 and β_4 (TCDA)

6.3.2 SR vs %IR

Next, we assess the effect of SR on %IR by means of the LQR. The variable SR is transformed into the square root of SR (SQRT(SR)) to strengthen the relationship between the two variables. After transformation, the relationship of SR on %IR is more evident. The purpose of the assessment between the SR and %IR is to determine whether the variable is relevant in terms of the coefficient's magnitude for the final combined model. SR serves as an important model since it has influence on the %IR reading from a DCVG

assessment [64]. Generally, it can be said that as SR increases so does the %IR value. The following is the model between SR and %IR and its corresponding equation from the analysis.

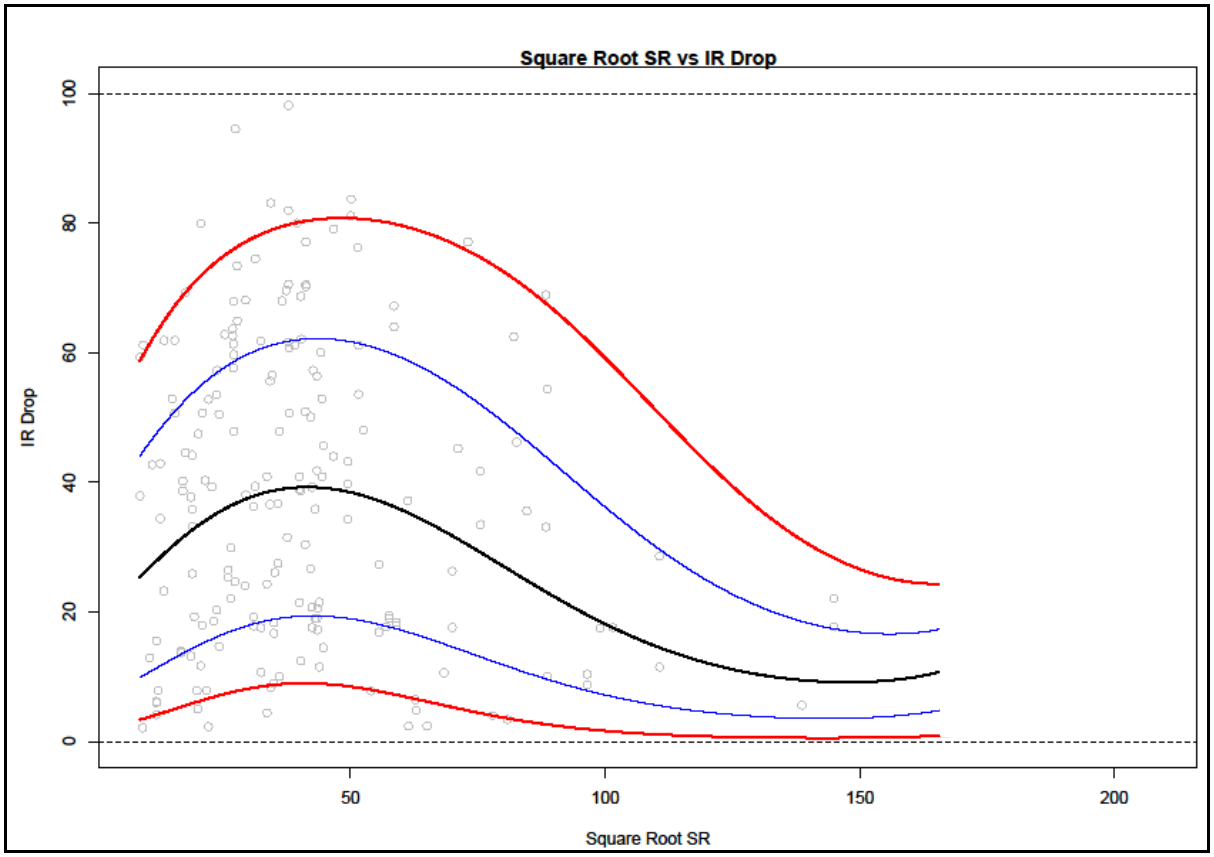


Figure 6-5: LQR on SQRT(SR) vs %IR

Figure 6-5 shows the outcome of the modelling process by using the LQR on the variable of interest. The relationship between SQRT(SR) and %IR does not seem to be straightforward and in general, seems to imply an increase in SQRT(SR) means a decrease in %IR. Note that for the lowest quantile, the model is limited by 0 which illustrates the advantage of using LQR. The models corresponding to each quantile are further clarified in the equation below.

$$\boldsymbol{Q_{logit(\%IR)}(\tau) = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3} \quad \textbf{6.7}$$

β_1 is the intercept and β_2, β_3 and β_4 are the estimated coefficients for X where X is the SQRT(SR). τ is the quantile of the various models. The coefficient values for each quantile are given in the table below.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-4.07163	-2.72031	-1.51144	-0.69876	-0.23727
$\boldsymbol{\beta_2}$	0.09503	0.06745	0.05713	0.06057	0.0768
$\boldsymbol{\beta_3}$	-0.00149	-0.00103	-0.00088	-0.00089	-0.00103
$\boldsymbol{\beta_4}$	0.00001	0	0	0	0

Table 6-3: Estimated LQR Coefficient Values for SQRT(SR) vs %IR

Judging by the coefficient estimates in Table 6-3, β_2 (SQRT(SR)) shows a decreasing trend from the 0.05 quantile up to the median quantile. After this the estimated coefficient value increases until the 0.95 quantile. The highest estimated value is at the 0.05 quantile which is 0.095. This can also be summarised as, lower values of %IR are most affected by the SQRT(SR) while the median value of %IR is least affected by SQRT(SR). Upper values of %IR seem to be affected by the higher SQRT(SR) values but not as much as the 0.05 quantile. β_3 and β_4 showed trends which are different to that of the β_2 coefficient. All the β_3 showed negative values while the β_4 coefficient is virtually zero. The trends of each coefficient follow the trend we see in Figure 6-5. The quantiles for the estimated coefficients are given below.

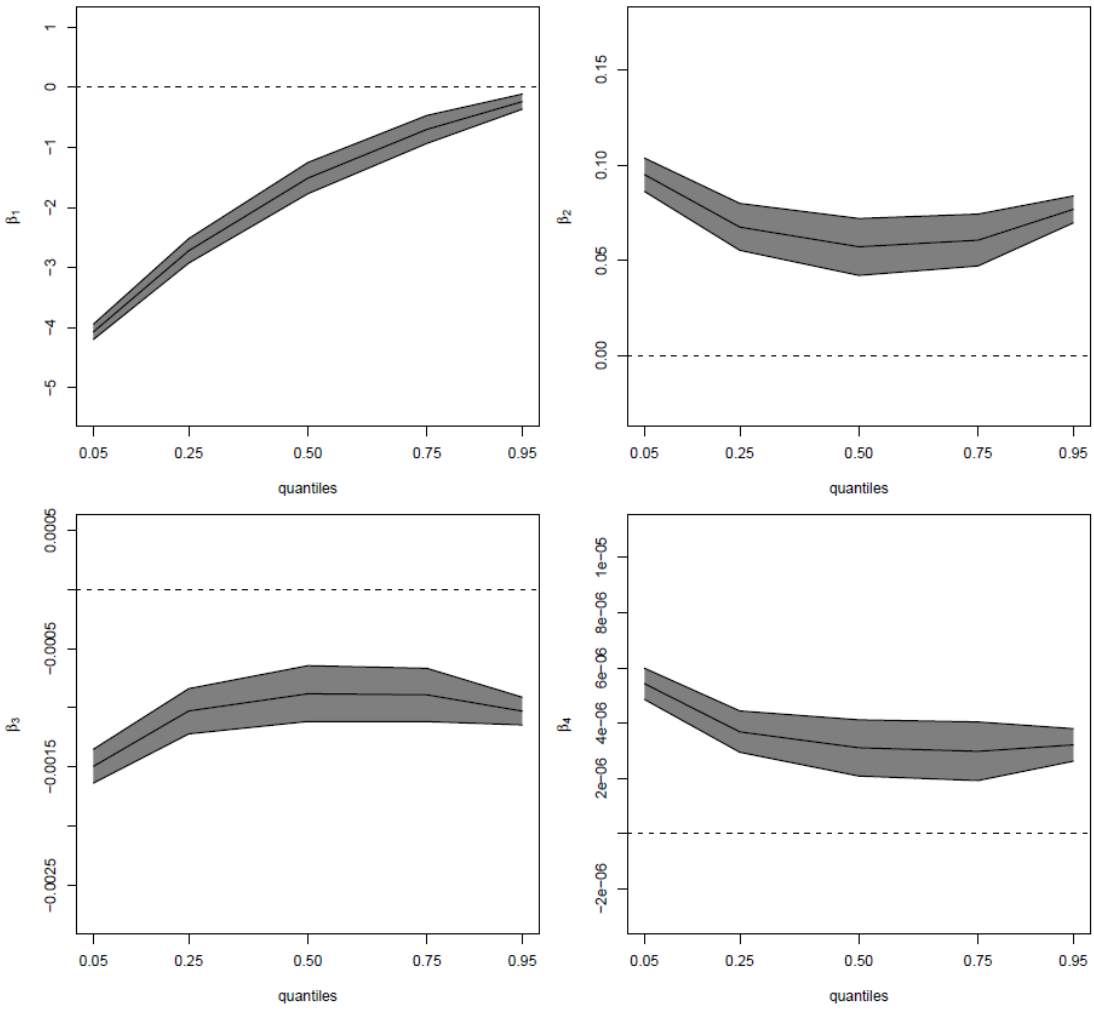


Figure 6-6: Estimated Coefficients across the Quantiles by LQR for β_1, β_2, β_3 and β_4 (SQRT(SR))

From Figure 6-6, it can be clearly seen that the LQR estimated coefficients confidence intervals do not contain zero which is statistically significant. The result also seems to contradict the findings from [64] where higher SR translates into higher %IR readings. The initial part of the assessment in Figure 6-6, below the value of 50 of the SQRT(SR) does suggest this is the case but after the model peaks a decreasing trend is observed. From this set of data, it can be considered, generally that as SQRT(SR) increases, the contribution to %IR decreases. This is also echoed in Figure 6-6 where a general decreasing trend is observed.

6.3.3 POPD vs %IR

To continue our objective of constructing a combined model which includes various variables that give rise to %IR, the independent variable POPD is assessed against %IR. POPD represents the extent of external corrosion experienced by the pipeline. The DCVG assessment is not known or built for the purpose of identifying corrosion on pipelines. The main purpose of the technique is to identify and quantify the severity of coating defects. Although this is the case, the slight R^2 value for the Total Corrosion Area (TCA) and Total Corrosion Volume (TCV) in Chapter 4 gives an indication that the two variables might have a connection. No prior transformation was done on the variable POPD. The models for the various quantiles corresponding to the two variables are as follows.

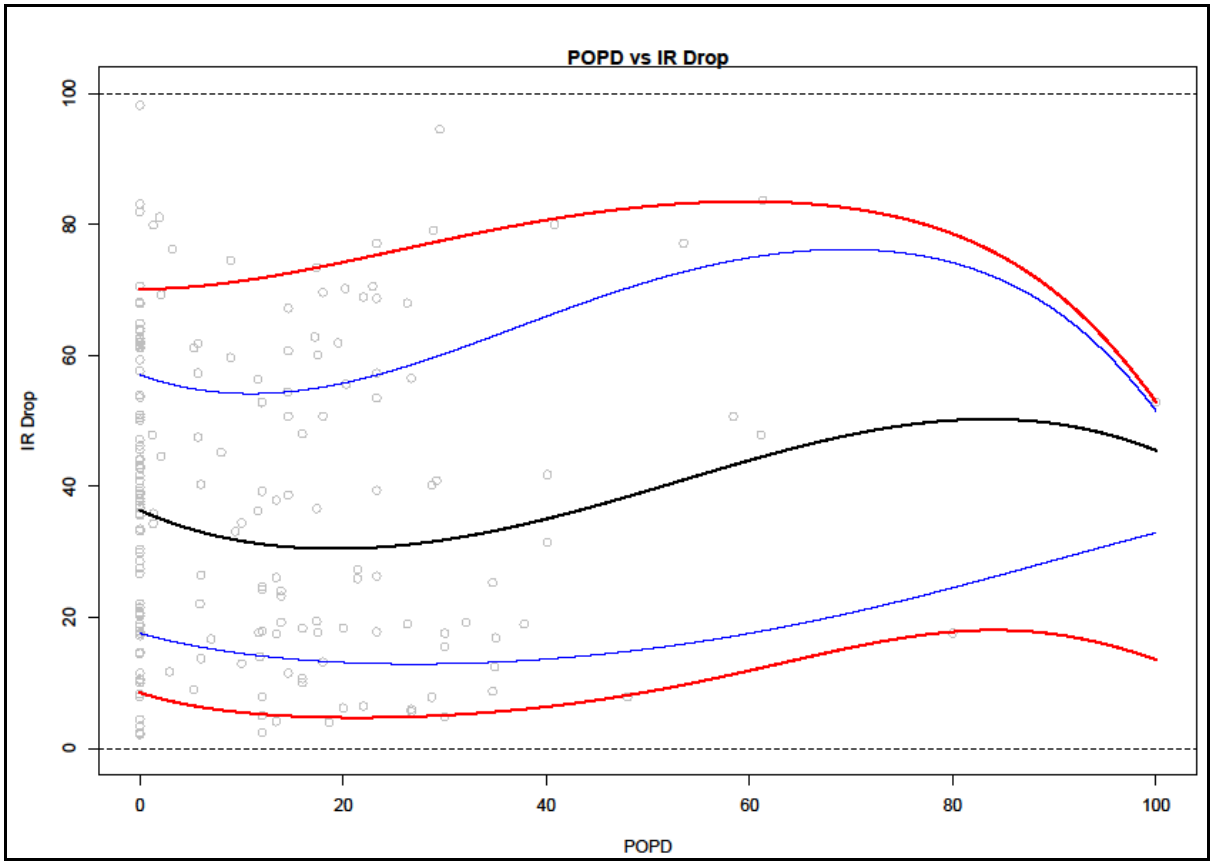


Figure 6-7: LQR on POPD vs %IR

Figure 6-7 shows the models of the corresponding quantiles by LQR between the POPD and %IR variable. The models show a trend which suggests the relationship to be considerable. Generally, it can be suggested that with increasing POPD there is a slight increase in %IR. To quantify this relationship further, equations of the corresponding models are given below.

$$\boldsymbol{Q_{logit(\%IR)}(\tau) = \beta_1 + \beta_2\, X + \beta_3\, X^2 + \beta_4\, X^3} \qquad \textbf{6.8}$$

β_1 is the intercept coefficient and β_2, β_3 and β_4 are the estimated coefficients for X. X is the represented variable for POPD. τ is the quantile of the various model. The coefficient values for each quantile are given in the table below.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-2.37372	-1.54716	-0.56016	0.28276	0.85231
$\boldsymbol{\beta_2}$	-0.06517	-0.02905	-0.02994	-0.02253	0.00045
$\boldsymbol{\beta_3}$	0.00192	0.00065	0.00096	0.0012	0.00063
$\boldsymbol{\beta_4}$	-0.00001	0	-0.00001	-0.00001	-0.00001

Table 6-4: Estimated LQR Coefficient Values for POPD vs %IR

Table 6-4 illustrates the LQR estimated coefficient values for POPD vs %IR which shows a small but positive relationship between POPD and %IR. The maximum estimated positive coefficient value by LQR for the coefficient β_2 occurs at the 0.95 quantile with a value of 0.00045. The minimum, which is also the highest value in terms of magnitude, occurs at the 0.05 quantile with a coefficient value of 0.065 in the logit of %IR. Based on simply looking at the β_2 coefficient, the finding here does seem to suggest there exists a marginal effect on %IR by POPD. The effect of deeper pits is felt at the higher end of the %IR values. However, for shallower pits, the estimated values seem to suggest that %IR readings lose their detection ability due to the pitting effect. For the non-linear terms, β_3 and β_4 showed low values across the quantiles. The effect is also different between the two where β_3 shows positive estimates while β_4 illustrates negative values except for quantile 0.25 across the quantiles of the logit of %IR. The quantiles corresponding to the LQR estimated coefficients in Table 6-4 are given as follows.

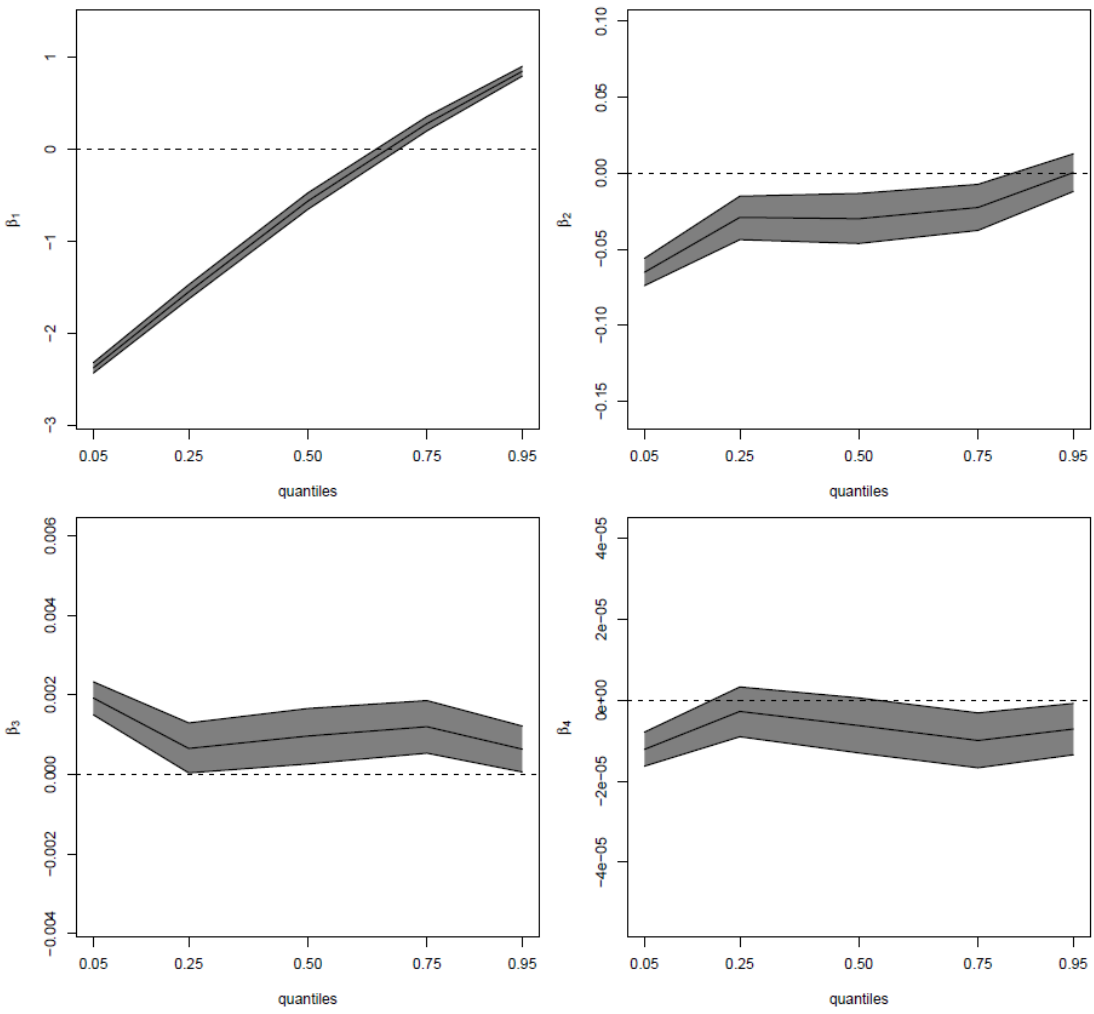


Figure 6-8: Estimated Coefficients across the Quantiles by LQR for β_1, β_2, β_3 and β_4 (POPD)

Looking at Figure 6-8, it can be said that generally all the estimated coefficients for POPD are significant. For the β_2 coefficient, the confidence band has captured the zero value at the 0.95 quantile thus rendering it insignificant. As for the β_4 coefficient, the estimates at the 0.25 to the median quantile included zero in its confidence interval. Overall considering the finding here and referring to Figure 6-7, it can be said that the effect of POPD on %IR is rather small due to the modest peaks and valleys observed in the quantile models observed in Figure 6-7 .

6.3.4 DOC vs %IR

The DOC of the pipeline represents the depth of the soil above the pipeline as to which is used to cover them (depth of the buried pipe). The depth of pipes has an effect on the %IR signal since the equipotential lines “emitted” by the pipe are affected by the vertical distance between the pipe and the inspector using the DCVG probe. The deeper the pipe, the larger this distance is and the smaller the %IR signal measured. This industry understanding is supported by works such as [64], [65].

Moving on with our construction of the combination of variables for the %IR model, the effect of the depth of pipe is tested with the %IR indication using the LQR. Prior to the logit transformation of the %IR variable, the DOC was also transformed by applying a natural log to the variable. The resulting transformation results in a better relationship between the DOC and %IR. The produced models are given below.

