



Probability of Failure Feedback Loop Due to Internal Corrosion in Oil Production Systems

A thesis submitted for the degree of Master of
Philosophy by

Martin Quinn

Department of Mechanical and Aerospace Engineering
Brunel University London
London
United Kingdom

Declaration

I have not plagiarised this work

.....

.....

Martin Quinn

Date

Dedication

I would like to thank everyone who has supported me during the course of my research, including my colleagues in TWI, Ujjwal, Payam and Shervin. I would like to thank Amanda who has supported me throughout the process.

Finally, I would like to thank my supervisor Dr. Bin Wang, for his unending patience and support throughout the course of my research.

Abstract

This dissertation aims to provide a methodology for creating a dynamic feedback loop to predict failures due to internal corrosion on high-pressure process piping systems in the oil and gas industry. This has been aided by applying machine learning capabilities to data available in a standard integrity management system.

While great advances have been made in integrity management in recent years a significant number of failures still occur due to pipes corroding from the inside out. Although real time or “Online Monitoring” systems are available, they are usually only used on critical piping systems where there are known issues or where the loss of integrity would cause massive disruption to business or loss of life.

In a typical facility, such as the one used as the basis of this dissertation, the total length of the pipework can easily exceed 20km. For piping systems not covered by online monitoring the structural integrity is managed through periodic inspection where the remaining wall thickness is measured by ultrasonic thickness (UT) inspection at specific points known as Condition Monitoring Locations (CMLs). This thickness is then compared to the minimum required thickness (T_{min}) for that CML.

Given that the number of CMLs governed by an integrity management system can easily exceed 100,000 with anywhere between 25,000 to 200,000 UT readings taken on these CMLs being added annually, patterns in data that may have been recognised in smaller sets can go unnoticed. Machine Learning can offer the capability of detecting these patterns in larger datasets.

The algorithms developed for this study aim to provide the basic input to screen for different types of corrosion in the entire system by

- 1) Manually identifying corrosion patterns through inspection
- 2) Training the system to classify this specific corrosion pattern
- 3) Training the system to determine corrosion rate
- 4) Generating the output, i.e. a ranking of all the CMLs in a system by how likely they are to experience the same corrosion pattern
- 5) Determining the corrosion rate from the trained network

| | |
|--|----|
| Abstract..... | i |
| 1. Introduction | 1 |
| 1.1. Structural Integrity Overview..... | 1 |
| 1.1.1. Oil and Gas Industry | 1 |
| 1.1.2. Structural Integrity | 2 |
| 1.1.3. Damage Mechanisms..... | 3 |
| 1.1.4. Inspection Methods | 4 |
| 1.2. Risk Based Inspection..... | 8 |
| 1.2.1. Background | 8 |
| 1.2.2. Current Approach..... | 9 |
| 1.3. Experience..... | 12 |
| 1.4. Aim | 14 |
| 1.5. Objective | 16 |
| 1.6. Contribution of this research | 16 |
| 1.7. Condition Monitoring Locations (CMLs) | 16 |
| 2. Literature Review | 17 |
| 2.1. Introduction | 17 |
| 2.2. Theory of Asset Integrity Management | 18 |
| 2.2.1. Ageing Equipment..... | 18 |
| 2.2.2. Inspection Frequency..... | 19 |
| 2.2.3. The Bathtub Curve | 20 |
| 2.3. Evolution of Risk Based inspection | 21 |
| 2.3.1. Risk | 21 |
| 2.3.2. Probability of Failure..... | 21 |
| 2.3.3. Thinning Damage Factor | 22 |
| 2.4. Risk based inspection..... | 23 |
| 2.4.1. Risk of Failure | 23 |
| 2.4.2. Risk Matrix | 23 |
| 2.4.3. Assessment Periods | 23 |
| 2.4.4. Risk Target..... | 23 |
| 2.5. API 581 Approach to internal thinning | 26 |
| 2.6. ANOVA in the Nuclear Industry | 29 |
| 2.7. Norsok Models v API 581 models | 30 |
| 2.8. Failure investigations | 31 |
| 2.9. Distribution of Corrosion Rates..... | 31 |
| 2.9.1. Distribution Selection..... | 33 |

| | | |
|---------|---|----|
| 2.10. | Determining PoF(t) of internal corrosion..... | 33 |
| 2.10.1. | Dataset..... | 33 |
| 2.10.2. | Remaining Life..... | 34 |
| 2.11. | Machine Learning..... | 37 |
| 2.12. | Summary..... | 37 |
| 3. | Data Structure..... | 38 |
| 3.1. | Introduction..... | 38 |
| 3.2. | Risk Based Inspection..... | 38 |
| 3.3. | PoF Calculation..... | 39 |
| 3.3.1. | Effect of inspection on PoF..... | 40 |
| 3.4. | Summary..... | 41 |
| 4. | Machine Learning Approach to Corrosion Rates..... | 42 |
| 4.1. | Introduction..... | 42 |
| 4.2. | Current Approach..... | 42 |
| 4.2.1. | Corrosion rate characteristics..... | 44 |
| 4.2.2. | Linear Corrosion..... | 44 |
| 4.2.3. | Error in measurement..... | 44 |
| 4.2.4. | Short Term Corrosion..... | 46 |
| 4.2.5. | Long Term Corrosion..... | 46 |
| 4.3. | Data Normalisation..... | 47 |
| 4.4. | Corrosion classification..... | 47 |
| 4.5. | Corrosion Rate Calculation..... | 51 |
| 4.6. | Summary..... | 51 |
| 5. | Discussion of results..... | 52 |
| 5.1. | Results Overview..... | 52 |
| 5.2. | Accuracy..... | 53 |
| 5.2.1. | Overall Accuracy..... | 53 |
| 5.2.2. | Accuracy of Neural Network Entire Dataset..... | 54 |
| 6. | Conclusions and Future Work..... | 57 |
| 6.1. | Conclusions..... | 57 |
| 6.1.1. | Current Approach..... | 57 |
| 6.1.2. | Feedback..... | 58 |
| 6.1.3. | CR Calculation..... | 59 |
| 6.1.4. | Classification..... | 60 |
| 6.1.5. | Validation..... | 60 |
| 6.2. | Comparison of PoF vs API 581..... | 61 |

| | | |
|------------|--|----|
| 6.2.1. | Effectiveness A | 61 |
| 6.2.2. | Effectiveness B | 62 |
| 6.2.3. | Effectiveness C | 62 |
| 6.2.4. | Effectiveness D | 63 |
| 6.3. | Future Work | 63 |
| 6.3.1. | Supplementary operation and construction data..... | 64 |
| 6.3.2. | Combining supplementary non-UT inspection results | 64 |
| 6.3.3. | Classification Potential | 64 |
| 6.3.4. | Corrosion Classification Potential | 65 |
| 6.3.5. | Validation of Results | 65 |
| 6.4. | Recommendations | 65 |
| | Feedback Loop | 65 |
| | Neural Network..... | 66 |
| | References | 67 |
| APPENDIX A | Table of Sample Piping System Results..... | 1 |
| APPENDIX B | Neural Network Results | 2 |
| | Accuracy of Neural Network With 3 Available Readings | 2 |
| | Accuracy of Neural Network With 4 Available Readings | 5 |
| | Accuracy of Neural Network With 5 Available Readings | 7 |
| | Accuracy of Neural Network With 6 Available Readings | 10 |
| APPENDIX C | Corrosion Rate Distributions..... | 13 |
| | Straight Sections | 13 |
| | Fittings 14 | |
| | Full Dataset | 15 |
| APPENDIX D | Code Used for Calculation..... | 17 |
| | SQL Code For Data Normalisation | 17 |
| | R Code for corrosion rate calculation | 20 |
| | R Code for PoF calculation | 30 |
| | R Code for corrosion classification..... | 30 |