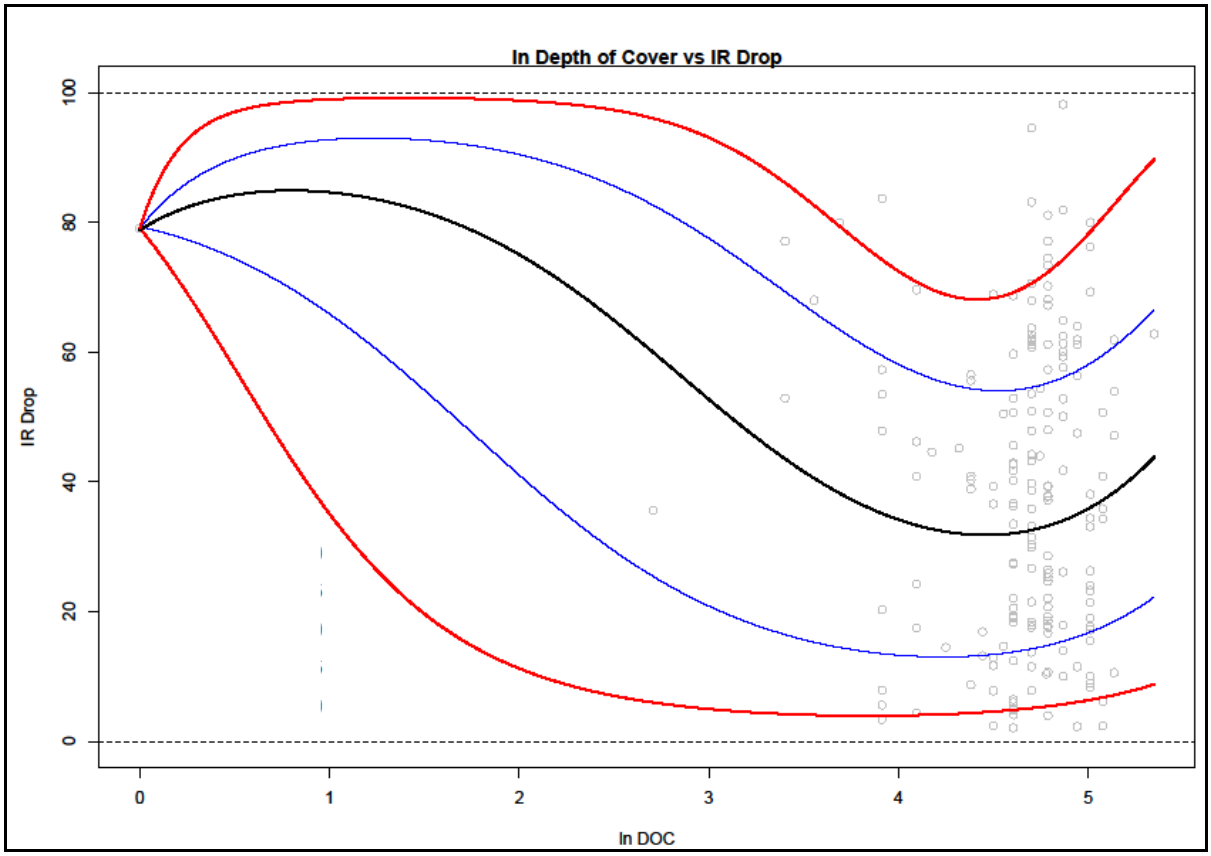


Figure 6-9: LQR on ln(DOC) vs %IR

Figure 6-9 has shown that the estimated models by the LQR resonate with the common understanding in the aforementioned statements that the ln(DOC) has a negative effect on the %IR signal. The assessment conducted here is much superior over traditional regression methods such as the OLS where it provides a solution that is within the %IR bounds and also has the added advantage of observing the effects of different quantiles across the changing DOC. Traditional mean response regression such as OLS would clearly miss this insight. To delve into this further, the corresponding equations for the generated models are given below.

$$\boldsymbol{Q_{logit(\%IR)}(\tau) = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3} \quad \textbf{6.9}$$

The β_1 coefficient is the intercept and β_2, β_3 and β_4 are the LQR estimated coefficients for X. ln(DOC) is denoted by X. τ is the quantile for each of the corresponding model. The coefficient values for each quantile are given in the table below.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	1.33153	1.34467	1.31433	1.32882	1.33082
$\boldsymbol{\beta_2}$	-2.16976	-0.39198	1.09343	2.24786	5.51194
$\boldsymbol{\beta_3}$	0.21038	-0.35796	-0.80429	-1.16402	-2.58555
$\boldsymbol{\beta_4}$	0.01253	0.06367	0.10195	0.13486	0.29621

Table 6-5: Estimated LQR Coefficient Values for ln(DOC) vs %IR

Table 6-5 shows the estimated values for the coefficients by using LQR. It is surprising to see that the β_2 coefficient provides an upward trend of the %IR signal going across the quantiles. This maximum occurs at the 0.95 quantile where the coefficient value is 5.512. At the 0.95 quantile, the estimated coefficient value is 5.512. The minimum value occurs at the 0.05 quantile with an estimated value -2.170. Results based on these findings seem to contradict the previous statements that deeper pipes will give a much weaker %IR signal. The results for β_2 describe the linear part of the model which is the initial portion of the model in Figure 6-9. The downward trend can be seen “generally” when including all the non-linear terms into consideration. It

comes as some surprise that β_3 possesses this trend with a maximum magnitude occurring at the 0.95 quantile with a value of -2.586. Quantile plots relating to the LQR models are given below.

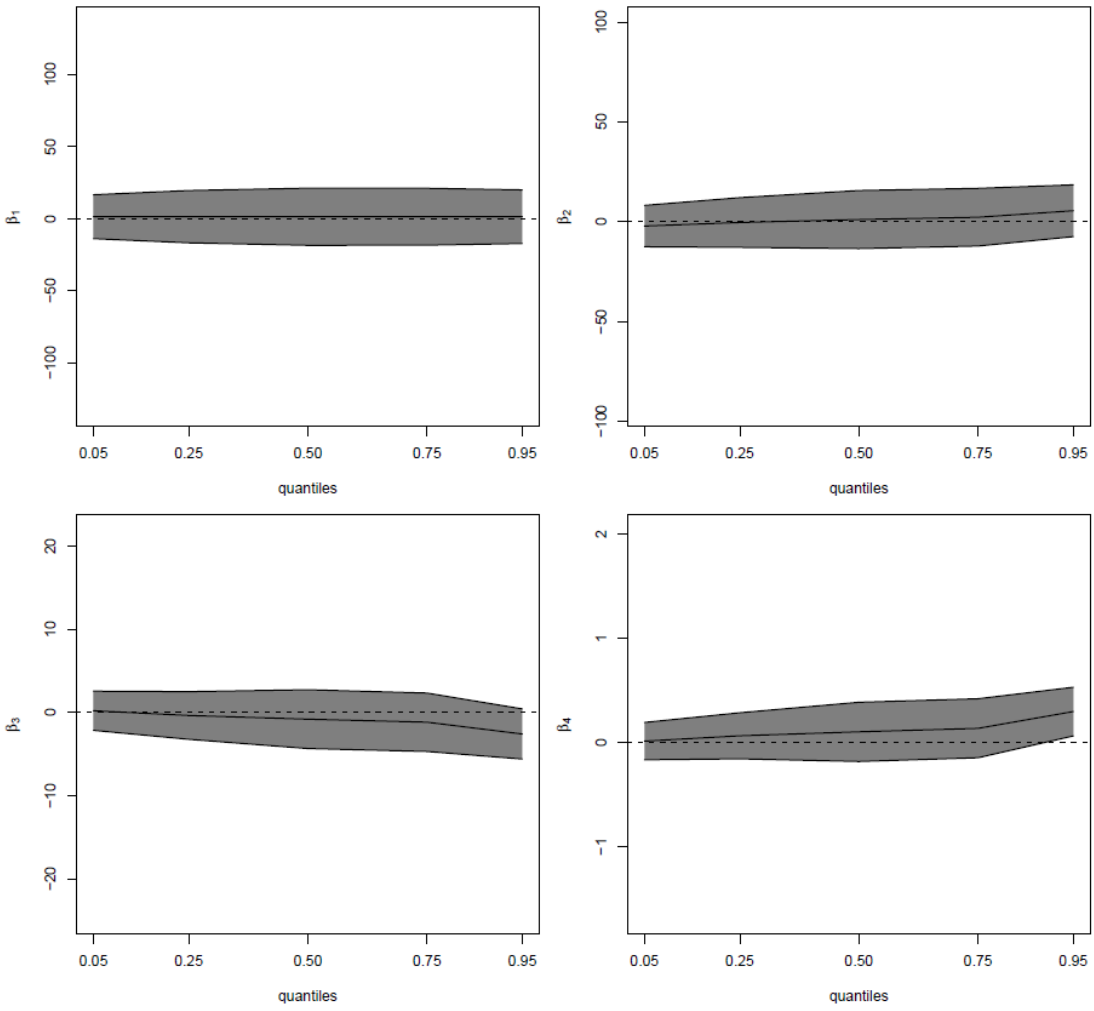


Figure 6-10: Estimated Coefficients across the Quantiles by LQR for β_1, β_2, β_3 and β_4 (ln(DOC))

The estimated quantiles in Figure 6-10 show that not all the estimated coefficient values are statistically significant (at the 0.3 level). At quantile 0.95 for β_3, the confidence band nearly escapes the zero value with the edge barely touching the zero line. Although this is the case, it does still include the zero value within its confidence interval which means the estimate is still

insignificant. Only at the 0.95 quantile of the β_4 coefficient is the estimate statistically significant.

However, the results from Figure 6-10 do correlate well with industry's understanding on the matter. The confidence band represented here shows the level of confidence of the estimation. But it is not 100% certain that the true effect between ln(DOC) and %IR is non-existent. Therefore, the variable DOC is chosen (due to the industry understanding of the variable) to be included in the final combined model despite it being statistically insignificant.

6.3.5 DUC vs %IR

The variable DUC can be considered as the corrosion product of the pitting process on the pipe with or without the additional foreign material such as sand, stones, etc. occurring underneath the coating of the pipe. The DUC is thought to have no significant contribution to the signal generated for %IR. However, the presence of DUC can correlate with pitting and has the potential to affect the %IR signal (based on previous assessments on POPD where a small effect is observed). For the purpose of constructing the combined %IR, the DUC is examined to see whether it has anything to contribute in terms of signal strength. The following is a representation of the models corresponding to each quantile constructed based on the LQR for DUC vs %IR. The variable DUC was not transformed before the assessment.

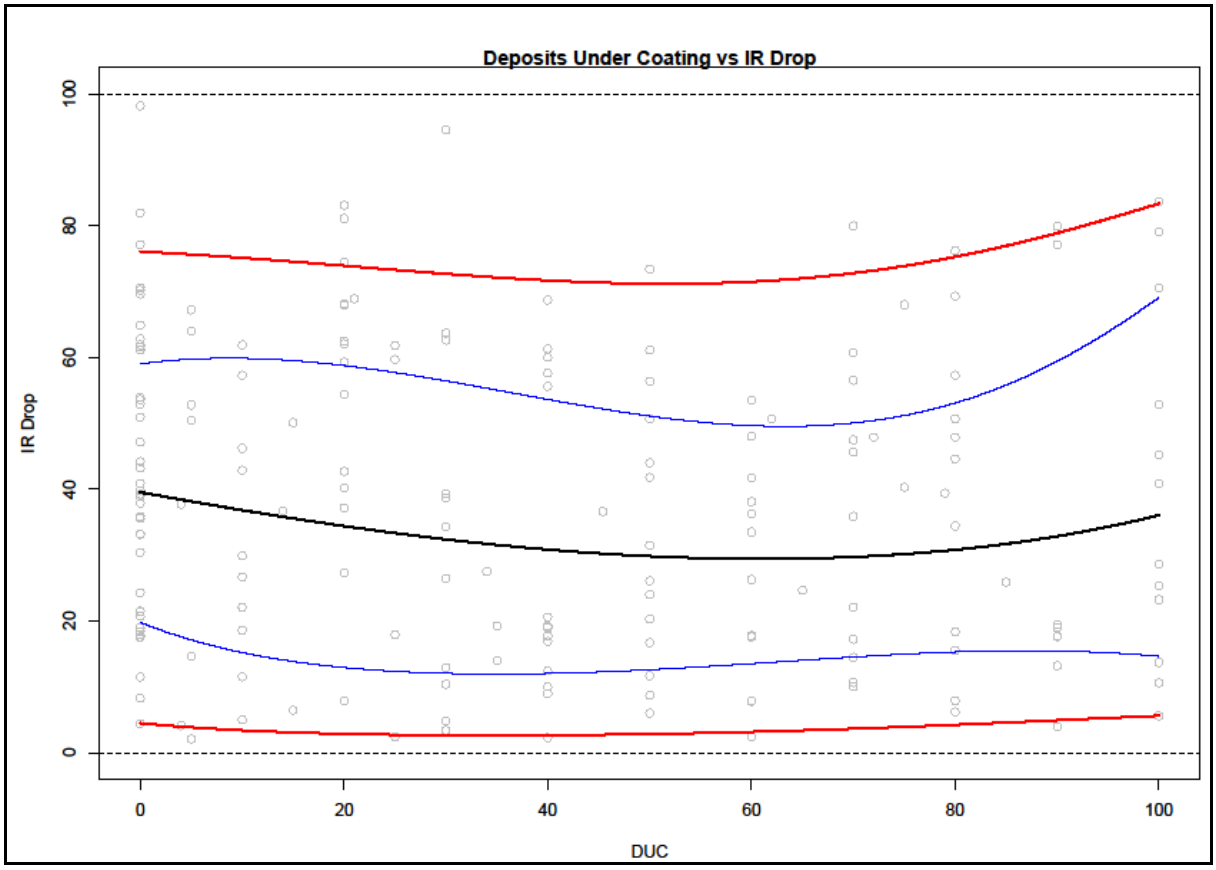


Figure 6-11: LQR on DUC vs %IR

The trend in Figure 6-11 indicates low correlation between the two variables across the quantile. Observing the 0.05 and 0.95 quantile, the apparent "flatness" is similar. However, regarding the quantiles within these two bands shows a pronounced relationship especially the 0.75 quantile. This would be overlooked if one were to do an OLS based regression on the same dataset where only the mean (similar to the median) response is generated. The median also shows similar trend to the two-outermost quantiles (0.05 and 0.95). Equations relating to these models are given below.

$$\boldsymbol{Q_{logit(\%IR)}(\tau) = \beta_1 + \beta_2\, X + \beta_3\, X^2 + \beta_4\, X^3} \qquad \textbf{6.10}$$

The β_1 coefficient is termed the intercept. β_2, β_3 and β_4 represent the LQR estimated coefficients for X. X is defined to be the DUC and τ is the quantile for each of the corresponding models. The coefficient values for each quantile are given in the table below.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-3.05626	-1.39934	-0.42485	0.36487	1.15589
$\boldsymbol{\beta_2}$	-0.03509	-0.03937	-0.01184	0.00853	-0.00409
$\boldsymbol{\beta_3}$	0.00065	0.00079	0.00003	-0.00056	-0.00012
$\boldsymbol{\beta_4}$	0	0	0	0.00001	0

Table 6-6: Estimated LQR Coefficient Values for DUC vs %IR

The coefficient values estimated by using LQR are given in Table 6-6. β_2 of the DUC variable represents the linear part of the model, where an uncertain trend can be seen. The coefficient values dips between the 0.05 quantile to the 0.25 quantile. After this it increases to the 0.75 quantile and drops again at the 0.95 quantile. The maximum estimated coefficient occurs at the 0.75 quantile with an estimated value of 0.009. Non-linear terms showed small values which suggests little to no effect on the contribution to the logit of %IR. The following quantile plots are the derivation from the estimated quantile values in Table 6-6.

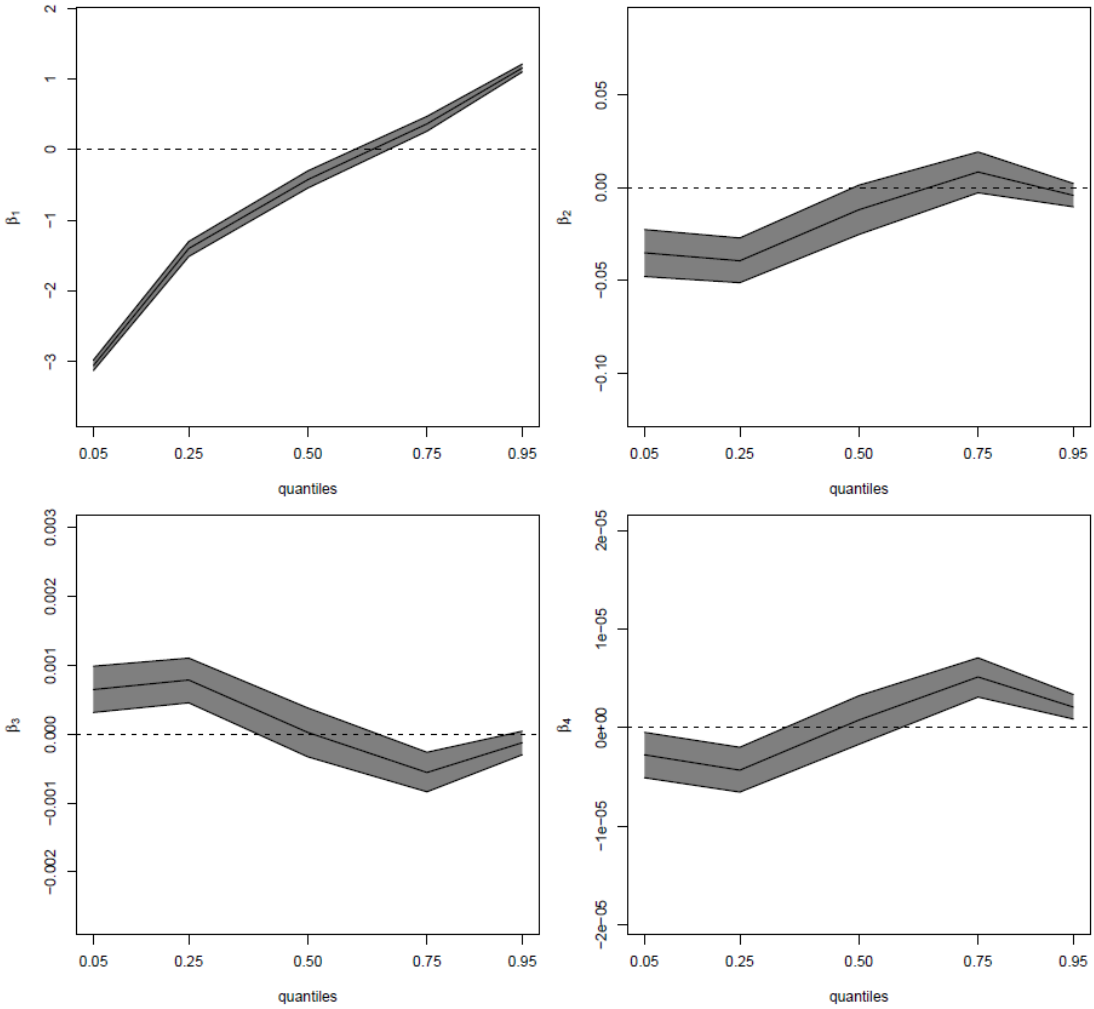


Figure 6-12: Estimated Coefficients across the Quantiles by LQR for β_1, β_2, β_3 and β_4 (DUC)

The quantile trend for β_2 in Figure 6-12 shows generally an increasing trend with a dip occurring at the 0.95 quantile. The increasing trend is suggestive that there might exist an increasing effect of DUC on the higher quantiles of the %IR. However, the higher quantiles for β_2 lie in a region where zero terms are included within the confidence interval. From a statistical standpoint, this means that the estimation is not significant and may be due to sampling error. For the lower quantiles, zero terms are not included in the confidence band which suggests significance. While in these quantiles, may look to increase, the estimated values are negative in character which can

only mean that higher amounts of deposits underneath the coating affect the %IR signal in a negative way (decreasing signal – similar to the one obtained for POPD). Coefficients for β_3 and β_4 generally show estimates which are statistically significant. However, the coefficient values for these non-linear terms are small which implies little effect. Taking this and Figure 6-11 into consideration, it can be summarised that the effect of DUC on the generation of the %IR signal is minute.

6.3.6 Combined Model for %IR

A combined model was constructed based on the previous assessments of each variable. Each of the variables were assessed separately to identify its relevance to %IR. Based on these separate assessments, all the variables were initially assumed (whether linear or nonlinear) to be part of the model. The outcomes of the relationship were later looked at and its relevance for %IR is considered (this stage is still within the separate assessments). After all the variables from these assessments are considered, a careful selection process was done to determine which of the assessed variables are required for the combined model. The criterion of such selection was firs based on the general engineering (industry) understanding of the variable and secondly on its statistical relevance by looking at its coefficient's magnitude and statistical significance. An example would be where it is commonly understood within the engineering field that a variable is correlated to the dependent variable but its relevance in terms of statistical inference is insignificant – the variable

will still be selected regardless of its negative statistical outcome. The following is the combined model after the selection process.

$$Q_{logit(\%IR)}(\tau) = \beta_1 + \beta_2\, lTC + \beta_3\, lTC^2 + \beta_4\, lTC^3 + \beta_5\, SRSR + \beta_6\gamma + \beta_7\varepsilon + \beta_8\varepsilon^2 + \beta_9\varepsilon^3 + \beta_{10}\,\delta \quad \textbf{6.11}$$

where:

lTC = ln(TCDA)

$SRSR$ = Square Root SR

γ = POPD

ε = ln DOC

δ = DUC

τ = Representing the quantile in question

It can be seen from the model above that all the variables that have gone through separate assessments regarding %IR have been selected. However, the linear and nonlinear terms for each variable were selected based on the magnitude of their coefficients. This, and the consideration of variables from an engineering viewpoint results in the chosen variables. A good example of this is the variable ln(TCDA). It is known as a fact that the DCVG instrument was built for the purpose of identifying coating defects and to quantify their severity in terms of size. As such, the variable ln(TCDA) is crucial in examining its effect on the %IR signal. On top of this, we can see that the magnitude of the ln(TCDA) variable, linear or nonlinear, is substantial and makes it suitable for the combined model. Another example of choosing variables based on engineering judgement is the square SQRT(SR) and the

POPD. These two variables are important in the eyes of a researcher where excluding one would lead to a less meaningful model. The variable ln(DOC) is another important variable in determining the %IR reading. The concept of depth of pipe and how it relates to %IR readings is already mentioned in the previous sections. Thus, all the linear as well as the nonlinear terms of the ln(DOC) variable are included. However, if one observes the separate assessment of the ln(DOC) and %IR, it is found that the estimated coefficients confidence interval included the zero term. Although this is the case, the author believes that leaving out such an important variable in the combined model would lead to a model which is incomprehensive. As previously mentioned, the variable DUC is chosen due to the linkages it has with variables such as the POPD which could lead to the %IR signal being different. The combined model's estimated coefficients are given below.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	1.78796	-0.37094	-1.4325	-2.06227	-3.9055
$\boldsymbol{\beta_2}$	-1.1501	0.36574	1.01971	1.32537	2.29261
$\boldsymbol{\beta_3}$	0.20895	-0.02844	-0.12883	-0.16524	-0.32022
$\boldsymbol{\beta_4}$	-0.00954	0.00123	0.00564	0.0069	0.01396
$\boldsymbol{\beta_5}$	-0.00412	-0.00615	-0.00455	-0.00406	-0.00219
$\boldsymbol{\beta_6}$	-0.00919	0.00197	0.00237	0.00717	0.01062
$\boldsymbol{\beta_7}$	-0.23256	2.09442	2.53372	4.23192	5.05638
$\boldsymbol{\beta_8}$	-0.76513	-1.47309	-1.47657	-2.10135	-2.22377
$\boldsymbol{\beta_9}$	0.13855	0.18969	0.18002	0.24347	0.24172
$\boldsymbol{\beta_{10}}$	0.00092	-0.00348	-0.00362	-0.00483	-0.00164

Table 6-7: Estimated LQR Coefficient Values for the Combined Model

Based on Table 6-7, the most important variable, ln(TCDA), shows the coefficient values β_2 with an increasing trend peaking at the 0.95 quantile. The maximum estimated coefficient value for ln(TCDA) is 2.293. Conversely, the minimum value occurs at the 0.05 quantile with a coefficient value of 1.150. A closer look at all the coefficient values related to ln(TCDA), β_2, β_3 and β_4 reveals they have different characteristics. The β_2 coefficient starts with a positive estimate with negative predictions all the way until the 0.95 quantile. The trend is quite the opposite for the β_3 coefficient estimates. For the β_4 estimates, the trend is similar to that of β_2. The different trends seem to suggest that the model is fluctuating. This can only be clarified by conducting a derivative on the %IR combined model to make sense of what

this actually means. This is clarified in greater detail in the discussion section.

For the variable SQRT(SR), the maximum estimated coefficient value, β_5, occurs at the 0.05 quantile at a value of 0.00412. As one scans across the quantiles, β_5 estimated values decrease in magnitude. This can be interpreted as the effect of the SQRT(SR) on larger %IR being less negative than the lower %IR readings. Based on the increasing trend of β_5, it can also be stipulated to mean the effect of the SQRT(SR) is larger (less negative) on larger %IR readings and lower (more negative) on lower %IR readings, despite the estimated values being negative across all the quantiles. As suggested in these findings, the effect of SQRT(SR) on %IR is supported by previous studies such as [64] but only as a trend. Based on this MEOC data set, the values of the coefficient β_5 potrays a different outlook on a common engineering understanding (%IR is affected positively by increasing SR values).

The variable POPD which is represented by the coefficient β_6, shows an increasing trend of estimates along the quantiles. This seems to say that deeper pits have increasing effect on larger %IR readings compared to shallower pits which have a negative effect on lower %IR indication. The maximum estimated β_6 is found at the 0.95 quantile with a value of 0.011. The lowest value (magnitude) obtained from the LQR analysis for the POPD variable β_6, occurrs at the 0.25 quantile with a value of 0.002. Based on the estimated coefficient values across the quantiles, the variable POPD shows an increasing trend to contribute to %IR where deep corrosion pits affect

mostly higher readings of the %IR. Previously, it was thought that corrosion does not contribute to the %IR reading from a DCVG indirect assessment. However, the results presented here, and the results obtained in Chapter 4, suggested quite the opposite perception. It also highlights that POPD affects the %IR in a more complex manner than previously thought. This is because, during the initial separate assessment, non-linear terms were also included in the modelling which resulted in a trend which is quite the opposite to what was found here. After several other variables were combined with the POPD variable and non-linear terms were not considered, only then the resulting estimates portray a different picture. Therefore, the variable POPD reacts with other variables to give a positive increasing effect to %IR. However, the resulting coefficient for the POPD in the %IR combined model points to an effect which is minor. Further research is needed to verify this claim where a physical experiment and simulation are needed to prove the findings obtained in this thesis.

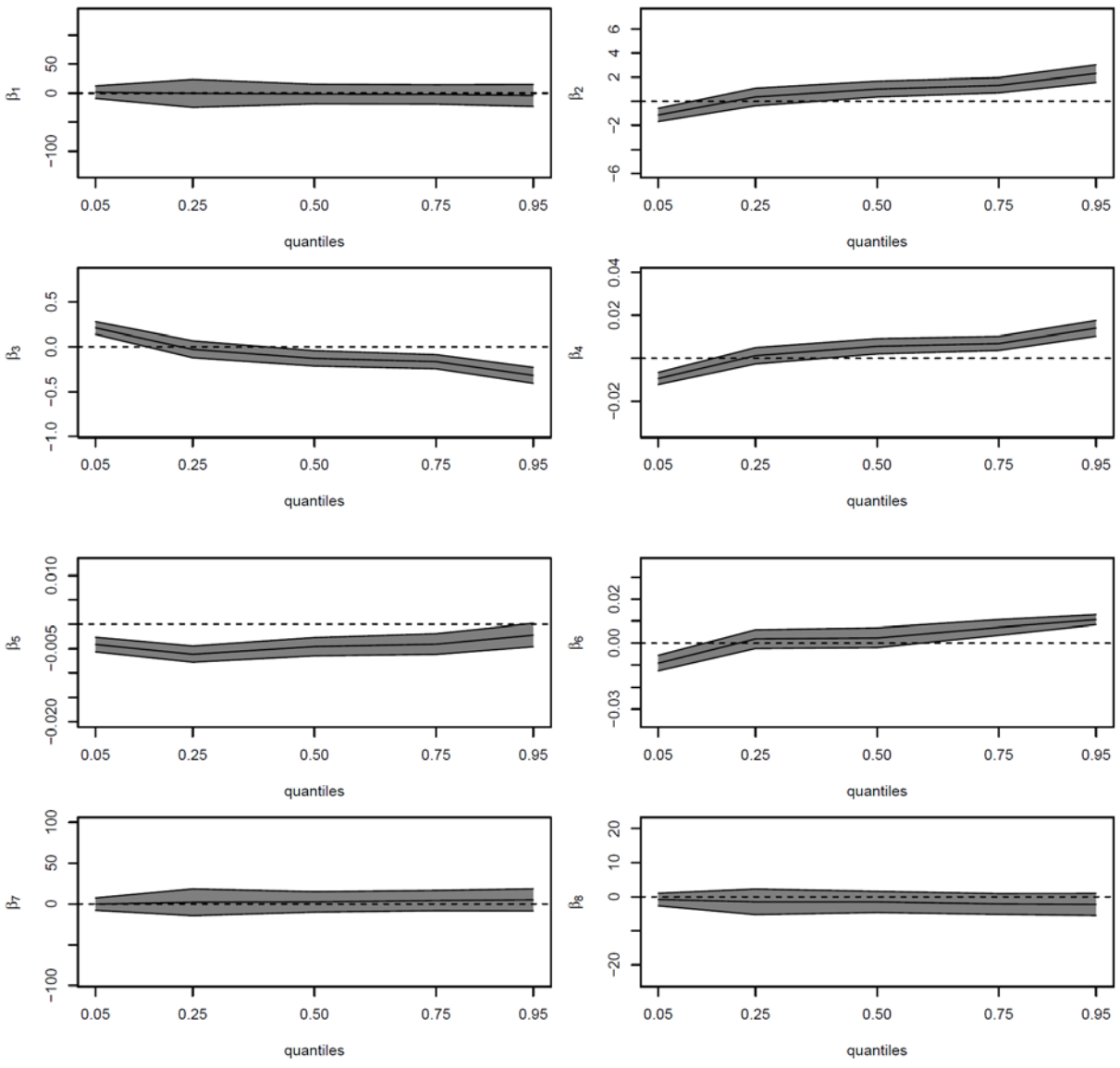
For the variable, ln(DOC), the results obtained in Table 6-7 are similar to the ones obtained in the separate assessment of ln(DOC) and %IR. The trend of the estimated coefficient β_7, is seen to increase. The highest estimated value is at the 0.95 quantile with a value of 5.056. The lowest estimated value of β_7 is at the 0.05 quantile with a coefficient value of 0.233. As for the nonlinear terms related to the ln(DOC), the trend is opposite to the linear terms. The maximum estimated coefficient in terms of magnitude for β_8 occurred at the 0.95 quantile with a value of -2.223. The smallest estimated β_8 occurred at the 0.05 quantile. As stated earlier, based on research by

[65], the depth of the buried pipe plays a huge role in the amount of obtained %IR signal coming from the cathodic protection (CP) system. If we were to take this into consideration, both the variable β_7 and β_8 should be considered together to obtain a more rounded picture on the effect of DOC on %IR signals. Analysing each variable separately leads to different conclusions.

It is not known either in the industry understanding or in literature that DUC has an effect on the %IR readings. DUC can be the result of the accumulation of corrosion products from the corroding pipe underneath the coating. It can also include other foreign material such as soil which can also be corrosive depending on type and composition. Therefore, leaning towards to the findings for the POPD variable, an increase in corrosion (deeper pits) will produce more corrosion products and therefore an effect on the %IR indication. The effect is not direct but is thought to have some relation to the POPD variable. Further analysis is needed to verify this. However, if this thinking is correct, there is a degree of relationship between DUC and %IR. The highest predicted coefficient was found at the 0.75 (magnitude) quantile with a value of 0.00483. The trend of β_{10} show increasing negativity from the 0.05 quantile up to the 0.75 quantile where it peaks and drops again at the 0.95 quantile. The low values of the estimated coefficients show an effect which is marginal with regards to the %IR signal. A positive effect is observed at the 0.05 quantile with a β_{10} value of 0.00092. On this basis, we can say the effect of small amounts of DUC has a positive effect on the lower readings of %IR. However, after this quantile all the other estimated coefficients are negative which suggests that the amount of DUC has an

inverse effect on the %IR value. If more DUCs are present on a given pipe, the more negative the effect will be (except for a slight dip at the 0.95 quantile) on higher values of the %IR.

The quantiles of the estimated coefficients in Table 6-7 above are given below.



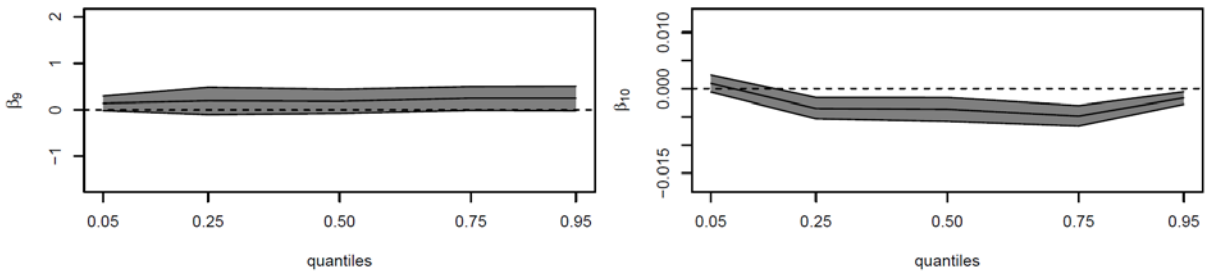


Figure 6-13: Estimated Coefficients across the Quantiles by LQR for β_1 to β_{10} (Combined Model for %IR)

From Figure 6-13, generally it can be said that the estimated coefficients are significant with each of them only occasionally crossing the zero line. This is an exception for coefficients β_7 to β_9 which represents the ln(DOC), where across the quantiles, the term zero is included within the confidence band. The coefficient β_6 is slightly better where only the estimated coefficients at the 0.25 and 0.5 quantiles do include zero within its confidence interval. As was previously stated, although this is the case, the variable remains as one of the more important variable where the exclusion of such variables for determining %IR would be incomplete.

6.4 Logistic Quantile Regression on POPD

POPD is described as the percentage of the depth of the pit over the pipe wall thickness. It is a measure of corrosion on the pipe where higher percentages equate to higher corrosion activity. Corrosion is one of the factors that affects a pipeline's integrity. By controlling this phenomenon, the safe continual operation of the pipe is achieved. Therefore, the

understanding of its contributors in the context of the MEOC data is key to promoting safety. Since the POPD is represented by percentages, it too is a form of outcome which has pre-specified bounds. This makes it possible and ideal to model corrosion in terms of pit depth by using the LQR method. Again, the variables chosen for the model have to be justified from an engineering and scientific perspective based on established literature. Only continuous variables are chosen for assessment and the best ones are included in the final model. The flow of the assessment is similar to the previous section where initially the variables under consideration are set against POPD separately and finally combined to form a model which has the ability to predict the extent of pitting in terms of its depth. Logistic transformation of the POPD variable is given as follows.

$$\boldsymbol{logit(POPD) = \log\left(\frac{POPD + \Delta}{100 - POPD + \Delta}\right)} \qquad \textbf{6.12}$$

where Δ represents a small value, in our case 0.001 to ensure that all possible values of POPD are defined.

6.4.1 TCDA vs POPD

In the previous assessment, both the POPD and the TCDA are assessed separately to see each contribution to %IR. This time, however, TCDA is also assessed separately but with the dependent variable POPD. It is thought that TCDA has a substantial contribution to POPD. When coating breakdown on

a pipe occurs, the metal substrate is exposed to the environment. Although the cathodic protection system is there to protect the pipe from corroding, in the presence of a coating defect, the current supplied by the CP can sometime “wander” off to adjacent pipes. Other factors such as the SR, CP level output and the efficiency of ground beds (the anode in a CP system) can affect the level of current supplied to the pipe. Over time, coating will break down and the possibility of corrosion happening is high if the supplied CP current is insufficient. If the level of current supplied is insufficient and the potential of the pipe becomes less negative, then the pipe is susceptible to corrosion. The level of the pipe’s potential can be inspected by a method call the Close Interval Potential Survey (CIPS) which is similar to a DCVG survey, but the objective of the assessment is different. The CIPS data is not available for the MEOC dataset and the data in the form of the DCVG assessment will have to suffice for this assessment.

Prior to the LQR modelling, the variable TCDA is transformed to be ln(TCDA). This is done to facilitate modelling where transformation will yield better relationship between the two variables. The figure below illustrates this relationship by means of a LQR.

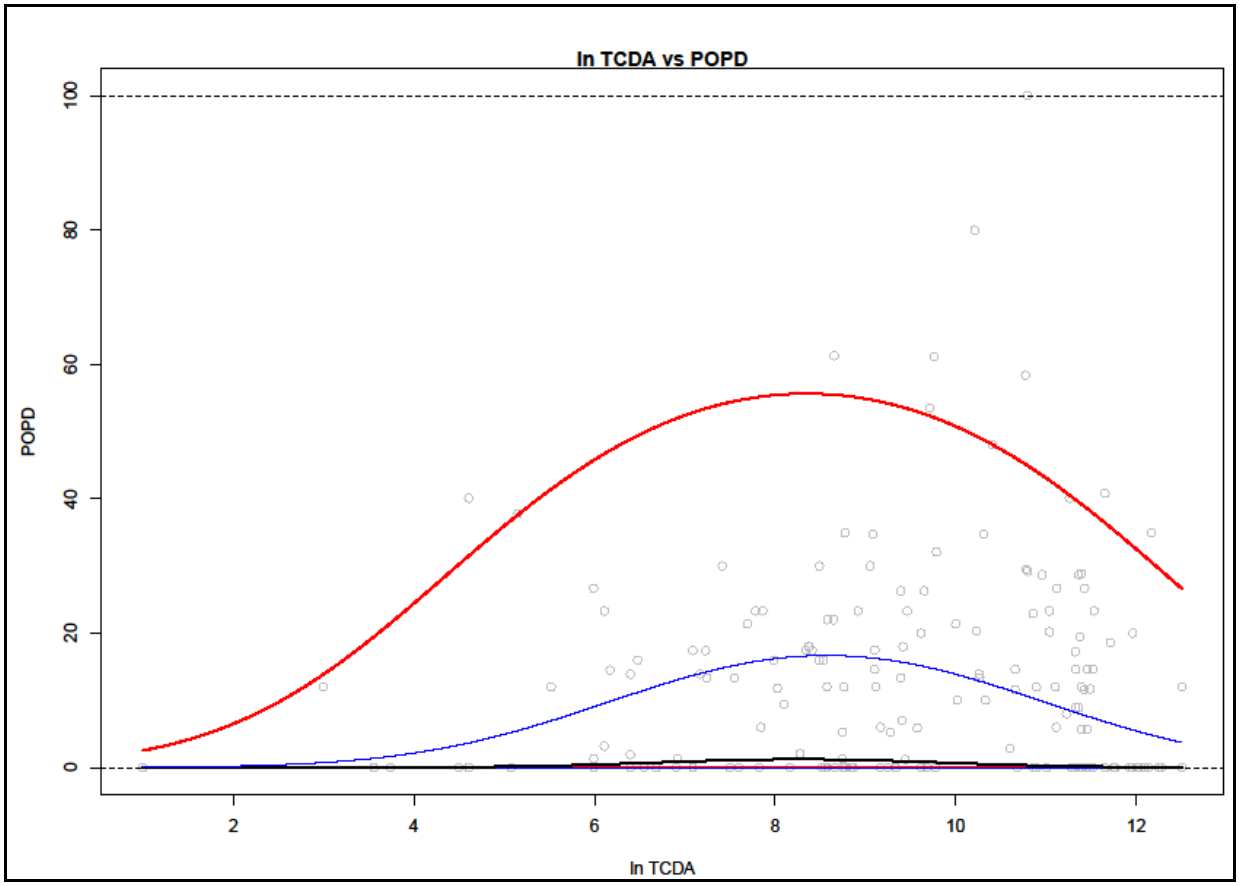


Figure 6-14: LQR on ln(TCDA) vs POPD

The LQR models in Figure 6-14 shows that the relationship between ln(TCDA) and POPD is not linear. Generally, it can be seen that the models are increasing to a peak before dropping back down again. If one were to disregard the quantile lines, a 100% POPD (punch through of the pipe wall) can be seen at around 59,000 cm^2 of TCDA which seems to illustrate the rising value of POPD with respect to increasing values of TCDA. However, more data points are observed beyond this TCDA size which helps prevent the quantile line increasing. The LQR also captures the whole distribution of the POPD by plotting quantiles where an OLS would only measure the mean response, that in this particular case is quite low since the majority of the observation is low. The following LQR models corresponding to Figure 6-14 and its estimated coefficient values are given as follows.

$$\boldsymbol{Q_{logit(POPD)}(\tau) = \beta_1 + \beta_2\, X + \beta_3\, X^2} \qquad \textbf{6.13}$$

where $\boldsymbol{\tau}$ is representing each quantile and ln(TCDA) is termed X. $\boldsymbol{\beta_1, \beta_2\ and\ \beta_3}$ the estimated coefficients of the model represent the $intercept$, ln(TCDA) and (ln(TCDA))^2 respectively. Below, are the estimated coefficient values for the LQR models.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-11.708	-16.127	-18.163	-9.374	-4.772
$\boldsymbol{\beta_2}$	0.061	1.606	3.337	1.812	1.198
$\boldsymbol{\beta_3}$	-0.004	-0.096	-0.202	-0.106	-0.072

Table 6-8: Estimated LQR Coefficient Values for ln(TCDA) vs POPD

The estimated coefficient values in Table 6-8 show interesting results for the β_2 estimation. Starting at the 0.05 quantile, the value steadily rises until the median. After this, the trend starts to decrease all the way to the 0.95 quantile. The maximum value estimated is at the median with a value of 3.337. The lowest estimated coefficient is observed at the 0.05 quantile with

a value of 0.061. For the β_3 coefficient, the negative estimated values across the quantiles complement the β_2 coefficient which represents the downward slope seen in Figure 6-14. The trend highlights that ln(TCDA) has a maximum positive effect at the median depth of the pits. Pits that are deeper than this will have a lesser effect coming from larger ln(TCDA) sizes. The results here show that pitting depth is affected by the size of ln(TCDA) the most when the depth of the pit is at its median and also when ln(TCDA) is also at its median. The following quantile plots show the estimated coefficient values in Table 6-8 with regards to respective quantiles.

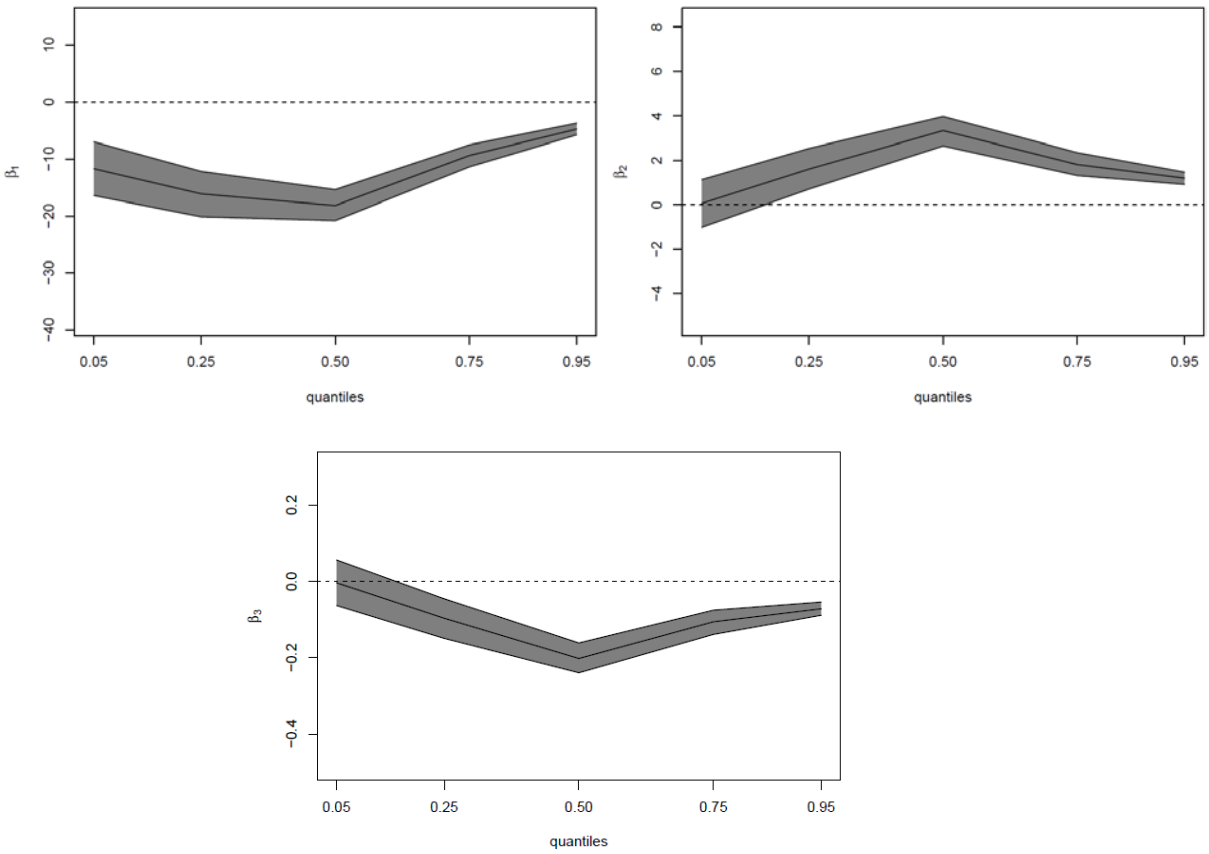


Figure 6-15: Estimated Coefficients across the Quantiles by LQR for β_1, β_2 and β_3 (ln(TCDA))

The quantile plots in Figure 6-15 shows many of the quantile confidence bands do not include the zero term. This is especially true for the intercept. As one can see from the figure above, the trend of β_2 shows a peak at the median which suggests the effect of the size of the median ln(TCDA) on the depth of corrosion pits is greatest when the pit depth is at its median.

6.4.2 SR vs POPD

The variable SR plays an important role in facilitating corrosion of buried pipes. This is supported by works from [57], [87], [101] which state that higher SR will slow the corrosion process. Higher SR has higher electrical resistivity. Since corrosion currents are needed for the corrosion process to happen, higher resistance to the flow of these current will slow the corrosion process. A table developed by [82] has classified the ranges of SR and related it to a range of aggressiveness for the corrosion process. Previous studies have also identified that SR is a major factor in determining corrosion [102]. A CP system's performance can also be affected by the SR. If highly resistant soil surrounds the pipe, the intended CP current supplied to halt corrosion is blocked thus leaving the pipe vulnerable.

The LQR for POPD with respect to SR was conducted to investigate the relationship between the two variables and to see whether it corroborates the statements made above. A transformation of SR into the square root for SR (SQRT(SR)) was done to facilitate modelling. The following models are illustrated below.

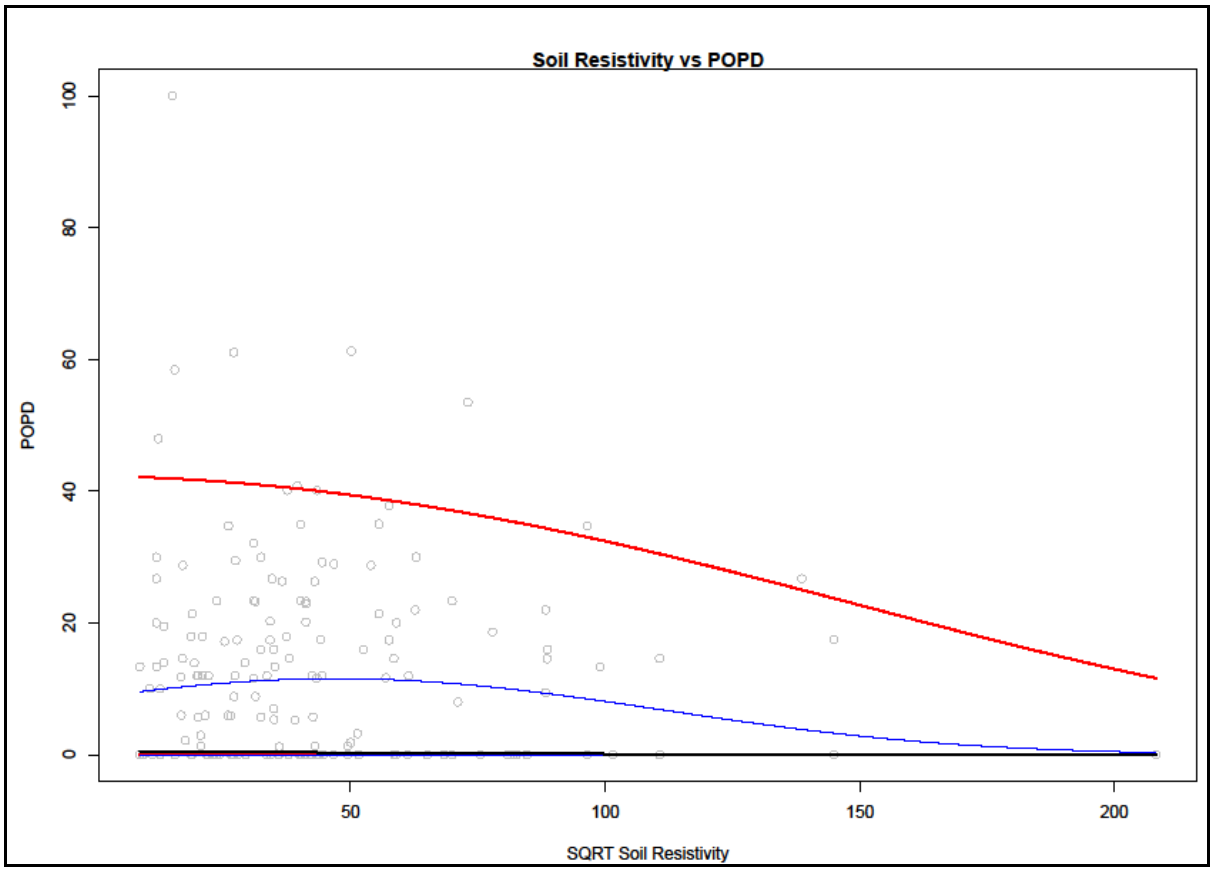


Figure 6-16: LQR on SQRT(SR) vs POPD

The general trend of the models found in Figure 6-16 indicate a decreasing trend. For the median quantile and below, a straight line is observed due to the large amount of “no corrosion” present on the pipes at the selected excavation sites. These quantiles are also bounded by the logit transform of the POPD. Without any transform (e.g. logistic), these models, representing the median and below, would have portions below the zero percent line due to nonlinearity nature of the models. The 0.75 and especially the 0.95 quantiles do seem to support the statements made above (decreasing trend). The logit transformation of the POPD and its corresponding model are given as follows.

$$\boldsymbol{Q_{logit(POPD)}(\tau) = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3} \quad \textbf{6.14}$$

X represents the variable in question SQRT(SR) while β_1 is the model's intercept and β_2, β_3 and β_4 are the model's estimated coefficients. The following table gives the estimated coefficients for the model.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-11.432	-8.982	-3.118	-1.058	1.021
$\boldsymbol{\beta_2}$	-0.002	-0.059	-0.134	-0.070	-0.098
$\boldsymbol{\beta_3}$	0.000	0.001	0.002	0.001	0.002
$\boldsymbol{\beta_4}$	0.000	0.000	0.000	0.000	0.000

Table 6-9: Estimated LQR Coefficient Values for Square Root SR vs POPD

Estimated coefficient values given in Table 6-9 for β_2 show gradual negativity with maximum (magnitude) occurring at the 0.5 quantile at an estimated value of -0.134. At the lowest quantile (0.05) of the β_2 coefficient, the estimated coefficient value is -0.002. The remaining β_3 and β_4 are observed to be zero or close to zero which is represented clearly in Figure 6-16. The trend across the quantile illustrates that the SR's effect on POPD is largest (in terms of magnitude) at the median pit depth compared to shallower pits. The effect is also negative in character which means that higher SQRT(SR) levels have the most negative effect on the median depth of pits. It is

therefore summarised that SQRT(SR) negatively affects the POPD most at its median. Higher SQRT(SR) values will lead to slower corrosion and this effect is greatest when the pit depth is at its median. The following are the quantile plots of the estimated coefficient values from Table 6-9.

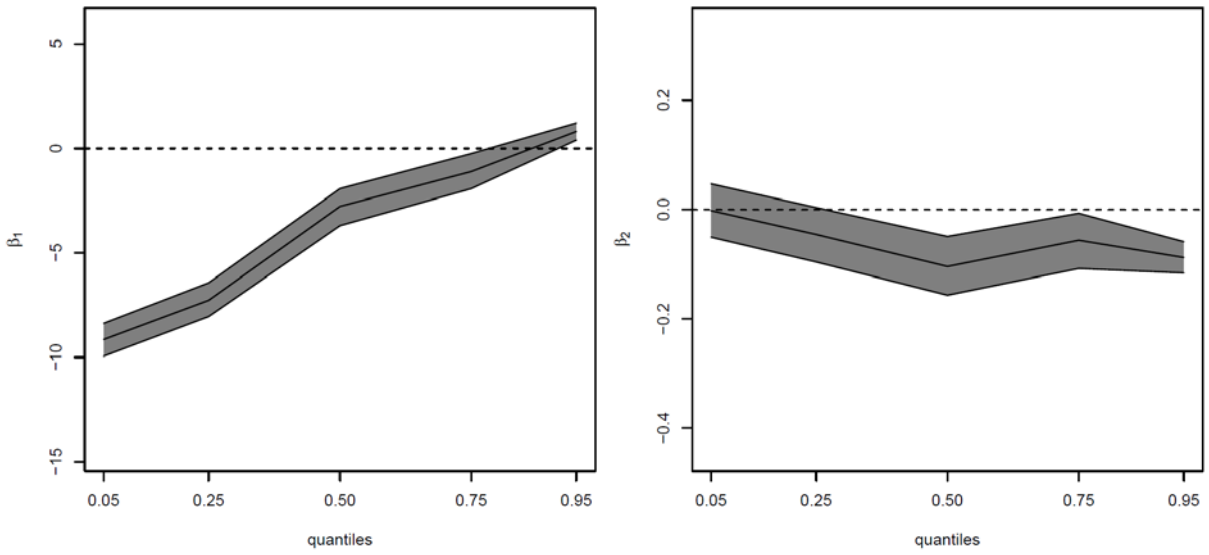


Figure 6-17: Estimated Coefficients across the Quantiles by LQR for β_1 and β_2 (SQRT SR)

In Figure 6-17 most of the estimated coefficient values for β_2 are significant, with the exception of the estimated value at the 0.05 quantile where the zero term is captured within the confidence band. The trend of β_2 echoes the statements made previously where the upper quantiles of a POPD (median to 0.95) shows the biggest negative effect coming from the SQRT(SR) factor. Figure 6-17 further clarifies this by illustrating the negative effect of SQRT(SR) on corrosion.

6.4.3 DUC vs POPD

The DUC variable is thought of as the result of the corrosion process where the corrosion products are accumulated underneath the coating. If this is the case, then a positive relationship should occur between the two variables. DUC can also be made up of foreign materials coming from the surrounding environment such as sand and stones. For this assessment the variable DUC was not transformed as the data is sufficient (in terms of trend) for modelling. The DUC variable through its estimated coefficients will later be judged for its inclusion in the final POPD combined model, based on magnitude and engineering relevance. The constructed models for DUC and POPD are given below.

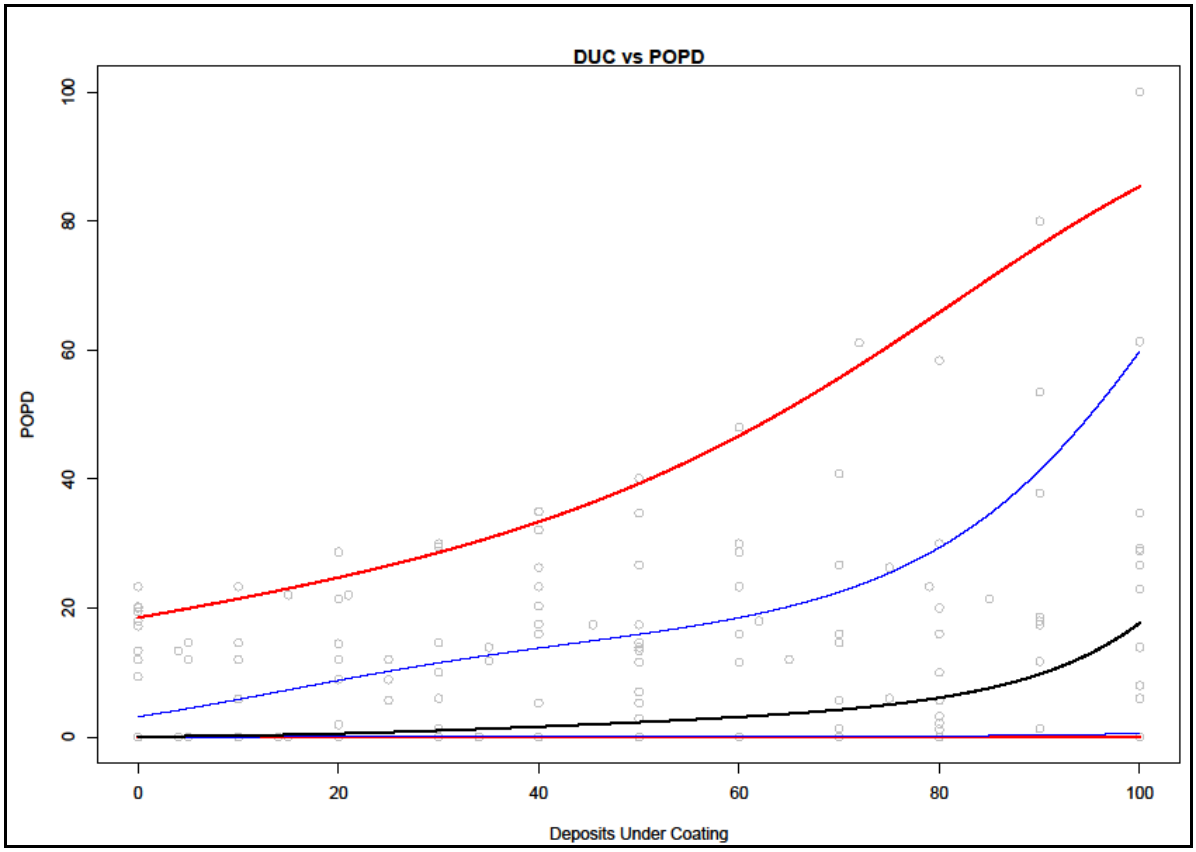


Figure 6-18: LQR on DUC vs POPD

Figure 6-18 points to an increasing trend between DUC and POPD. It can be seen that as DUC increases, the variable POPD also increases. This somewhat supports the idea that the two variables are indeed connected. The relationship observed is nonlinear where the higher quantiles show a steeper slope compared to other quantiles. It can also be observed that for quantiles 0.05 and 0.25 the models look flat. This is mainly due to no corrosion activity occurring at the selected site of excavation. Interestingly, there are cases where no corrosion is found at higher values of DUC. This as earlier stated is due to foreign material getting stuck and accumulating underneath the coating (not corrosion products). The models in Figure 6-18 also illustrate the superiority of the LQR method where it restricts the bounds

between 0 and 100 and also characterises the whole of the POPD distribution with respect to every increasing DUC value. The following is the logit-transformed POPD variable after which is presented a table that highlights the estimated coefficients for the LQR models in Figure 6-18.

$$\boldsymbol{Q_{logit(POPD)}(\tau) = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3} \qquad \textbf{6.15}$$

The τ term above represents the quantile in question and X represents the variable DUC while β_1 is the model's intercept and β_2, β_3 and β_4 are the model's estimated coefficients DUC, $(DUC)^2$ and $(DUC)^3$.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-11.503	-11.135	-8.917	-3.937	-1.505
$\boldsymbol{\beta_2}$	0.001	0.036	0.169	0.101	0.019
$\boldsymbol{\beta_3}$	0.000	0.000	-0.002	-0.002	0.000
$\boldsymbol{\beta_4}$	0.000	0.000	0.000	0.000	0.000

Table 6-10: Estimated LQR Coefficient Values for DUC vs POPD

Table 6-10 shows the estimated coefficient β_2 (DUC) with an increasing trend as one scans through the quantiles and decreasing down to 0.019 at the 0.95 quantile. The highest estimated value is at the 0.5 quantile whereas the lowest estimated value is at the 0.05 quantile. Based on Table 6-10, the DUC has the largest effect on the median pit depth where its effect is lowest

for shallower pits. The figure given below is the quantile plots illustrating the estimated coefficients found in Table 6-10.

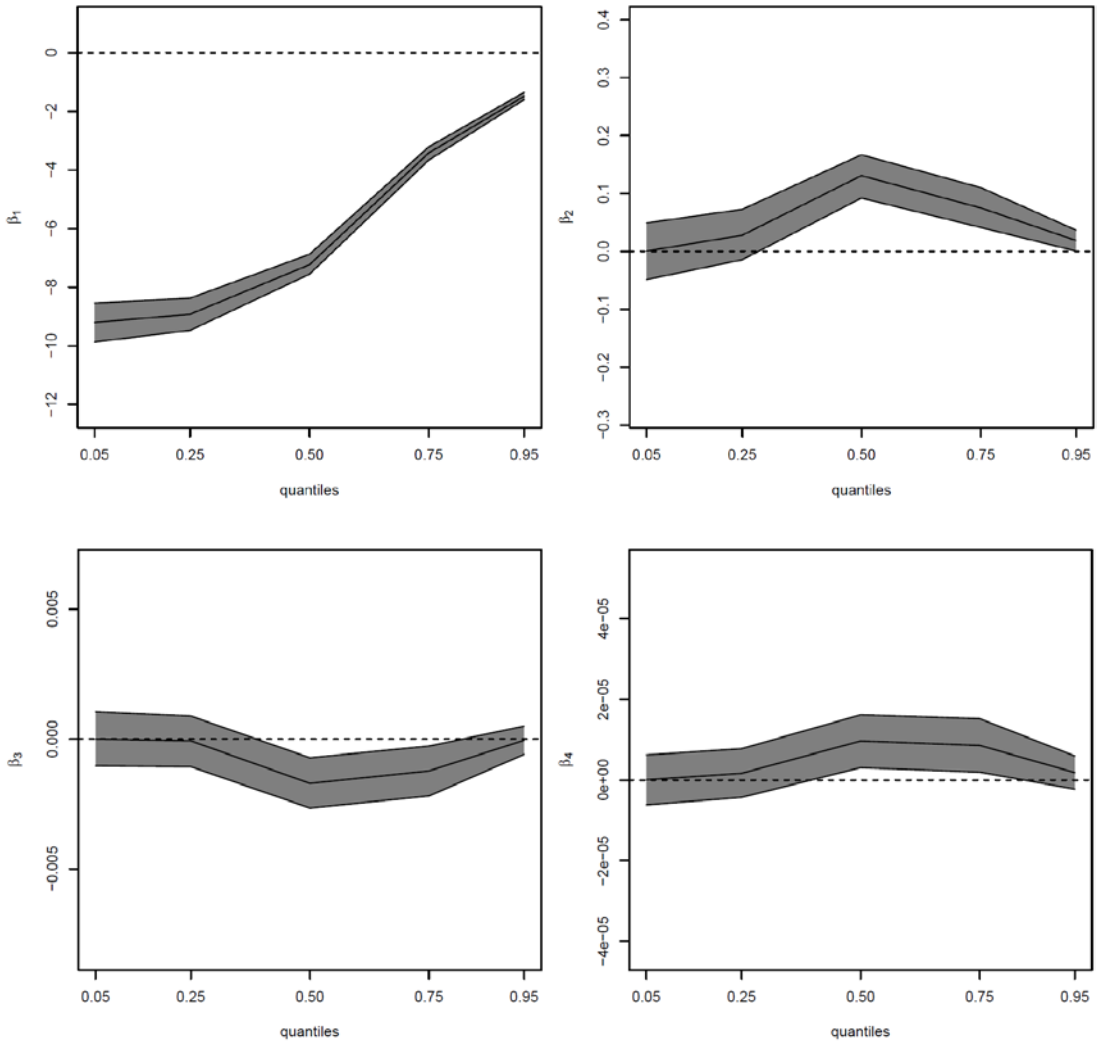


Figure 6-19: Estimated Coefficients across the Quantiles by LQR for β_1, β_2, β_3 and β_4 and β_2 (DUC)

Similar to the results obtained in the previous assessment, all the estimated values across the quantiles for β_2 are significant except for the estimated value at the 0.05 quantile. The trend portrayed here says a lot in terms of the linear part of the LQR models where the effect of the DUC on POPD is strongest at its median depth. Based on this result, it seems that the DUC

does have an effect on POPD which parallels the theory that some of the DUCs are the result of corrosion products trapped underneath the pipeline coating.

6.4.4 Combined Model for POPD

A model which combines all the variables under consideration is developed for the purpose of predicting the extent of corrosion through the depth of pitting. The criteria for the selection of such variables remains the same as the previous combined model for %IR. Engineering and scientific justifications are needed for selection with the additional consideration of statistical outcomes. The magnitude of each variable is observed and the ones having an effect on POPD is chosen. All other variables which have no effect on the response variable are discarded. The following is the LQR model developed based on the selected variables.

$$Q_{logit(POPD)}(\tau) = \beta_1 + \beta_2\, lTC + \beta_3\, lTC^2 + \beta_4\, SRSR + \beta_5 SRSR^2 + \beta_6 \delta + \beta_7 \delta^2 \quad \textbf{6.16}$$

where:

lTC = ln(TCDA)

$SRSR$ = Square Root SR

δ = DUC

τ = Representing the quantile in question

From an engineering standpoint, all the variables considered should be selected for the combined model since all of them influence the formation of corrosion either directly or indirectly. However, only certain nonlinear terms were considered for selection. No cubic terms from the separate assessments are included due to the majority of the estimated coefficients being zero or close to zero.

The variables ln(TCDA) and SQRT(SR) are considered important factors in facilitating corrosion. In the presence of a good coating system where the coating isolates the metal substrate from the environment, corrosion is expected to be zero. Modern coating such as the 3-layered system which includes fusion bonded epoxy (FBE) will maintain the pipe from corroding if the application of the coating is done correctly. This newer coating system is not found in the MEOC data. However, if the coating does break down and the level of CP current applied to the pipe is insufficient, then corrosion is expected. The size of these coating failures does have an effect on the depth of the pitting. This is observed in Chapter 5 (although the analysis conducted there is the effect of POPD on TCDA) where the disparity between the rate of pitting and coating defect growth occurs at the median region. No effect of POPD (towards TCDA) is seen within this area. SR has the effect of influencing corrosion by restricting or allowing the flow of corrosion currents. This combined with the TCDA variable will yield different results based on the different mix of the two variables. For example, a situation where a large

coating defect is present does not guarantee corrosion if the pipe is surrounded by resistant soil (other factors will also contribute to this equation e.g. the CP current being applied to the pipe – this data is not provided in the MEOC data set). The DUC however, is thought to be the by-product of corrosion and other foreign material which does not contribute to corrosion directly. The estimated coefficients for the model above are given in the following table.

	Quantiles				
	0.05	**0.25**	**0.5**	**0.75**	**0.95**
$\boldsymbol{\beta_1}$	-11.757	-16.247	-18.961	-10.587	-3.032
$\boldsymbol{\beta_2}$	0.074	1.487	2.731	1.627	0.264
$\boldsymbol{\beta_3}$	-0.004	-0.084	-0.159	-0.093	-0.013
$\boldsymbol{\beta_4}$	-0.002	-0.027	-0.004	0.015	0.012
$\boldsymbol{\beta_5}$	0.000	0.000	0.000	0.000	0.000
$\boldsymbol{\beta_6}$	0.001	0.025	0.095	0.041	0.013
$\boldsymbol{\beta_7}$	0.000	0.000	0.000	0.000	0.000

Table 6-11: Estimated Coefficients across the Quantiles by LQR for $\boldsymbol{\beta_1}$ to $\boldsymbol{\beta_7}$ (Combined Model for POPD)

In Table 6-11, most of the estimated coefficients for the nonlinear terms in the model are found to be zero with the exception of the $(\ln(TCDA))^2$ variable. The estimated coefficients for ln(TCDA) across the quantile are found to peak at the median where its effect is greatest. The estimated coefficient value at this quantile is 2.731. The lowest estimated value for ln(TCDA) occurs at the 0.05 quantile with a value of 0.074. The nonlinear

term of the ln(TCDA) variable, β_3 also shows similar trend albeit with a negative effect.

For the variable SQRT(SR), the trend of the estimated coefficient for the linear part of the model is not straightforward. The maximum estimated coefficient value in terms of its magnitude is at the 0.25 quantile with an estimated value of 0.027.

Across the quantiles for the variable DUC, it is observed that the trend is similar to the ln(TCDA) variable where the highest estimated coefficient value occurs at the median. The value is recorded to be 0.095. The minimum captured value is at the 0.05 quantile with a coefficient value of 0.001. Below are the quantile plots for the estimated coefficient value based on the findings in Table 6-11.

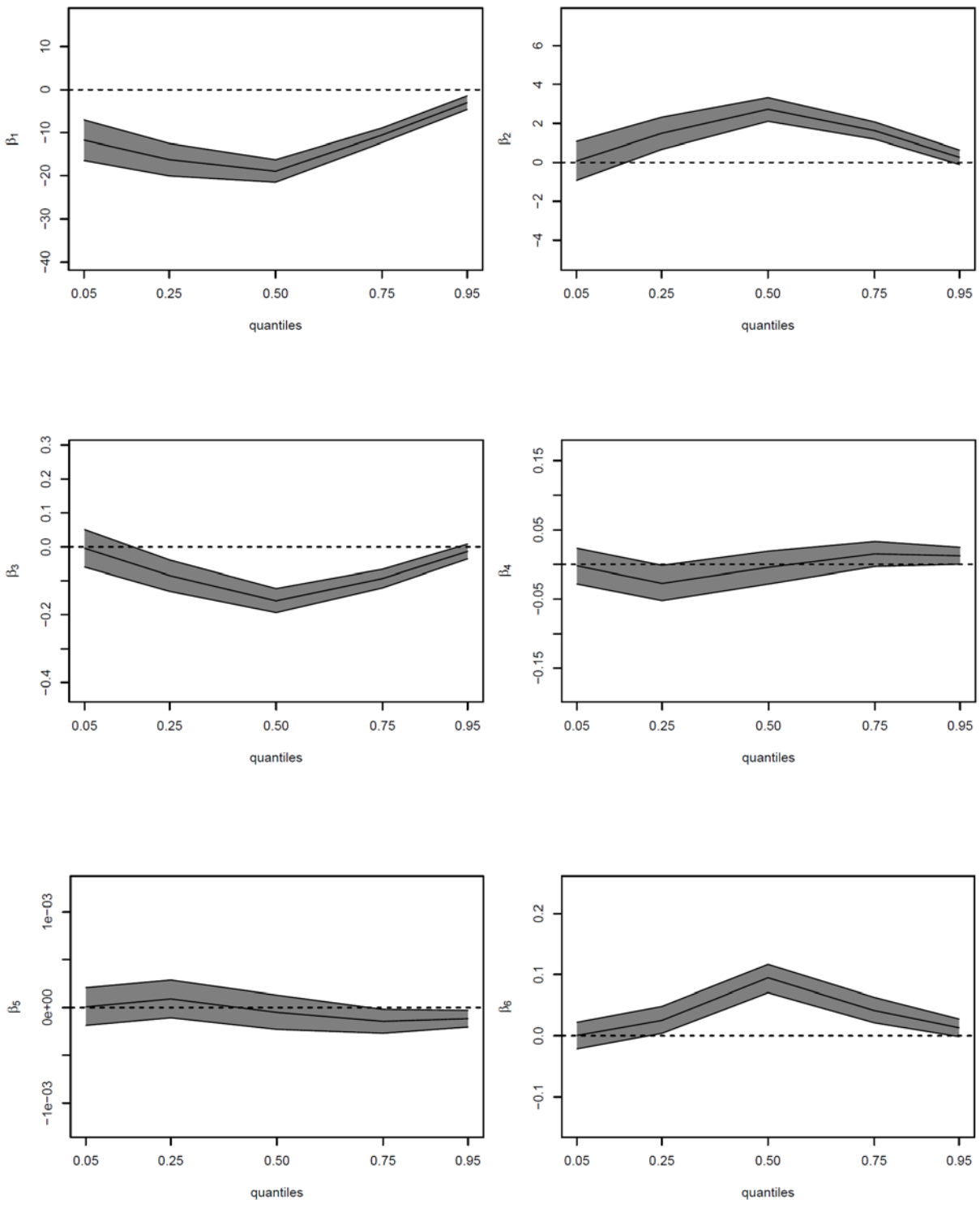


Figure 6-20: Estimated Coefficients across the Quantiles by LQR for β_1 to β_6 (Combined Model for POPD)

The confidence band of each estimated coefficient value for the combined model for POPD is highlighted in Figure 6-20. Most of the estimated coefficient values and the associated confidence interval do not contain the zero term. However, for the variable SQRT(SR) almost all the estimated values for each of the quantiles do contain the zero term within its confidence interval. As for the variable ln(TCDA) and DUC only the estimated coefficients predicted at the 0.05 and 0.95 quantiles are insignificant. Furthermore, these two variables exhibit a similar trend of the estimated coefficients. Both of the estimations peak at the median which can mean the greatest effect by the ln(TCDA) variable and the DUC variable occurs at the median pit depth. Before and after this peak, the effect is not as strong. This can be explained by the fact that the median pit depth is where the corrosion is most critical. At the start of the corrosion process, the size of the ln(TCDA) is not a major factor for corrosion to initiate. Corrosion does not rely on the size of the defect for it to happen but rather the chemistry that reacts within a corrosion cell. This can be seen in the lower estimated coefficient values for ln(TCDA) at the shallower pits (lower quantiles). For larger ln(TCDA) where more metal substrate is exposed to the environment, its effect is not strong for deeper pits. Once a corrosion pit reaches this stage, the influence of large sized coating defects is small due to the pit growth being independent from it. This can also be seen at the higher quantiles of the POPD where estimated coefficients for ln(TCDA) are relatively low. The coating defect has most influence at the median depth suggesting that beyond this point, the influence of coating defect size towards pit depth diminishes. Additionally, in Chapter 4, Figure 4-2, it can be seen that clusters of data points are

observed to be below the mid-way point (the median of TCDA is lower than this value) of TCDA. A through-wall pit is also observed within the proximity of the median sized TCDA. What this means is that most of the deeper pits occur at smaller to medium sized TCDA. This finding leads the Author to speculate that at some point in time, the growth of TCDA has slowed down (when compared to POPD) or stopped completely whereas pitting continues to grow at a much faster rate leading to the existence of deeper pits occurring at smaller to medium sized TCDA.

6.5 Discussion

The analysis started with the usage of a logistic regression for the determination of the probability of excavation based on the %IR readings. The reason for this exercise is to obtain a guideline for pipeline operators which can be used as an additional indicator for the decision-making process of where to excavate. It is discussed thoroughly in previous chapters that determining the location of excavation sites is a challenge since the indications available to the assessor are not totally reliable. For the analysis, two new variables were generated based on the Author's categorisation of justified and unjustified excavations where it takes a binary form. These variables are not fixed and can later be varied depending on the level of conservatism that a particular project requires. The analysis shows that the number of unjustified excavations for the MEOC project according to the $TExx$ variable is quite substantial. This illustrates the unreliability of the current method as practised by the pipeline industry. The major part of the

method requires expert engineering judgment which can be erroneous if not properly applied. The application of this model should be treated with care. Since the models were constructed with the MEOC dataset, the Author is inclined to suggest the models are only to be applied to the MEOC pipelines. To make it applicable to other pipelines, new ECDA data from other assessment projects is needed to combine it with the data we have here to make the model more generalised.

Following the logistic regression, the LQR was conducted by taking the %IR as the dependent variable. The %IR is a suitable candidate for LQR since the boundary is fixed at 0 to 100%. Additionally, %IR serves as the first indicator that an inspector would obtain from the DCVG assessment. By having a practical model, within the specified range, an inspector could use the model as a supplementary guideline to interpret what the measurements of a DCVG indication are signalling. Thus, a more reliable decision-making process for further assessments is enabled.

In Chapter 5, attempts were made to model the relationship of various variables with %IR by using Bayesian quantile regression to estimate the coefficient values of each variable. Four models were generated based on the Oriset and the Filtset data where the latter had 4 outliers removed from the dataset. The approach also considered the relationship between the TCDA and the %IR to be linear. The resulting models after the refinement and the omission of the outliers produce good predictions of %IR based on the TCDA values. Model 2a highlighted models which are acceptable based

on industry common understanding regarding the DCVG technique. However, due to the linear approach taken in that model, predictions outside of the %IR boundaries were observed. The work in this Chapter aims at resolving this by the usage of the LQR to ensure the predicted values are within the specified bounds. The determination of the variables for the combined model of the %IR is selected differently from Chapter 5. Each variable is scrutinized and was set against the %IR separately to see each underlying relationship. Based on the coefficient's magnitude and the variable's relevance to the dependent variable, independent variables are chosen for the inclusion within the final combined model. Additionally, categorical variables which are subjective in nature are not included in the final model to avoid any subjectivity.

The ln(TCDA) variable shows an increasing effect across the quantiles with respect to %IR. This echoes well with current industry understanding of the DCVG principle. To gain access to the full extent of TCDA's effect on %IR, a plot of TCDA *vs* %IR based on the combined %IR model is given as follows. The resulting models in Figure 6-21 are the derivation of %IR with respect to TCDA and treating all other variables in the model as constants.

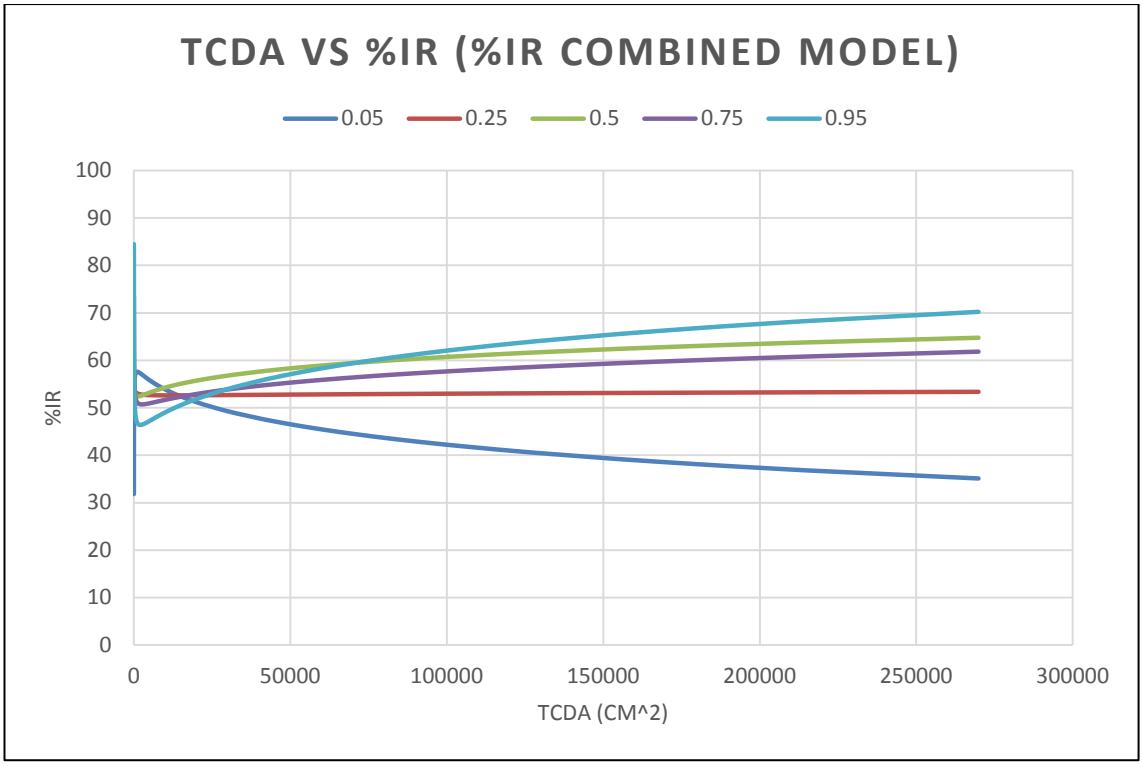


Figure 6-21: TCDA vs %IR for the %IR Combined Model

It can be seen in Figure 6-21 that the predicted %IR values based on the TCDA are well within the %IR bounds. Since Figure 6-21 uses the original dataset, TCDA sizes of below 17,800 cm^2 see an unusual trend which can be stipulated as the effect of the inconsistencies highlighted in Chapter 5 (presence of higher TCDAs at lower readings of %IR). This portion of the dataset was removed in the Filtset data for the analyses of Models 2 and 2a in Chapter 5. Above this size, a stable trend emerges where the highest predicted %IR value occurs at the 0.95 quantile. From an engineering perspective, this is satisfactory. The median quantile shows a somewhat surprising trend where the predicted values of %IR are higher than the 0.75 quantile. However, when considering Models 1 and 1a (using original dataset) in Chapter 5, a similar trend emerges where the median %IR is

shown to have the most effect from the TCDA. Due to the non-linear approach (through modelling and the logistic transform) taken in this Chapter, the resulting estimation of %IR in Figure 6-21 shows small differences between the quantiles (compared to Models 1 and 1a in Chapter 5) where the 0.95 quantile of the %IR received the greatest effect from the TCDA. The 0.25 quantile in Figure 6-21 also shows a flat response indicating marginal effects are taking place between the TCDA and %IR. This is similar to the flat response we observed in Models 1 and 1a in Chapter 5. At the 0.05 quantile, there seems to be a downward trend where increasing TCDA results in a decreasing %IR reading. This is also similar to the downward trend of the 0.05 quantile for Model 1a of Chapter 5. This might be due to the "inconsistencies" found in Chapter 5 where higher TCDAs are related to lower %IR readings and *vice versa*. Overall, the trend predictions made here are similar to the ones obtained in Chapter 5 (for Models 1 and 1a) due to the same dataset being used for the analyses. Based on these models, the general application to other pipelines are rather limited due to the observations of the models being too dependent on the "noise" within the original data set. As was stated earlier Model 2a (Chapter 5) appeals to a more general application where the "noise" was removed for more general prediction applicability. It is advisable to keep in mind that this "noise" can be the result of a unique situation of the environment which is not encountered elsewhere. Unique situations can include the presence of stray currents constantly provided by adjacent CP systems or the nature of the surrounding soils which yields the %IR reading in a certain way. Thus, if one were to conduct assessments based on the same MEOC pipelines in future

iterations, the models generated in this Chapter would be sufficient for the prediction of %IR based on the TCDA. The 0.95 quantile seems to be the best candidate for %IR prediction while quantiles 0.25 and below indicates a less confident approximation (based on downward trend). This may be due to external environmental forces at play where it influences the outcome of the model. As was stated before, application to other pipelines is not recommended due to the “uniqueness” of the MEOC pipelines and its environment (dataset).

Recent literature, notably [64], stated that other factors such as the SR, has a positive influence on %IR signal. However, the results shown in the work presented here seems to suggest otherwise. The trend of the estimated SQRT(SR) coefficient does show an increase as one scans across the quantiles with all the estimated values being negative. This can generally be viewed as the effect of SQRT(SR) on %IR being negative. The apparent contradiction is probably due to the different approach taken for each study. In Chapter 5, similar trends can be seen where highly resistant soil gives a negative effect on the %IR. However, the negative trend observed in that model is complemented by another soil variable (also highly resistant) which shows an opposite trend. Since the model constructed here does not consider categorical variables, it is assumed that if there were a variable of that sort, the same scenario would also apply for this model. Furthermore, the work done here uses datasets that comes from real life projects and employs the LQR method for the construction of the models whereas the

work conducted by McKinney relies on simulations of data and is based on the Finite Element Method.

An interesting finding was the effect of variable POPD on %IR. The results obtained seem to suggest that corrosion has a positive effect on the %IR signal. This idea was initially suspected in Chapter 4 where corrosion variables such as the total corrosion area and volume demonstrated a marginal correlation with %IR. In the LQR analysis, the trend across the quantiles for the combined %IR model, shows increments of the estimated coefficients which points to a positive effect. However, if one looks at the quantile plot for the POPD coefficient, only the median and higher quantiles illustrate statistical significance. Nevertheless, the remaining significant estimates show increasing trends which supports our initial assumption that corrosion has a positive effect on the %IR signal.

The variable DOC generally shows a negative effect on %IR. The linear part of the constructed combined %IR model shows a positive trend in terms of the estimated coefficients. From this, it is suggested that the pipe depth has an increasing effect on %IR. However, after observing the LQR models, the non-linear quadratic terms show negative effects on %IR. For clarification of the variable's opposing trend, a derivation of %IR with respect to the DOC was carried out. This is illustrated in Figure 6-22 below.

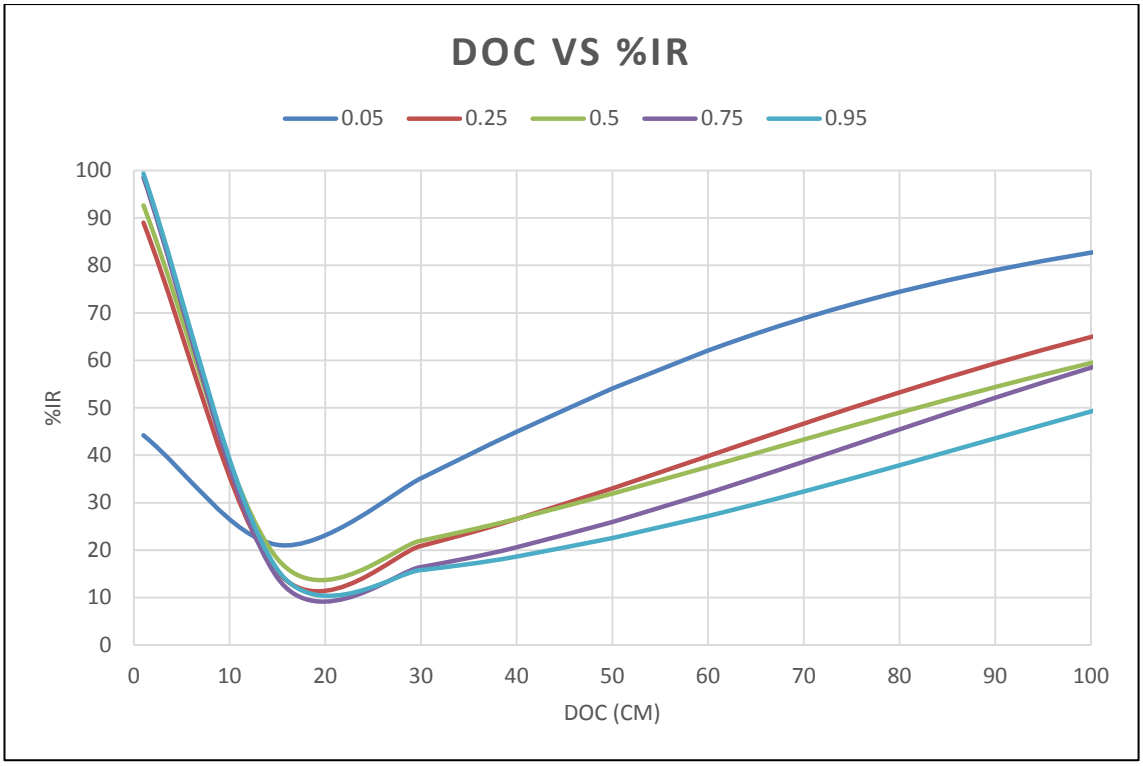


Figure 6-22: DOC vs %IR for the %IR Combined Model

Taking both observations together, it can be summarised that generally, the relationship between DOC and %IR is negative which suggests that increasing pipe depth will result in a weaker %IR signal only up to a depth of approximately 20 cm. The finding before this point is corroborated by papers such as [65] and also the industry's understanding of the system. It can be said that the models are only applicable up to this depth. If one were to go beyond this point, the effect of the DOC is reversed where deeper depths will give higher readings of %IR. As discussed earlier this trend is unexpected. One possible explanation can be directed towards the interference of stray currents. At deeper depths, the interference is more pronounced since the soils are much more compacted making it less resistant to current flow. As such, the deeper a pipe is, the more vulnerable it is to interference. Another

possible reason for this is the attachment of steel rods (for previous sacrificial anodes) which give off current from the CP system. The currents that leave through the steel rod(s) can interfere with the equipotential lines generated by the defect. At shallower depths, the influence of the rod is less apparent due to the lower exposed area of the metal with the surrounding soil.

Corrosion in the form of the POPD variable was also considered as the dependent variable for the LQR method. This is due to its values which are bounded by 0 to 100%. The models generated from this analysis can be of use when an assessor is trying to predict the amount of corrosion based on data available such as the coating defect area, SR and the amount of deposits found underneath the coating. The predicted corrosion based on the coating defect size is given in Figure 6-23 below by deriving the POPD combined model with respect to ln(TCDA). The ln(TCDA) was converted back to its original form prior to the derivation.

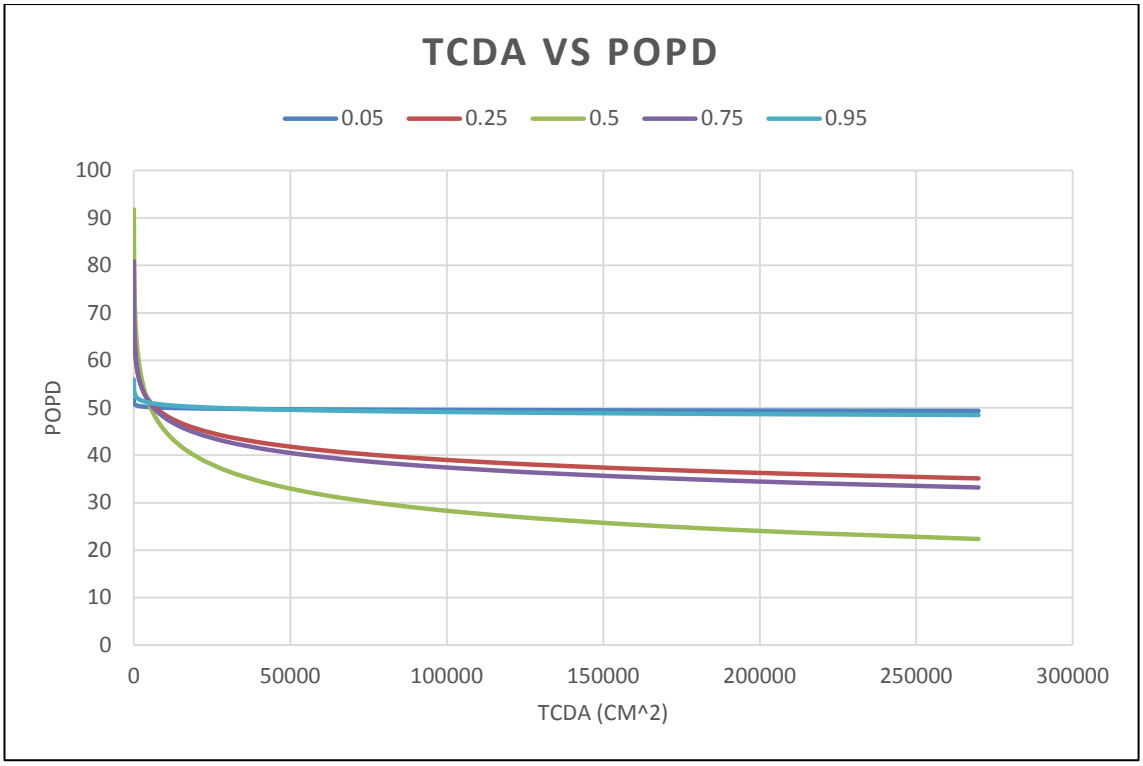


Figure 6-23: TCDA vs POPD for the POPD Combined Model

It is shown that the size of TCDA has the strongest effect on corrosion when the pits depths are at their median. (This is illustrated in Figure 6-23). The median quantile of the POPD shows that it is affected more by TCDA than any other quantiles. Other extreme quantiles such as the 0.05 and the 0.95 quantiles show little effect coming from the TCDA on the shallowest or the deepest of corrosion pits. The start and end (close to punch through) of the pitting process require lesser effect from the size of TCDA to grow. The most important phase of the life of pits is when the depth is at its median as this is the start of when the rate of corrosion is higher than the rate of growth for the TCDA. It is therefore suggested that repairs on coating defects should be made before it reaches this stage where the corrosion process is most critical. Therefore, by repairing coating defects before it reaches its median

size will have the benefit of promoting the pipe's integrity. However, this recommendation is only applicable to the MEOC data since the median coating defect size is already known. To generalise the model further, more data from other ECDA projects are needed. It is tempting to say that the median depth is synonymous to the pipe wall's half-thickness, but this is not always the case.

It can be seen that the SR factor does contribute to corrosion (POPD) but in a negative way. The extent of the truth of this statement can be demonstrated by the derivation of the POPD combined model with respect to SR. The resulting models are given as follows.

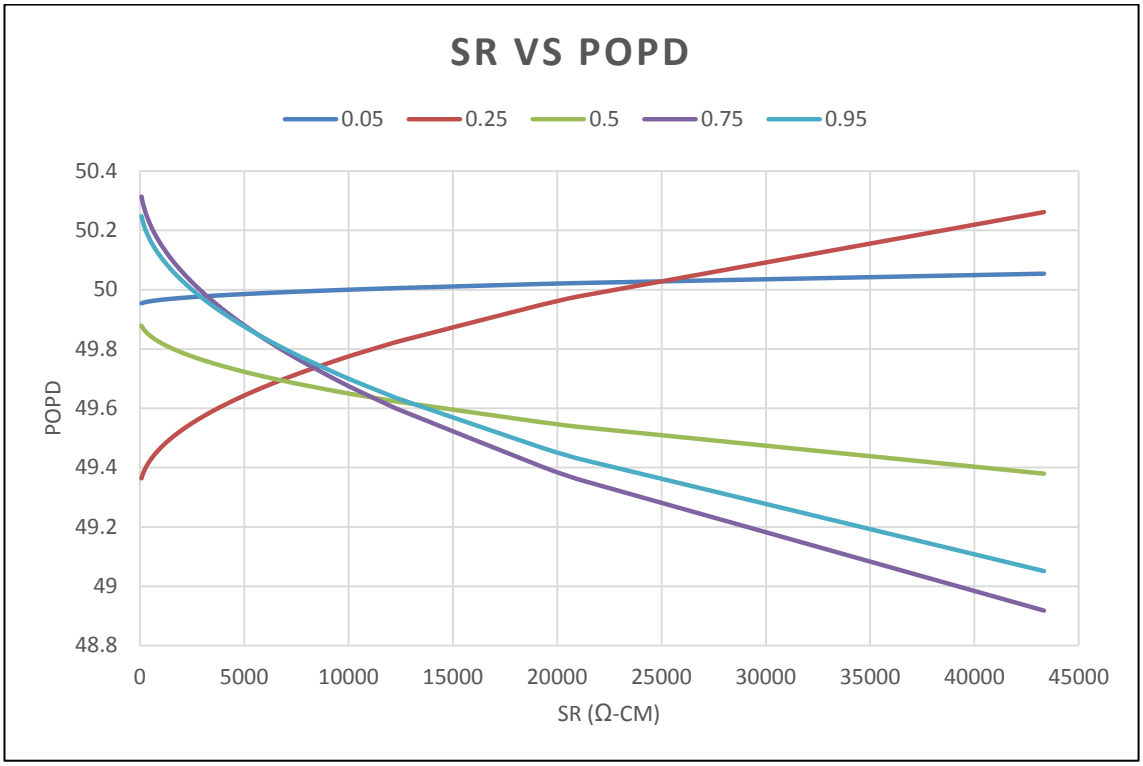


Figure 6-24: SR vs POPD for the POPD Combined Model

It was long thought that SR is an indicator of corrosion (see [57], [87], [101], [102]). Based on the models generated in Figure 6-24, this is only partly true. The upper quantiles of the POPD (0.5 and above) show a decreasing trend which is supported by the literature. However, for shallower depths of the POPD, low SR levels effects corrosion in a reversed manner where increasing SR results in corrosion. The reason for this reverse phenomenon can be explained by the shielding effect of the soil itself. From the trend above, it can be said that increased SR affects the corrosion of shallower pits by preventing the CP current from reaching the metal substrate. Therefore, corrosion activity is promoted. At greater pit depths (after the initial phase of shielding) the corrosion is most active (based on the previous assessment on TCDA) and will need the soil to facilitate the process of corrosion by transferring currents in and out of the corrosion cell. At shallower depths (initial corrosion) the transfers of such corrosion currents are less thus negating the strong SR effect. The parallel in the findings from literature with regard to the upper quantiles of the POPD suggest that the constructed model is sufficient for use in future corrosion assessments of the MEOC pipelines.

The results of the analysis on DUC however, illustrate that its amount depends on the corrosion activity. To see this more clearly, the derivation of the POPD combined model was based on the DUC. The following Figure 6-25 illustrates this relationship.

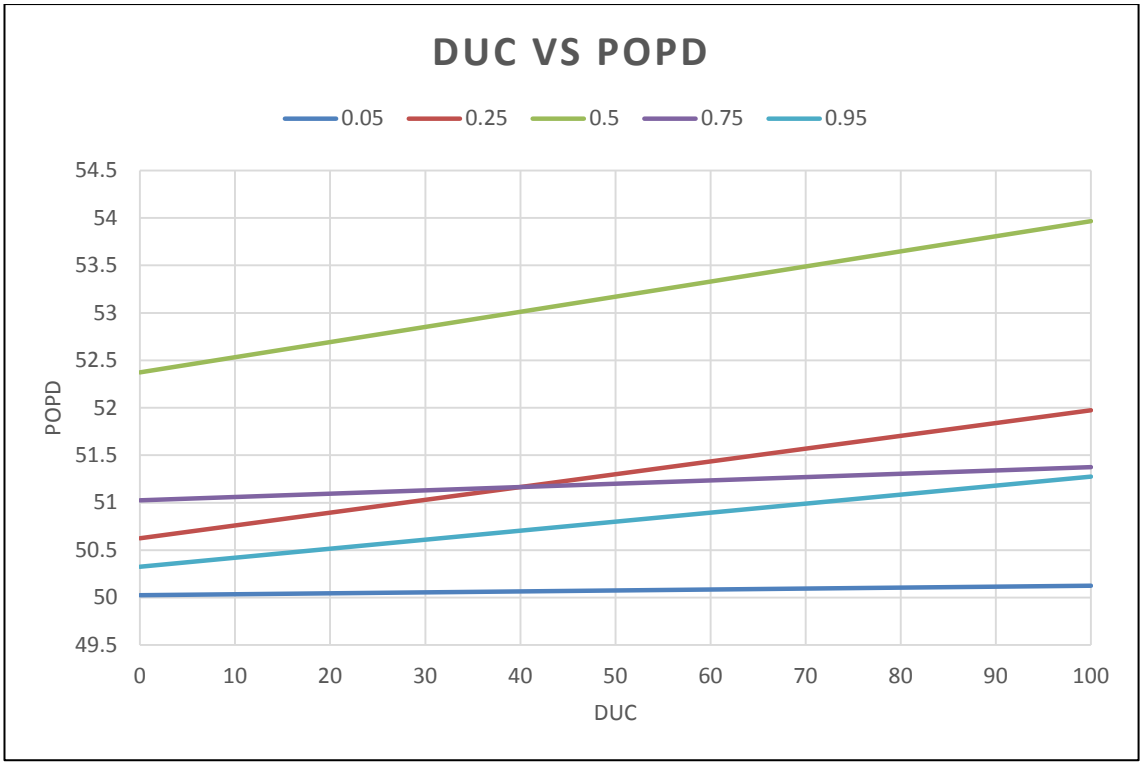


Figure 6-25: DUC vs POPD for the POPD Combined Model

The corrosion activity is represented as the POPD where greater depths translate as more corrosion products. Over time, accumulation of these products will form part of the deposits found underneath the coating. This assumption is in agreement with the trend illustrated in Figure 6-25 above. However, the median POPD quantile is the most affected by the DUC. Since it is already known that at the median pit depths corrosion is the most critical corrosion products are also expected to be high during this phase, hence the strong effect. DUC is not a direct measurement of corrosion since other foreign materials can also be present in the deposits. This is explained by the distribution of the POPD variable where the variation tends to increase with increasing DUC. At higher POPD represented by the upper quantiles (0.75 and 0.95), the DUC is made up of corrosion products plus other foreign

material. The relationship between the two variables is thus not proportional. Since this is the case, higher DUC does not translate into a direct amount of POPD hence the lesser effect being observed. At lower POPD represented by the lower quantiles (0.05 and 0.25) it is suggested that the composition of the DUC is mostly the product of the resulting corrosion. Since the POPD is shallower and the amount of DUC is also less, the effect of the DUC on the POPD is also less. Another possible explanation is in the initial phase of a corrosion pit where the corrosion products tend to accumulate within the pits themselves and not spread out underneath the coating. If this is the case, very little DUC is present and will have very small effect on the POPD. The almost flat line for the 0.05 quantile in Figure 6-25 illustrates this by highlighting very small effect of DUC on POPD.

LQR has provided a meaningful method in the inference of the bounded dependent variable's distribution with respect to each independent variable. In the case of the MEOC data, bounded outcomes such as %IR and POPD were inferred practically as the method ensures the interpretation is within a specified boundary. On top of the logistical transformation of the dependent variable, LQR employs the quantile regression method where it characterizes the dependent variable's distribution much more thoroughly than the traditional central tendency methods where typically the mean or median are used for inference.

Chapter 7

Conclusion and Future Work

Based on the MEOC ECDA project, the reliance on the DCVG technique alone to identify coating defects and quantify severity seems to be insufficient. The significant level of subjectivity that is applied to the determination of excavation sites is rather misleading. This is illustrated by the number of unjustified excavations resulting in finding no coating defects within the bell hole. To add to this uncertainty, the so called “undiscovered” defects which might or might not be present elsewhere on the pipe are also a prime concern since the current method for interpreting DCVG data lacks accuracy. The work presented in this thesis sets out to improve these deficiencies by developing statistical models which improve interpretation of

the ECDA process. Two properties of a pipeline which require attention in terms of repairs and ensuring the safe continual operation of the pipe are the coating defect size, TCDA, and the corrosion that the pipe is currently undergoing, represented by the variable POPD. These two variables are the prime indicator as to which this thesis focuses on.

The relationship established between %IR and the coating defect size (TCDA) showed low linear correlation values signalling a relationship which is complex. A nonlinearity approach was taken to address this, and a novel technique was used to investigate the range of defects across the coating size spectrum. Two models were constructed in the form of the Non-Interaction model and an Interaction model which included interaction variables as part of the model structure. The results found that the interaction variables play an important role in the prediction of the size of coating defects. The interaction model produces predictions which mimic the boundary of the range of the size of coating defects within the MEOC data, whereas the Non-Interaction model predicts lower values. The lower quantiles in both models indicate a relationship which is less apparent due to the large number of zero or close to zero TCDA sizes appearing across the range of %IR. Relying on these quantiles for TCDA approximation should be conducted with caution. The inverse parabola, and the negative estimates of TCDAs found, further supports the idea that smaller defect area gives a poor representation of %IR. For the larger defect areas, the approach taken in this thesis seems to be sufficient. This can be summarised as the interaction model seems to “follow” more from the MEOC data which can lead to

overfitting as opposed to the Non-Interaction variable which seems more generalised.

To account for the possible overfitting of the interaction model, the LASSO regression technique was applied to each quantile of the Non-Interaction and Interaction models. Some of the coefficients were shrunk to zero with the aim of generalising the models which makes variable selection faster and easier. The results suggest a more interpretable model with accuracy being maintained. Based on the findings, it is safe to suggest that if one were to apply these models to other pipelines (which is not part of MEOC) then the Non-Interaction model (R) is more suited. As in the case of subsequent ECDA assessments on the MEOC pipelines, the Interaction model (R) is more suitable since it “follows” the structure of the MEOC dataset.

A Bayesian approach to QR was used to predict the size of TCDA and also investigate the factors that may contribute to the %IR readings. Bayesian techniques allow an assessor to quantify the full spectrum of uncertainty in the prediction of parameters. In certain countries, the law dictates that an ECDA should be performed on a periodic schedule to ensure the safe continual operation of the pipeline. The NACE SP0502-2010 [97] highlights the importance of periodic assessments where “through successive applications of the ECDA method, an operator will be able to identify and address locations of corrosion activity which has occurred, is occurring and at locations where there is a potential to occur”. This makes the ECDA a continuous updating process. The Bayesian principle fits this very well since

initial assessments from this thesis can be used for future iterations of the ECDA. The main objective from this initial assessment is thus to produce data in the form of a prior distribution which can be used for future ECDA assessments of the buried pipelines. Models were constructed based on the linear approach and refinements made in the form of reducing variables within the models. To compensate the low linear correlation value obtained between %IR and TCDA, 4 outliers were taken out to facilitate the relationship. Additionally, an uninformative prior was used to reduce the level of subjectivity applied to the model. According to the results, Model 4 produces the best estimates of TCDA based on increasing %IR. Model 4 uses both the “filtset” data and also the refinement method mentioned earlier. As before, the interpretation based on the %IR of smaller defects should be treated with care due to large numbers of zero readings present at lower quantiles. For the prediction of %IR based on the TCDA (and other variables), Model 2a showed the best prediction aligned with current industry understanding. Model 2a also uses the “filtset” data and the refinement technique. In summary, it can be said that for a linear approach to the modelling of the MEOC data to be successful, careful judgement of outliers is critical and that also the refinement method where only meaningful variables are included in the model is needed. Since the models were built by the filtset data, it is recommended that these models are applicable for future assessments of the MEOC pipelines. For the comparisons between QR and BQR, the estimated coefficients by both approaches were quite similar. This is due to the uninformative prior used in the construction of the model, and

since the main objective of the research was done to produce a prior knowledge for subsequent inspection, this is not of a concern.

The investigation of the contributing variables to %IR continued with the LQR method. The LQR method is attractive for the investigation of %IR since it limits the interpretation of the dependent variable to within a defined range. Unlike Model 2a from the Bayesian approach, the interpretation of %IR is between 0 to 100. To keep the level of subjectivity low, only continuous variable were used and based on the results obtained from the Bayesian method, only relevant variables were chosen for the construction of the models. The resulting model reveals a pattern similar to those produced using the Bayesian method (Models 1 and 1a) because the same dataset was used (not “filtset” data). Model 1a (Figure 5-13) in the BQR method showed trends similar to those found in the LQR method with differences in the linearity and the range of %IR predicted. The differences can be attributed to the different linearity approach, the variables used within each model and the constraining effect of the LQR method. Nevertheless, the consistency in the trends between this approach and the BQR, shows that the model is applicable for the assessment of %IR in future ECDA projects. However, the model’s applicability is recommended only for the MEOC pipelines since it is heavily influenced by the data’s structure (similar to Model 1a). The general application of these models to other pipelines is possible but should be treated with care.

After considering all the models produced to predict TCDA, it is concluded that the Bayesian approach of prediction is the more attractive method for

future assessments of coating defects. This is due to Model 4 being concise with the added benefit of generating useful information (in the form of a prior distribution) for the next subsequent assessment of MEOC pipelines. Other methods lack this advantage. Moreover, the credible intervals produced are much narrower giving the assessor more confidence in the predicted values. For the prediction of %IR, it is concluded that the Bayesian approach (Model 2a) is the more practical for the next subsequent ECDA on the MEOC pipelines. The reason for this echo the previous conclusion, i.e. the model's simplicity and the incorporation of the current model for future assessments. However, the 0.05 quantile of Model 2a showed negative estimates which are better approximated with the LQR method. Both Models 4 and 2a have the added advantage of using the "filtset" data and hence the resulting models are more general in that they do not incorporate some of the "noises" present in the dataset. Undoubtedly, the "filtset" data still maintains the trait of the MEOC pipelines where other "noises" are still present. The generalisation does not mean the absolute application to every existing buried pipeline but as was stated earlier the application of the models to other pipelines beyond the MEOC scope should be done with care. One can argue that the usage of such a dataset is somewhat biased based on the Author's interpretation. While this is true to a certain extent, only four points were removed from the dataset and since the results we have here have the possibility to be incorporated into future iterations of the ECDA process using the Bayesian approach, this so-called bias will diminish with each iteration of assessment.

Corrosion is represented by the variable POPD. The POPD variable is one of the most important variables determining a pipeline's structural integrity. Similar to TCDA and %IR different modelling methods were used to determine its effect on other variables and *vice versa*. An initial linear assessment was done between TCDA and POPD. The result suggested that with every increasing TCDA, the depth of POPD decreases. Similar results were obtained with the interaction model where the derivation of the model with respect to POPD yields a decreasing trend. The results from Model 4 of the Bayesian approach is also in line with the earlier findings. Moreover, the results of Model 4 see a consistent upward trend across the TCDA quantiles. Between the 0.4 and the 0.6 quantiles emerged a plateau of estimates suggesting that for these defect sizes, the effect of an increasing POPD is minimal. The apparent flatness is an indication of the pit growth rate being faster than the growth of TCDA. This will produce deeper pits at smaller TCDA which solidifies the findings from the initial linear assessment in Chapter 4. The trait can also be seen in the LQR model where the POPD was modelled as the dependent variable. The results suggest that it is at the median pit depth that corrosion is most critical and that at some point in time, the growth of TCDA slows down or stops completely while the growth of pits carries on. Therefore, repair of such defects should be made before this. The consistencies in the findings from different approaches and different datasets (Model 4 uses filtset data) suggests that both models based on different variables are acceptable for predicting corrosion. This finding is also in parallel with findings from [42] which concluded that at coating defects, potential differences will occur due to the cathodic current applied to the

pipeline. If the current density provided to that particular coating defect is insufficient, then corrosion is highly probable. The potential at the bare metal tends to be anodic and the potential of metal underneath the coating surrounding the defect will be cathodic. Cathodic currents will protect the cathodic site (under coating) even further while leaving the exposed metal to undergo accelerated corrosion. This, according to the author's opinion, is what we are seeing with the MEOC dataset. In conclusion, for the sole purpose of determining the amount of corrosion in terms of POPD, the LQR POPD combined model seems to be the more appropriate since it is dedicated to predicting corrosion and is bounded within a particular interval. However, if iteration of subsequent assessments where the Bayesian method is required, then Model 4 is more suitable.

Another interesting finding in Chapter 4 was that there are indications where corrosion have some degree of correlation with the %IR reading. The suspicion is supported by the findings from Chapter 6 where the %IR combined model suggested that corrosion has a positive effect on the %IR signal across the quantiles. The results of the relationship between TCDA and POPD highlights a negative trend that suggests corrosion is not directly proportional to coating defect severity. Since the %IR is thought to be a relative measure of coating defect size, this further exemplifies the notion that %IR has some relation to corrosion (meaning that the signal being picked up by the DCVG assessment is not entirely from coating defects but also from corrosion).

7.0 Future Work

There are a few suggestions for future work that could improve the work presented here namely:

- The main limitation of this research is the usage of the MEOC pipelines data. Models were generated based on the assessments made from the indirect and direct analysis of an ECDA. Due to this, the models generated are attached to the data's structure. Further research is required when new data is made available in the pursuit of developing a generic equation applicable to a wider variety of situations. Although the statistical analysis of determining the p-values and the confidence intervals is designed to acquire confidence on the predicted coefficients, more datasets (samples) are needed to obtain an estimate which is ever closer to the true parameter value. Moreover, with new datasets, further refinement is a possibility which will make the models much more appealing in terms of application. The TCDA and %IR is contributed to by many more factors than highlighted in this thesis, e.g. weather, defect orientation, CP current capacity, etc. but that were not included due to the unavailability of data.

- It can be seen that some of the models suffer from quantile crossing. This might be due to the method of estimating the quantile which is done separately (one after the other). A suggestion for future work could be to employ methods which estimate the quantiles simultaneously and applying a penalty term which acts as constraints thus ensuring the quantiles do not cross paths.
- With the availability of more ECDA data, there is the possibility of dividing the data into 2 sets. One set could be the training set where the data is used to construct the models. Another set could be the test set where the previously constructed model from the training set is tested in terms of its performance and how it fits the test set. After going through this process, the determination of whether the models are underfitting (bias) or overfitting (high variance) is a possibility. Underfitted models tend to have higher error with both the training and the test set whereas for overfitted models, the errors are higher with the test set but lower in the training set. The testing of the models will add to the confidence of the general usability of the models. This approach was not done in this thesis due to the limited data.
- The incorporation of the CIPS data with the DCVG data of an ECDA report could lead to better estimation of coating defects and corrosion. CIPS data looks at the pipeline's potential and identifies defects based on location where the pipeline is under protected. When both the DCVG and the CIPS data are combined in a model, more accurate predictions of location and severity of defects is expected.

- Further research can go into looking at the usage of a nonlinear approach to the Bayesian analysis of ECDA data which could account for the apparent outliers (without the need to manually remove them as in the case of the “filtset” data). It is also suggested that a faster approach to the sampling method of the posterior distribution than the current sampling method of the Metropolis-Hastings algorithm. One such algorithm could be the Gibbs sampling method which is considered faster and uses less computational power.
- When new data is made available, the prior distribution obtained from this thesis could be used for better prediction of the TCDA and %IR.
- An experimental approach could be a possible future research where the main goal of the experiment is to replicate the results obtained here. The experiment could start in the lab with investigation on how currents travel in soil and react with corrosion and coating defects found in pipelines. From the results, it is expected that the modelling of the CP current will be made easier. A scaled-up experiment from the lab to the field application is also suggested.
- An experimental method to verify the claim that the %IR readings can detect corrosion on pipelines should be attempted. Based on the data and the analysis conducted here, this seems to be the case.
- New methods have been developed in the field of Bayesian Statistics. In particular the Bayesian Belief Network (BBN) is thought to be appropriate for considering all the variables. As stated earlier, the limitation of the MEOC data where it does not account for other variables which could contribute to TCDA and %IR, e.g. weather,

defect orientation etc., could have a significant effect on the final result. In the event of the availability of such data, a BBN could be a suitable way (catering for the large numbers of variables) of mapping out these variables and see how they interact. The network maps out the underlying independent variables (child nodes) that contribute to the dependent variable (parent nodes). The result will be in the form of a posterior distribution where the uncertainties of each parameter are also demonstrated. More on this can be found in papers such as [103–105].

- Another approach that is suggested is the use of machine learning algorithms such as the artificial neural network (ANN) for the ECDA data. The inputs of the network can be the contributing variables, e.g. SR, DOC etc., which go to neurons which process them and return an output that is the dependent variable of interest, e.g. TCDA. The connecting synapses and contributing weights adapt (through training) and the network learns to make useful computations. By training the model iteratively, and with the addition of more data, the estimates of the final output are expected to improve.

References

[1] K. P. Green and T. Jackson, *Safety in the transportation of Oil and Gas: pipelines or rail?* Fraser Institute, 2015.

[2] A. W. Dawotola, P. van Gelder, J. Charima, and J. Vrijling, “Estimation of failure rates of crude product pipelines,” *Applications of Statistics and Probability in Civil Engineering–Faber, Köhler & Nishijima (eds) Taylor & Francis Group, London*, 2011.

[3] N.Noor.et.al., “Methodology for Soil-Corrosion Study of Underground Pipeline,” *Conference: 2nd International Conference on Plant Equipment and Reliability (ICPER 2010), At Kuala Lumpur Convention Centre (KLCC)*, 2010.

[4] M. Romano, “The Ins and Outs of Pipeline Coating : Coatings Used to Protect Oil and Gas Pipelines,” *Journal of Protective Coatings & Linings*, 2005.

[5] K. R. Larsen, “Using Pipeline Coatings with Cathodic Protection Protective Coatings Need to Maintain Integrity under Operating Environments and in the Presence of CP,” *Materials Performance*, vol. 55, no. 1, pp. 32–35, 2016.

[6] S. Papavinasam, M. Attard, and R. W. Revie, “External polymeric pipeline coating failure modes,” *Materials performance*, vol. 45, no. 10, pp. 28–30, 2006.

[7] J.M. Esteban K.J. Kennelley M.E. Orazem and R. M. Degerstedt, “Mathematical Models for Cathodic Protection of an Underground Pipeline with Coating Holidays: Part 1 — Theoretical Development.” .

[8] N. SP0502, “Standard Practice—Pipeline External Corrosion Direct Assessment Methodology,” *NACE International, Houston, TX, Item*, no. 21097, 2008.

[9] F. Varela, M. Yongjun Tan, and M. Forsyth, “An overview of major methods for inspecting and monitoring external corrosion of on-shore

transportation pipelines," *Corrosion Engineering, Science and Technology*, vol. 50, no. 3, pp. 226–235, 2015.
[10] U. S. DOT, *49 CFR 192.925 - What are the requirements for using External Corrosion Direct Assessment (ECDA)?* U. S. Department of Transportation,, 2011.
[11] Z. Ahmad, *Principles of corrosion engineering and corrosion control*. Butterworth-Heinemann, 2006.
[12] P. P. Deshpande and D. Sazou, *Corrosion protection of metals by intrinsically conducting polymers*. CRC Press, 2016.
[13] M. G. Fontana, *Corrosion engineering*. Tata McGraw-Hill Education, 2005.
[14] J. Qiu, "Stainless steels and alloys: why they resist corrosion and how they fail," *Corrosion special topical papers. Nanyang Technological University, Singapore*, 1995.
[15] V. S. Sastri, *Green corrosion inhibitors: theory and practice*, vol. 10. John Wiley & Sons, 2012.
[16] M. Baker and R. R. Fessler, "Pipeline corrosion," *Integrity Management System*, 2008.
[17] G. H. Koch, M. P. Brongers, N. G. Thompson, Y. P. Virmani, and J. H. Payer, "Corrosion cost and preventive strategies in the United States," 2002.
[18] G. Koch, J. Varney, N. Thompson, O. Moghissi, M. Gould, and J. Payer, "International Measures of Prevention, Application, and Economics of Corrosion Technologies Study," *NACE International IMPACT Report*, 2016.
[19] H. H. Uhlig, "The Cost of Corrosion to The United States★," *Corrosion*, vol. 6, no. 1, pp. 29–33, 1950.
[20] S. F. Daily, *Understanding corrosion and cathodic protection of reinforced concrete structures*. Corrpro Companies, Incorporated, 1999.
[21] Y.-S. Choi, S. Nesic, and S. Ling, "Effect of H 2 S on the CO 2 corrosion of carbon steel in acidic solutions," *Electrochimica Acta*, vol. 56, no. 4, pp. 1752–1760, 2011.
[22] L. S. McNeill and M. Edwards, "The importance of temperature in assessing iron pipe corrosion in water distribution systems," *Environmental monitoring and assessment*, vol. 77, no. 3, pp. 229–242, 2002.
[23] K. Zakowski, M. Narozny, M. Szocinski, and K. Darowicki, "Influence of water salinity on corrosion risk—the case of the southern Baltic Sea coast," *Environmental monitoring and assessment*, vol. 186, no. 8, pp. 4871–4879, 2014.
[24] R. W. Revie, *Corrosion and corrosion control*. John Wiley & Sons, 2008.
[25] J. Hobbs, "Reliable Corrosion Inhibition in the Oil and Gas Industry," *Health and Safety Executive, UK*, 2014.
[26] R. M. Burns and A. E. Schuh, *Protective coatings for metals*. Reinhold Publishing Corporation.; New York, 1967.
[27] H. Davy, "On the corrosion of copper sheeting by sea water, and on methods of preventing this effect; and on their application to ships of

war and other ships," *Philosophical Transactions of the Royal Society of London*, vol. 114, pp. 151–158, 1824.
[28] F. Kajiyama, "STRATEGY FOR ELIMINATING RISKS OF CORROSION AND OVERPROTECTION FOR BURIED MODERN PIPELINES."
[29] A. P. I. (API) Association of Oil Pipe Lines (AOPL), "PIPELINE 101," *Pipeline101 - Where-Are-Pipelines-Located*. .
[30] R. J. Kuhn, "Cathodic protection of underground pipe lines from soil corrosion," *Proc. Am. Petroleum Inst.[IV]*, vol. 14, pp. 153–167, 1933.
[31] D. A. Jones, *Principles and prevention of corrosion*. Macmillan, 1992.
[32] DOT, *49 CFR 192.461 - External corrosion control: Protective coating*. U. S. Department of Transportation,, 2017.
[33] E. Akbarinezhad, J. Neshati, and F. Rezaei, "Investigation on organic pipeline coating effectiveness via electrochemical impedance spectroscopy," *Surface Engineering*, vol. 23, no. 5, pp. 380–383, 2007.
[34] I. Thompson and J. R. Saithala, "Review of pipeline coating systems from an operator's perspective," *Corrosion Engineering, Science and Technology*, vol. 51, no. 2, pp. 118–135, 2016.
[35] D. P. Riemer, "Modeling cathodic protection for pipeline networks," State University System of Florida, 2000.
[36] F. Bellucci and L. Nicodemo, "Water transport in organic coatings," *Corrosion*, vol. 49, no. 3, pp. 235–247, 1993.
[37] A. Korzhenko, M. Tabellout, and J. Emery, "Dielectric relaxation properties of the polymer coating during its exposition to water," *Materials chemistry and physics*, vol. 65, no. 3, pp. 253–260, 2000.
[38] R. Norsworthy, "Coatings Used In Conjuction With Cathodic Protection Shielding Vs Non-Shielding Pipeline Coatings," *CORROSION 2009*, 2009.
[39] PHMSA, "Pipeline Safety Stakeholder Communications," *PHMSA: Stakeholder Communications - Excavation Damage*. .
[40] S. Lee, W. Oh, and J. Kim, "Acceleration and quantitative evaluation of degradation for corrosion protective coatings on buried pipeline: Part I. Development of electrochemical test methods," *Progress in Organic Coatings*, vol. 76, no. 4, pp. 778–783, 2013.
[41] J. Wagner, "Cathodic Protection Design I," *NACE International, Houston, TX*, p. 6, 1993.
[42] D. P. Riemer and M. E. Orazem, "Modeling coating flaws with non-linear polarization curves for long pipelines," *Corrosion and Cathodic Protection Modelling and Simulation*, vol. 12, pp. 225–259, 2005.
[43] K. G. D. Eur Ing R. L. Kean, "Principles of Cathodic Protection," *National Physics Lab*.
[44] A. Bahadori, *Cathodic corrosion protection systems: a guide for oil and gas industries*. Gulf Professional Publishing, 2014.
[45] P. Francis, "Cathodic Protection," *National Physical Laboratory, dispon√vel em: http://www. npl. co. uk/lmm/corrosion_control, acedido em Abril de*, 2007.
[46] J. Priebe, "Close interval surveys are used to assess the effectiveness of installed cathodic protection systems on buried and submerged pipelines and ensure the CP systems are operating properly.," *Corrosionpedia*. Sep-2017.

[47] A. Francis, P. Geren, and B. Phillips, “New method helps operators assess line integrity,” *Oil & gas journal*, vol. 99, no. 48, pp. 73–73, 2001.
[48] E. McAllister, *Pipeline rules of thumb handbook: a manual of quick, accurate solutions to everyday pipeline engineering problems*. Gulf Professional Publishing, 2013.
[49] N. Yahaya, K. Lim, N. Noor, S. Othman, and A. Abdullah, “Effect of Clay and Moisture Content on Soil Corrosion Dynamic,” *Malays. J. Civil Eng*, vol. 23, no. 1, pp. 24–32, 2011.
[50] A. Ekine and G. Emujakporue, “Investigation of corrosion of buried oil pipeline by the electrical geophysical methods,” *Journal of Applied Sciences and Environmental Management*, vol. 14, no. 1, 2010.
[51] N. Md Noor, K. S. Lim, Y. Nordin, and A. Abdullah, “Corrosion study on X70-carbon steel material influenced by soil engineering properties,” in *Advanced Materials Research*, 2011, vol. 311, pp. 875–880.
[52] M. Norhazilan, Y. Nordin, K. Lim, R. Siti, A. Safuan, and M. Norhamimi, “Relationship between soil properties and corrosion of carbon steel,” *Journal of Applied Sciences Research*, vol. 8, no. 3, pp. 1739–1747, 2012.
[53] R. Petersen, R. Melchers, and others, “Long-term corrosion of cast iron cement lined pipes,” *Centre for Infrastructure Performance and Reliability*, vol. 23, pp. 1–10, 2012.
[54] A. Ismail and A. El-Shamy, “Engineering behaviour of soil materials on the corrosion of mild steel,” *Applied Clay Science*, vol. 42, no. 3, pp. 356–362, 2009.
[55] A. Benmoussa, M. Hadjel, and M. Traisnel, “Corrosion behavior of API 5L X-60 pipeline steel exposed to near-neutral pH soil simulating solution,” *Materials and corrosion*, vol. 57, no. 10, pp. 771–777, 2006.
[56] S. Gupta and B. Gupta, “The critical soil moisture content in the underground corrosion of mild steel,” *Corrosion science*, vol. 19, no. 3, pp. 171–178, 1979.
[57] O. Pritchard, S. H. Hallett, and T. S. Farewell, “Soil corrosivity in the UK–Impacts on Critical Infrastructure,” *Infrastructure Transitions Research Consortium Working paper series*, pp. 1–55, 2013.
[58] *ASTM G187-12a Standard Test Method for Measurement of Soil Resistivity Using the Two-Electrode Soil Box Method*. West Conshohocken, PA.: ASTM International, 2017.
[59] Y. Kleiner, B. Rajani, and D. Krys, “Impact of soil properties on pipe corrosion: Re-examination of traditional conventions,” in *Water Distribution Systems Analysis 2010*, 2010, pp. 968–982.
[60] C. Von Wolzogen Kühr and L. Van der Vlugt, “Graphitization of cast iron as an electrobiochemical process in anaerobic soils,” ARMY BIOLOGICAL LABS FREDERICK MD, 1964.
[61] P. R. Roberge, *Handbook of corrosion engineering*. McGraw-Hill,, 2000.
[62] J. Perdomo and I. Song, “Chemical and electrochemical conditions on steel under disbonded coatings: the effect of applied potential, solution resistivity, crevice thickness and holiday size,” *Corrosion Science*, vol. 42, no. 8, pp. 1389–1415, 2000.

[63] F. Anes-Arteche, K. Yu, U. Bharadwaj, C. Lee, and B. Wang, “Challenges in the application of DCVG-survey to predict coating defect size on pipelines,” *Materials and Corrosion*, vol. 68, no. 3, pp. 329–337, 2017.
[64] J. P. McKinney, “Evaluation of Above-Ground Potential Measurements for Assessing Pipeline Integrity.” 2006.
[65] O. C. Moghissi, J. P. McKinney, M. E. Orazem, D. D’Zurko, and others, “Predicting Coating Holiday Size Using Ecda Survey Data,” in *CORROSION 2009*, 2009.
[66] G. Ruschau and P. Kowalskiphmsa, “Evaluation and validation of above ground techniques for coating condition assessment,” Final report DTRS56-05-T-0004. CC Technologies (A DNV Company), 2006.
[67] M. Roche, D. Melot, and G. Paugam, “Recent experience with pipeline coating failures,” *JOURNAL OF PROTECTIVE COATINGS AND LININGS*, vol. 23, no. 10, p. 18, 2006.
[68] O. Olabisi, M. Miller, and L. Rankin, “PHMSA-SPONSORED RESEARCH: IMPROVEMENTS TO ECDA PROCESS-SEVERITY RANKING.”
[69] G. Matocha, J. F. Burns, K. C. Garrity, and A. Eastman, “ANALYSIS OF ILI, ACVG AND DCVG INDICATIONS IN CONJUNCTION WITH CIS A CASE STUDY,” in *2012 9th International Pipeline Conference*, 2012, pp. 9–16.
[70] Z. Masilela and J. Pereira, “Using the direct current voltage gradient technology as a quality control tool during construction of new pipelines,” *Engineering Failure Analysis*, vol. 5, no. 2, pp. 99–104, 1998.
[71] B. Al-Najjar, “On establishing cost-effective condition-based maintenance: Exemplified for vibration-based maintenance in case companies,” *Journal of Quality in Maintenance Engineering*, vol. 18, no. 4, pp. 401–416, 2012.
[72] D. An, N. H. Kim, and J.-H. Choi, “Practical options for selecting data-driven or physics-based prognostics algorithms with reviews.” .
[73] F. Ayello, N. Sridhar, G. Koch, V. Khare, A. Al-Mathen, and S. Safri, “Internal Corrosion Threat Assessment of Pipelines using Bayesian Networks,” in *Corrosion 2014 Conference, San Antonio, TX*, 2014.
[74] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and statistics for engineers and scientists*, vol. 5. Macmillan New York, 1993.
[75] K. Yu, Z. Lu, and J. Stander, “Quantile regression: applications and current research areas,” *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, no. 3, pp. 331–350, 2003.
[76] Q. Li, R. Xi, N. Lin, and others, “Bayesian regularized quantile regression,” *Bayesian Analysis*, vol. 5, no. 3, pp. 533–556, 2010.
[77] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
[78] K. Yu, P. Van Kerm, and J. Zhang, “Bayesian quantile regression: an application to the wage distribution in 1990s Britain,” *Sankhyā: The Indian Journal of Statistics*, pp. 359–377, 2005.

[79] K. Yu and R. A. Moyeed, “Bayesian quantile regression,” *Statistics & Probability Letters*, vol. 54, no. 4, pp. 437–447, 2001.
[80] M. Bottai, B. Cai, and R. E. McKeown, “Logistic quantile regression for bounded outcomes,” *Statistics in medicine*, vol. 29, no. 2, pp. 309–317, 2010.
[81] ISO, *ISO15589-1: Petroleum and Petrochemical and Natural Gas Industries - Cathodic Protection of Pipelines - Part 1: On-land pipelines*. Geneva, CH: International Organization for Standardization, 2015.
[82] ASTM, *ASTM G187-12a Standard Test Method for Measurement of Soil Resistivity Using Two Electrode Soil Box Method*. American Society for Testing Materials, 2013.
[83] J. Mora-Mendoza, L. Saucedo-Robles, H. Rodrıguez-Clemente, M. González-Núñez, G. Zavala-Olivares, and M. Hernández-Gayosso, “Integral diagnostic in the failure causes of external corrosion of a natural gas transport pipeline,” *Materials and Corrosion*, vol. 62, no. 8, pp. 796–801, 2011.
[84] L. S. Aiken, S. G. West, and R. R. Reno, *Multiple regression: Testing and interpreting interactions*. Sage, 1991.
[85] E. J. Pedhazur and L. P. Schmelkin, *Measurement, design, and analysis: An integrated approach*. Psychology Press, 2013.
[86] B. S. Everitt, *The Cambridge dictionary of statistics*. Cambridge University Press, 2006.
[87] J. H. Fitzgerald, “Evaluating soil corrosivity–Then and now,” *Materials Performance; (United States)*, vol. 32:10, Oct. 1993.
[88] L. Hao and D. Q. Naiman, *Quantile regression*, no. 149. Sage, 2007.
[89] R. Koenker, “Quantile Regression: 40 Years On,” *Annual Review of Economics*, vol. 9, no. 1, 2017.
[90] R. Koenker, *Quantile regression*, no. 38. Cambridge university press, 2005.
[91] M. Kutz, *Handbook of environmental degradation of materials*. William Andrew, 2005.
[92] M. Shwehdi and U. Johar, “Transmission line EMF interference with buried pipeline: essential and cautions,” in *International conference (ICNIR2003) EMFI*, 2003, pp. 20–22.
[93] N. S. TM0109, *Aboveground survey techniques for the evaluation of underground pipeline coating condition*. 2009.
[94] C. Ukiwe, S. McDonnell, and others, “Optimization Of The Coating Anomaly Detection And Prioritization Methodology Using Voltage Gradient Surveys,” in *CORROSION 2011*, 2011.
[95] B. C. Yen and G. D. Tofani, “Soil stress assessment can prevent corrosion, reduce pipeline coating damage,” *Oil Gas J.;(United States)*, vol. 83, no. 34, 1985.
[96] L. D. Jay and others, “Probability and statistics for engineering and sciences,” *Brooks/Cole Publishing Company Monterey, California*, 1982.
[97] NACE, *ANSI/NACE SP0502-2010: Pipeline External Corrosion Direct Assessment Methodology*. NACE International, 2010.
[98] N. Orsini, M. Bottai, and others, “Logistic quantile regression in Stata,” *Stata Journal*, vol. 11, no. 3, pp. 327–344, 2011.

[99] S. Columbu and M. Bottai, “Cognitive deterioration in cancer patients using logistic quantile regression,” in *Advances in Latent Variables-Methods, Models and Applications*, 2013.
[100] API, *579-1/ASME FFS-1 Fitness for Service*. 2007.
[101] T. S. Farewell, S. H. Hallett, J. A. Hannam, and R. J. Jones, “Soil impacts on national infrastructure in the United Kingdom,” *NSRI, Cranfield University, UK*, 2012.
[102] J. W. Palmer, W. Hedges, and J. L. Dawson, *The use of corrosion inhibitors in oil and gas production*. Maney Pub., 2004.
[103] G. Kabir, R. Sadiq, and S. Tesfamariam, “A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines,” *Structure and Infrastructure Engineering*, vol. 12, no. 8, pp. 874–889, 2016.
[104] G. Koch, F. Ayello, V. Khare, N. Sridhar, and A. Moosavi, “Corrosion threat assessment of crude oil flow lines using Bayesian network model,” *Corrosion Engineering, Science and Technology*, vol. 50, no. 3, pp. 236–247, 2015.
[105] O. Shabarchin and S. Tesfamariam, “Internal corrosion hazard assessment of oil & gas pipelines using Bayesian belief network model.” .

8.0 Appendix

	0.05		0.25		0.5		0.75		0.95	
	BQR	QR	BQR	QR	BQR	QR	BQR	QR	BQR	QR
(Intercept)	14.2	-13.13761	54.6	48.22268	86.1	98.85388	109	141.0879	23.6	102.7219
Total Coating Defect Area (TCDA)	0.0000022	0	0.00000631	-0.00001	0.0000687	0.00006	0.0000408	0.00003	0.0000532	0.00008
Soil Resistivity (SR)	-0.0000235	-0.00006	-0.000331	-0.00031	-0.000567	-0.00048	-0.00028	-0.00028	-0.000346	-0.00029
Percentage of Pit Depth to Wall Thickness (POPD)	0.00611	-0.01271	-0.0775	-0.09468	0.0439	0.06549	0.0892	-0.00068	0.108	-0.05744
Deposits under Coatings (DUC)	0.0079	0.04509	0.0234	0.0409	-0.0372	-0.06428	-0.135	-0.10804	-0.0704	-0.12529
Depth of Cover (DOC)	0.0549	0.03285	0.0887	0.09387	0.0933	0.10389	0.0454	-0.02672	0.0364	-0.00084
Time in Service (TIS)	-0.336	0.35863	-0.414	-0.25667	-0.374	-0.676	-0.778	-0.94074	1.19	-0.14359
Pipe Size (PS)	-0.0818	-0.00155	-0.865	-0.84935	-1.31	-1.35012	-0.697	-1.21011	0.285	-0.27746
Backfill Type (Rock)	5.2	61.03542	49.3	56.27788	50.8	53.49289	24.2	33.46651	10.7	53.12368
Backfill Type (Sand + Clay)	-1.03	5.42229	0.719	36.12577	16.3	13.42679	6.14	11.37001	-11.6	-23.56192
Backfill Type (Stones + Clay)	1.72	7.40428	3.1	4.89392	0.562	1.77951	-4.34	-6.88938	0.411	8.13043
Coating Type (Coal Tar)	-3.26	9.44636	2.11	4.64846	-0.215	-6.38968	-15.2	-15.07863	11	-23.40123
Coating Type (Polyethylene)	-6.28	3.99895	-12.1	-18.27787	0.368	-5.41162	-24.1	-32.55006	2.44	-40.23342
Backfill Geometry (Angular)	0.754	3.8084	-6.03	-7.5326	-19.9	-22.37914	-12	-26.75601	-8.69	-49.57319
Backfill geometry (Round + Angular)	-2.64	-4.59182	-9.22	-12.47078	-0.835	-3.92166	0.933	4.01135	-0.446	4.32595
pH Of Water in Soil (Acidic)	1.17	18.95119	1.98	6.01878	-8.1	-10.0818	-21.8	-26.26496	-4.8	-55.72907
pH Of Water in Soil (Alkaline)	8.41	12.88766	12.8	13.65743	0.753	0.8598	1.74	2.45899	-0.804	-13.1597
pH Of Water in Soil (Neutral)	7.24	12.1713	7.12	6.43559	7.03	7.31162	-0.415	0.17448	-14	-22.25217
pH Of Water underneath Coating (Acidic)	-0.943	-2.93952	-16.5	-20.01174	-3.24	-2.013	-24.6	-25.07575	-11	-37.14624
pH Of Water underneath Coating (Alkaline)	-2	-4.64393	-1.27	-2.37229	-7.78	-7.36889	-3.81	-4.85221	0.991	3.65309
pH Of Water underneath Coating (Neutral)	2.56	8.21353	0.381	-1.62633	-0.125	5.92506	10.3	20.15952	-1.99	-7.91314

Table 8-1: Bayesian Quantile Regression (BQR) Estimates along with Quantile Regression (QR) Estimates for Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 for Model 1

	0.05		0.25		0.5		0.75		0.95	
	BQR	QR	BQR	QR	BQR	QR	BQR	QR	BQR	QR
(Intercept)	14.700000	83.132320	77.900000	90.972630	79.400000	92.566150	30.900000	115.156400	22.000000	45.068440
Total Coating Defect Area (TCDA)	0.000074	0.000080	0.000048	0.000040	0.000062	0.000060	0.000142	0.000150	0.000229	0.000230
Soil Resistivity (SR)	0.000029	0.000260	-0.000107	-0.000060	0.000206	0.000220	0.000229	0.000300	0.000373	0.000140
Percentage of Pit Depth to Wall Thickness (POPD)	0.033400	0.045270	0.099400	0.129900	0.161000	0.196310	0.070000	0.052760	0.055800	0.007720
Deposits under Coatings (DUC)	-0.020900	-0.052210	-0.073600	-0.048550	-0.037300	-0.025340	-0.040400	-0.011620	-0.050000	-0.052110
Depth of Cover (DOC)	0.066800	0.086000	0.001990	-0.010590	-0.006960	-0.001840	-0.027100	-0.012090	0.068200	0.068940
Time in Service (TIS)	-0.116000	-0.130830	-0.088400	-0.187420	-0.234000	-0.326080	-0.199000	-0.314640	0.179000	0.104190
Pipe Size (PS)	-0.126000	-0.075030	-0.508000	-0.525720	-0.300000	-0.293320	-0.368000	-0.626540	0.186000	-0.551590
Backfill Type (Clay)	-2.230000	-38.348380	-0.338000	-6.882280	3.350000	-2.249960	60.800000	-1.579160	16.300000	22.248100
Backfill Type (Sand + Clay)	-1.630000	-32.617690	-10.100000	-17.796240	-16.700000	-21.625990	41.500000	-25.560030	4.900000	5.476250
Backfill Type (Stones + Clay)	11.400000	-22.535500	-2.180000	-10.973440	-3.290000	-10.132730	58.400000	-5.529310	25.400000	30.989040
Coating Type (PVC Cold Wrap)	-9.260000	-42.734530	-33.500000	-31.966310	-31.400000	-37.941470	-27.200000	-31.282620	-5.690000	-7.163900
Coating Type (Coal Tar)	7.570000	-26.436400	1.070000	5.887830	-1.450000	-8.369240	1.040000	-3.492240	29.700000	31.739610
Backfill Geometry (Round)	1.730000	2.402760	-0.364000	-5.000980	1.020000	2.231530	-2.950000	-10.478380	-1.580000	-4.825460
Backfill Geometry (Round + Angular)	-0.246000	-2.080580	-0.366000	-5.197000	-0.156000	0.672240	1.470000	-6.144310	1.580000	-0.068520
pH Of Water in Soil (Acidic)	3.980000	46.796290	15.200000	22.264600	7.400000	8.883040	2.440000	3.499160	9.200000	8.782310
pH Of Water in Soil (Alkaline)	-1.840000	-6.497280	-13.400000	-14.384740	-21.200000	-23.020260	-17.600000	-23.007100	2.620000	7.110790
pH Of Water in Soil (Neutral)	-11.400000	-13.758600	-13.000000	-12.392160	-11.200000	-12.078250	-11.200000	-14.119210	-16.800000	-19.397650
pH Of Water underneath Coating (Acidic)	2.800000	15.763990	0.825000	2.241060	1.020000	14.598710	1.130000	10.054170	-4.980000	-25.521390
pH Of Water underneath Coating (Alkaline)	-1.620000	-1.712290	-5.570000	-4.546200	-8.360000	-8.726300	-1.440000	-2.883960	-3.430000	-2.738770
pH Of Water underneath Coating (Neutral)	-4.960000	-5.807580	-29.400000	-36.952270	-6.670000	-8.779000	-12.400000	-14.218990	-13.100000	-29.966860

Table 8-2: Bayesian Quantile Regression (BQR) Estimates along with Quantile Regression (QR) Estimates for Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 for Model 2

	0.05		0.25		0.5		0.75		0.95	
	BQR	QR	BQR	QR	BQR	QR	BQR	QR	BQR	QR
(Intercept)	4.74	5.5141	42.156909	37.91505	87.5	84.82694	87.1	112.9995	64.9	79.98014
TCDA	-0.0000353	0	0.000012	0.00002	0.0000828	0.00009	0.0000678	0.00006	0.000073	0.00002
Soil Resistivity (SR)	0.000000565	0.00018	-0.000401	-0.00036	-0.000668	-0.00076	-0.0000789	-0.00015	-0.000296	-0.0006
Depth of Cover (DOC)	0.0508	0.04424	0.084797	0.09058	0.0722	0.07501	0.0433	0.03004	0.0228	0.03386
Pipe Size (PS)	-0.158	-0.15242	-0.960746	-0.85812	-1.77	-1.73404	-1.12	-1.79002	0.432	-0.04051
Backfill Type (Rock)	5.23	62.024	49.131534	55.3512	53.4	54.15413	39.7	57.59096	6.45	20.88416
Backfill Type (Sand + Clay)	-0.939	2.40057	0.40573	35.38389	25.4	15.96705	18.8	17.56455	-8.21	-21.09323
Backfill Type (Stones + Clay)	1.56	3.19906	0.816277	4.95639	0.619	1.6121	-0.0571	1.58495	-1.46	-3.57373
Coating Type (Coal Tar)	2	1.83555	8.088755	8.13012	5.54	5.11152	-0.604	-0.52326	-6.78	-4.55448
Coating Type (Polyethylene)	-0.113	-0.91281	-4.374617	-14.44378	6.77	8.63253	-0.561	-1.9322	-15.3	-31.76233
Backfill Geometry (Angular)	0.434	2.82207	-5.116034	-6.01787	-18.2	-16.65991	-18.1	-38.3265	-6.73	-26.70664
Backfill geometry (Round + Angular)	-3.65	-5.71419	-6.09831	-11.24719	0.251	1.66144	0.168	0.01712	-1.24	-0.83484
pH Of Water in Soil (Acidic)	1.2	20.13217	1.953859	7.31045	-6.07	-7.82088	-20.6	-27.82607	-4.67	-56.69553
pH Of Water in Soil (Alkaline)	8.11	10.85544	12.469238	15.72156	1.76	3.61206	4.57	-3.16313	-0.575	-6.19104
pH Of Water in Soil (Neutral)	5.06	9.08957	4.689159	7.40692	1.14	4.0128	-8	-8.64504	-12.6	-17.33608

Table 8-3: Bayesian Quantile Regression (BQR) Estimates along with Quantile Regression (QR) Estimates for Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 for Model 1a

	0.05		0.25		0.5		0.75		0.95	
	BQR	QR	BQR	QR	BQR	QR	BQR	QR	BQR	QR
(Intercept)	30.300000	86.797670	96.400000	97.640580	86.200000	96.027370	29.900000	103.189300	31.300000	44.699490
TCDA	0.000096	0.000100	0.000049	0.000060	0.000077	0.000090	0.000139	0.000160	0.000221	0.000220
Soil Resistivity (SR)	-0.000132	0.000040	-0.000025	0.000050	-0.000178	-0.000230	0.000330	0.000280	0.000482	0.000120
Depth of Cover (DOC)	0.056100	0.090840	-0.002670	0.001760	-0.066500	-0.060630	-0.043000	-0.034910	-0.093900	0.064510
Pipe Size (PS)	-0.301000	-0.239980	-0.702000	-0.831970	-0.452000	-0.427980	-0.405000	-0.631050	-0.005790	-0.435420
Backfill Type (Clay)	-18.900000	-41.563750	-16.700000	-12.980020	0.785000	-2.195680	57.000000	-0.185940	18.600000	20.723690
Backfill Type (Sand + Clay)	-19.100000	-38.162340	-26.400000	-24.253120	-20.300000	-22.137160	38.600000	-14.822950	10.900000	7.021800
Backfill Type (Stones + Clay)	-6.250000	-27.517880	-15.800000	-10.366370	-3.210000	-7.575040	54.600000	-4.387000	25.900000	27.062110
Coating Type (PVC Cold Wrap)	-6.480000	-47.243800	-37.800000	-38.470090	-32.100000	-36.438630	-28.100000	-33.995890	-8.590000	-8.357880
Coating Type (Coal Tar)	8.510000	-30.701130	-0.147000	-1.225670	-0.720000	-5.916940	0.433000	-4.990850	29.700000	31.984050
Backfill Geometry (Round)	1.220000	4.263210	0.296000	-0.496600	-0.279000	-3.654250	-2.510000	-6.354480	-0.829000	-4.983210
Backfill geometry (Round + Angular)	-0.060200	1.018380	-1.790000	-4.468820	-0.243000	-2.953660	1.070000	-1.757900	1.220000	0.802880
pH Of Water in Soil (Acidic)	4.130000	52.491990	23.000000	31.026370	6.630000	9.004480	5.160000	7.025120	11.900000	9.864600
pH Of Water in Soil (Alkaline)	-1.260000	-4.982980	-9.870000	-9.126160	-18.900000	-20.692270	-18.800000	-24.134410	4.000000	10.924820
pH Of Water in Soil (Neutral)	-11.300000	-14.624370	-6.340000	-4.029990	-11.600000	-13.155980	-11.600000	-14.112290	-16.800000	-19.629660

Table 8-4: Bayesian Quantile Regression (BQR) Estimates along with Quantile Regression (QR) Estimates for Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 for Model 2a

	0.05		0.25		0.5		0.75		0.95	
	BQR	**QR**	**BQR**	**QR**	**BQR**	**QR**	**BQR**	**QR**	**BQR**	**QR**
(Intercept)	-465	-457.17702	-4180	-4174.93367	78687.177	79808.12772	216000	217908.3647	189000	224992.2512
%IR	-0.0178	-0.00289	2.3	2.39424	84.428	81.74019	849	850.32554	-93.1	-185.38455
Soil Resistivity (SR)	-0.0034	-0.00343	0.0456	0.04597	-0.524	-0.52755	-0.00013	-0.02681	0.151	0.91576
Percentage of Pit Depth to Wall Thickness (POPD)	6.04	5.95071	169	169.30994	232.204	227.41219	20.1	23.50078	2740	2655.77653
Deposits under Coatings (DUC)	-0.901	-0.91188	-6.48	-6.57018	19.543	20.98267	250	258.12499	-257	-180.88631
Depth of Cover (DOC)	-0.00321	-0.0049	-1.31	-1.4892	-69.776	-71.17661	-317	-314.25913	-111	-116.58299
Time in Service (TIS)	4.92	4.94473	33.4	33.47407	-2351.485	-2366.39447	-6610	-6654.21784	-8030	-8229.21972
Pipe Size (PS)	10.4	10.12846	102	102.4972	707.098	696.70204	2130	2107.66508	6040	5226.80883

Table 8-5: Bayesian Quantile Regression (BQR) Estimates along with Quantile Regression (QR) Estimates for Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 for Model 3

	0.05		0.25		0.5		0.75		0.95	
	BQR	**QR**	**BQR**	**QR**	**BQR**	**QR**	**BQR**	**QR**	**BQR**	**QR**
(Intercept)	-464	-451.2233	-5710	-5964.69088	28978.772	29852.26172	100000	124534.4809	77655.6	82909.55
%IR	0.0221	0.01055	20.2	19.78903	475.876	467.86167	1320	1275.76375	1481.9	1486.526
Soil Resistivity (SR)	-0.00336	-0.00347	0.0576	0.0576	-0.453	-0.45064	0.679	0.5088	12.2	12.51588
Percentage of Pit Depth to Wall Thickness (POPD)	6.06	5.81388	177	178.06266	219.149	219.97055	237	242.43918	406	375.3882
Deposits under Coatings (DUC)	-0.916	-0.92355	-6.19	-6.19977	34.219	35.45087	152	192.12718	-218.8	-203.1032
Depth of Cover (DOC)	-0.00871	0.00592	-5.19	-4.93221	-20.033	-21.78868	-221	-210.58953	-97.8	-115.9905
Time in Service (TIS)	4.95	4.92626	18.9	25.61559	-1681.945	-1695.79043	-4820	-5333.01823	-5066.9	-5108.585
Pipe Size (PS)	10.3	9.95343	155	154.41867	963.069	960.79513	2750	2579.36638	4610.6	4529.043

Table 8-6: Bayesian Quantile Regression (BQR) Estimates along with Quantile Regression (QR) Estimates for Quantiles 0.05, 0.25, 0.5, 0.75 and 0.95 for Model 4

Variables	Quantiles								
	0.05			0.5			0.95		
		Credible Intervals			Credible Intervals			Credible Intervals	
	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975
(Intercept)	14.2	-2.95	38.4	86.1	77	95.5	23.6	-28	73.5
IR Drop (%IR)	0.0000022	-0.000012	0.0000384	0.0000687	0.0000517	0.0000837	0.0000532	0.00000513	0.0000788
Soil Resistivity (SR)	-0.0000235	-0.00045	0.000325	-0.000567	-0.000881	-0.000293	-0.000346	-0.00064	0.0000823
Percentage of Pit Depth to Wall Thickness (POPD)	0.00611	-0.102	0.118	0.0439	-0.0321	0.12	0.108	-0.0198	0.264
Deposits under Coatings (DUC)	0.0079	-0.0358	0.0511	-0.0372	-0.0764	0.00186	-0.0704	-0.139	-0.00845
Depth of Cover (DOC)	0.0549	0.0111	0.108	0.0933	0.0675	0.122	0.0364	-0.0251	0.109
Time in Service (TIS)	-0.336	-0.779	0.207	-0.374	-0.561	-0.189	1.19	0.137	2.65
Pipe Size (PS)	-0.0818	-0.329	0.104	-1.31	-1.53	-1.11	0.285	-0.159	0.739
Backfill Type (Rock)	5.2	-9.36	47.8	50.8	42.7	57.6	10.7	-2.52	47.9
Backfill Type (Sand + Clay)	-1.03	-12.7	3.38	16.3	5.88	30.6	-11.6	-27.8	1.94
Backfill Type (Stones + Clay)	1.72	-1.43	6.36	0.562	-1.46	3.69	0.411	-3.95	6.33
Coating Type (Coal Tar)	-3.26	-8.38	8.85	-0.215	-2.8	1.82	11	-5.26	35
Coating Type (Polyethylene)	-6.28	-20.7	5.22	0.368	-2.97	4.66	2.44	-25.9	21.6
Backfill Geometry (Angular)	0.754	-3.01	5.65	-19.9	-23.2	-16.4	-8.69	-36.8	2.98
Backfill geometry (Round + Angular)	-2.64	-7.12	0.369	-0.835	-4.32	1.22	-0.446	-6.5	5.37
pH Of Water in Soil (Acidic)	1.17	-10.3	16.5	-8.1	-14.8	0.286	-4.8	-44	9.5
pH Of Water in Soil (Alkaline)	8.41	-0.222	15.3	0.753	-1.25	4.34	-0.804	-7.01	4.32
pH Of Water in Soil (Neutral)	7.24	-0.67	15.8	7.03	1.03	11.1	-14	-20.3	-0.277
pH Of Water underneath Coating (Acidic)	-0.943	-10.3	3.26	-3.24	-20.6	1.32	-11	-42.4	4.92
pH Of Water underneath Coating (Alkaline)	-2	-5.33	0.492	-7.78	-10.1	-5.28	0.991	-1.36	5.77
pH Of Water underneath Coating (Neutral)	2.56	-4.74	13	-0.125	-3.71	3.08	-1.99	-11.9	4.81

Table 8-7: Bayesian Quantile Regression (BQR) Estimates With 95 % Credible Intervals for Quantiles 0.05, 0.5 and 0.95 for Model 1

Variables	Quantiles								
	0.05			0.5			0.95		
		Credible Intervals			Credible Intervals			Credible Intervals	
	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975
(Intercept)	14.7	-1.76	40.681215	79.4	43.7	92.1	22	-415	69.1
Total Coating Defect Area (TCDA)	0.0000741	0.0000676	0.000128	0.0000618	0.0000493	0.0000769	0.000229	0.000168	0.000232
Soil Resistivity (SR)	0.0000293	-0.000316	0.000383	0.000206	0.0000175	0.000329	0.000373	-0.0000612	0.000848
Percentage of Pit Depth to Wall Thickness (POPD)	0.0334	-0.085	0.144788	0.161	0.088	0.232	0.0558	-0.0308	0.238
Deposits under Coatings (DUC)	-0.0209	-0.0789	0.029226	-0.0373	-0.0655	-0.0111	-0.05	-0.115	0.00565
Depth of Cover (DOC)	0.0668	0.0269	0.108152	-0.00696	-0.0324	0.0161	0.0682	0.012	0.119
Time in Service (TIS)	-0.116	-0.328	0.085368	-0.234	-0.338	-0.128	0.179	-0.22	0.788
Pipe Size (PS)	-0.126	-0.31	0.079494	-0.3	-0.472	-0.153	0.186	-0.575	1.93
Backfill Type (Clay)	-2.23	-13.6	3.720633	3.35	-2.41	23.9	16.3	-7.2	260
Backfill Type (Sand + Clay)	-1.63	-12.8	5.18233	-16.7	-23.9	3.6	4.9	-26.6	246
Backfill Type (Stones + Clay)	11.4	-0.0934	18.838816	-3.29	-9.69	18.8	25.4	0.026	262
Coating Type (PVC Cold Wrap)	-9.26	-29.1	2.434434	-31.4	-38.2	-22	-5.69	-17.9	0.765
Coating Type (Coal Tar)	7.57	-11	20.359542	-1.45	-8.49	7.7	29.7	16.4	38.7
Backfill Geometry (Round)	1.73	-0.761	6.591509	1.02	-0.55	3.55	-1.58	-7.62	16.5
Backfill Geometry (Round + Angular)	-0.246	-4.13	3.671415	-0.156	-2.4	2.01	1.58	-2.14	10.4
pH Of Water in Soil (Acidic)	3.98	-10.1	38.354927	7.4	-0.427	14.5	9.2	-0.768	30.3
pH Of Water in Soil (Alkaline)	-1.84	-7.75	0.907865	-21.2	-24.3	-17.7	2.62	-1.7	12.2
pH Of Water in Soil (Neutral)	-11.4	-19.6	-0.207547	-11.2	-14.5	-8.07	-16.8	-26.6	-0.241
pH Of Water underneath Coating (Acidic)	2.8	-6.1	18.678275	1.02	-2.44	9.92	-4.98	-27.4	5.78
pH Of Water underneath Coating (Alkaline)	-1.62	-4.69	0.344062	-8.36	-10.1	-6.7	-3.43	-7.89	0.356
pH Of Water underneath Coating (Neutral)	-4.96	-16.5	0.619875	-6.67	-14.2	0.543	-13.1	-33.1	2.77

Table 8-8: Bayesian Quantile Regression (BQR) Estimates With 95% Credible Intervals for Quantiles 0.05, 0.5 and 0.95 for Model 2

Variables	Quantiles								
	0.05			0.5			0.95		
		Credible Intervals			Credible Intervals			Credible Intervals	
	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975
(Intercept)	4.74	-1.53	14.4	87.5	78.2	97	64.9	56	84.8
TCDA	-0.0000353	-0.00004	0.0000375	0.0000828	0.0000661	0.0000989	0.000073	0.0000008	0.0000833
Soil Resistivity (SR)	0.000000565	-0.000344	0.000364	-0.000668	-0.000863	-0.000416	-0.000296	-0.000641	0.0000428
Depth of Cover (DOC)	0.0508	0.0148	0.0888	0.0722	0.0447	0.099	0.0228	-0.0243	0.0562
Pipe Size (PS)	-0.158	-0.388	0.0759	-1.77	-2.01	-1.53	0.432	-0.199	0.64
Backfill Type (Rock)	5.23	-9.26	47.9	53.4	45.1	60.9	6.45	-4.09	36.7
Backfill Type (Sand + Clay)	-0.939	-7.93	3.08	25.4	16.1	36.8	-8.21	-22.3	2.47
Backfill Type (Stones + Clay)	1.56	-0.754	5.66	0.619	-0.927	3.38	-1.46	-8.41	2.96
Coating Type (Coal Tar)	2	-0.739	6.36	5.54	3	8.17	-6.78	-11.9	0.445
Coating Type (Polyethylene)	-0.113	-8.54	6.18	6.77	-0.593	14.7	-15.3	-34.4	2.78
Backfill Geometry (Angular)	0.434	-3.6	4.73	-18.2	-22.7	-14.4	-6.73	-30.5	2.56
Backfill geometry (Round + Angular)	-3.65	-7.75	-0.335	0.251	-1.89	2.89	-1.24	-10.2	5.22
pH Of Water in Soil (Acidic)	1.2	-10.2	16.7	-6.07	-12.7	0.598	-4.67	-42.2	9.47
pH Of Water in Soil (Alkaline)	8.11	-0.303	13.7	1.76	-1.08	5.82	-0.575	-6.11	3.91
pH Of Water in Soil (Neutral)	5.06	-1.05	12.4	1.14	-1.34	4.89	-12.6	-19.3	0.294

Table 8-9: Bayesian Quantile Regression (BQR) Estimates With 95% Credible Intervals for Quantiles 0.05, 0.5 and 0.95 for Model 1a

Variables	Quantiles								
	0.05			0.5			0.95		
		Credible Intervals			Credible Intervals			Credible Intervals	
	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975	Posterior Mean	0.025	0.975
(Intercept)	30.3	-2.29	501	86.2	77.7	94.1	31.3	-308	79.5
TCDA	0.0000956	0.0000936	0.000138	0.0000768	0.0000547	0.0000989	0.000221	0.000186	0.000222
Soil Resistivity (SR)	-0.000132	-0.000408	0.000202	-0.000178	-0.000395	0.0000902	0.000482	0.0000361	0.00108
Depth of Cover (DOC)	0.0561	0.0195	0.0981	-0.0665	-0.0907	-0.0337	0.0639	0.0178	0.0949
Pipe Size (PS)	-0.301	-0.333	0.00686	-0.452	-0.622	-0.309	-0.00579	-0.571	0.214
Backfill Type (Clay)	-18.9	-393	4	0.785	-2.71	5.56	18.6	-6.5	323
Backfill Type (Sand + Clay)	-19.1	-392	3.51	-20.3	-26.2	-13.6	10.9	-24.2	300
Backfill Type (Stones + Clay)	-6.25	-381	17.7	-3.21	-7.47	0.767	25.9	-0.151	324
Coating Type (PVC Cold Wrap)	-6.48	-51	2.72	-32.1	-36.8	-28.5	-8.59	-29.3	0.1
Coating Type (Coal Tar)	8.51	-37.8	18.5	-0.72	-5.65	2.67	29.7	9.06	38.9
Backfill Geometry (Round)	1.22	-5.87	5.43	-0.279	-2.77	1.51	-0.829	-6.18	2.5
Backfill geometry (Round + Angular)	-0.0602	-5.43	3.04	-0.243	-3.27	2.1	1.22	-1.56	7.78
pH Of Water in Soil (Acidic)	4.13	-9.69	39.6	6.63	-0.543	14	11.9	0.0936	31.7
pH Of Water in Soil (Alkaline)	-1.26	-6.84	1.42	-18.9	-21.5	-15.8	4	-1.39	12.3
pH Of Water in Soil (Neutral)	-11.3	-18.2	-0.912	-11.6	-15.9	-6.64	-16.8	-24.1	-0.0591

Table 8-10: Bayesian Quantile Regression (BQR) Estimates With 95% Credible Intervals for Quantiles 0.05, 0.5 and 0.95 for Model 2a

Variables	Quantiles								
	0.05			0.5			0.95		
		Credible Intervals			Credible Intervals			Credible Intervals	
	Posterior Mean	0.25	0.975	Posterior Mean	0.25	0.975	Posterior Mean	0.25	0.975
(Intercept)	-465	-496.4146	-445	78687.177	74052.516	79414.664	189000	128000	207000
%IR	-0.0178	-0.1091	0.0559	84.428	82.292	100.219	-93.1	-162	28.6
Soil Resistivity (SR)	-0.0034	-0.0037	-0.00304	-0.524	-0.529	-0.476	0.151	-0.0154	0.482
Percentage of Pit Depth to Wall Thickness (POPD)	6.04	5.4327	6.79	232.204	228.826	264.055	2740	2640	2790
Deposits under Coatings (DUC)	-0.901	-0.9835	-0.821	19.543	10.358	20.721	-257	-387	-212
Depth of Cover (DOC)	-0.00321	-0.0523	0.0392	-69.776	-71.118	-53.681	-111	-122	-70.9
Time in Service (TIS)	4.92	4.6761	5.16	-2351.485	-2362.104	-2319.907	-8030	-8200	-7690
Pipe Size (PS)	10.4	9.8316	11.3	707.098	702.75	728.063	6040	5670	7410

Table 8-11: Bayesian Quantile Regression (BQR) Estimates with 95% Credible Intervals for Quantiles 0.05, 0.5 and 0.95 for Model 3

Variables	Quantiles								
	0.05			0.5			0.95		
		Credible Intervals			Credible Intervals			Credible Intervals	
	Posterior Mean	0.025	0.975	Posterior Mean	0.25	0.975	Posterior Mean	0.25	0.975
(Intercept)	-4.64E+02	-4.95E+02	-4.45E+02	2.90E+04	2.91E+04	29374.475	7.77E+04	6.65E+04	78394.4
%IR	2.21E-02	-5.19E-02	1.07E-01	4.76E+02	4.72E+02	476.451	1.48E+03	1.46E+03	1828.3
Soil Resistivity (SR)	-3.36E-03	-3.65E-03	-2.98E-03	-4.53E-01	-4.54E-01	-0.452	1.22E+01	1.17E+01	12.3
Percentage of Pit Depth to Wall Thickness (POPD)	6.06E+00	5.46E+00	6.79E+00	2.19E+02	2.19E+02	219.652	4.06E+02	-1.72E+02	455.7
Deposits under Coatings (DUC)	-9.16E-01	-9.99E-01	-8.31E-01	3.42E+01	3.40E+01	34.787	-2.19E+02	-2.24E+02	-155.8
Depth of Cover (DOC)	-8.71E-03	-6.00E-02	2.84E-02	-2.00E+01	-2.07E+01	-19.534	-9.78E+01	-1.14E+02	-91.7
Time in Service (TIS)	4.95E+00	4.71E+00	5.19E+00	-1.68E+03	-1.69E+03	-1684.384	-5.07E+03	-5.17E+03	-3763.1
Pipe Size (PS)	1.03E+01	9.81E+00	1.12E+01	9.63E+02	9.60E+02	961.977	4.61E+03	3.54E+03	4720.8

Table 8-12: Bayesian Quantile Regression (BQR) Estimates With 95% Credible Intervals for Quantiles 0.05, 0.5 and 0.95 for Model 4