List of Figures

| | |
|---|----|
| Figure 1-1 Structural integrity will dictate what RBI/FFS/Repair actions are undertaken | 2 |
| Figure 1-2 Corrosion/Erosion is responsible for between 10% and 50% of piping failures | 3 |
| Figure 1-3 Morphology of Damage Mechanisms on pipes | 3 |
| Figure 1-4 A typical readout from a UT probe | 4 |
| Figure 1-5 Pitting will produce two responses from a single pulse | 5 |
| Figure 1-6 A response where the signal is bouncing off a defect and the back wall..... | 5 |
| Figure 1-7 Shear Wave inspection | 6 |
| Figure 1-8 A typical radiographic inspection | 7 |
| Figure 1-9 Risk Matrix from API 581 | 10 |
| Figure 1-10 Risk profile of a pipe over time projected onto the API 581 Risk Matrix | 11 |
| Figure 1-11 Multiple trains fed from the same pipeline..... | 12 |
| Figure 1-12 How corrosion rate propagates through a risk assessment | 15 |
| Figure 1-13 Typical piping isometric drawing showing the location of the CMLs | 17 |
| Figure 2-1 Bathtub curve as the sum of 2 Weibull Distributions..... | 20 |
| Figure 2-2 Risk Targets API 581 2016, each scale is logarithmic with base 10 | 24 |
| Figure 2-3 Risk curves in API 581 2008 as the categories are not consistently spaced..... | 25 |
| Figure 2-4 CML Integrity Management Data Model..... | 33 |
| Figure 2-5 Total Piping System thickness profile | 34 |
| Figure 2-6 Single Piping System thickness profile..... | 34 |
| Figure 3-1 PoF profile of a facility after 4 years vs corrosion rate | 38 |
| Figure 3-2 PoF profile of a facility after 10 years vs corrosion rate..... | 39 |
| Figure 3-3 PoF profile of 1,3 and 6-year inspection frequency | 41 |
| Figure 4-1 Visual Assessment of a Single CML..... | 42 |
| Figure 4-2 CR Distribution of Monte Carlo Simulation | 45 |
| Figure 4-3 Tc is lower than expected based on all previous readings | 46 |
| Figure 4-4 Thickness Readings Over Time | 46 |
| Figure 4-5 Neural Network to determine corrosion mode..... | 49 |
| Figure 5-1 The lognormal distribution curves for the entire dataset from the calculated and predicted corrosion rates are relatively close..... | 54 |
| Figure 5-2 All Corrosion rates vs Predicted Corrosion Rates | 54 |
| Figure 5-3 The lognormal distribution curves for the CMLs with no invalid readings..... | 55 |
| Figure 5-4 All Corrosion rates vs Predicted Corrosion Rates all valid readings | 55 |
| Figure 5-5 The lognormal distribution curves for the CMLs with no invalid readings..... | 56 |
| Figure 5-6 All Corrosion rates vs Predicted Corrosion Rates excluding readings | 56 |
| Figure 5-7 Predicted CR vs CR Including All Readings..... | 57 |
| Figure 6-1 Forecasted Risk with inspection effectiveness A..... | 61 |
| Figure 6-2 Forecasted Risk with inspection effectiveness B..... | 62 |
| Figure 6-3 Forecasted Risk with inspection effectiveness C..... | 62 |
| Figure 6-4 Forecasted Risk with inspection effectiveness D..... | 63 |
| Figure B-6-5 All Corrosion rates vs Predicted Corrosion Rates where 3 readings have been taken | 2 |
| Figure B-6-6 All Corrosion rates vs Predicted Corrosion Rates where 3 readings, all readings..... | 3 |
| Figure B-6-7 All Corrosion rates vs Predicted Corrosion Rates where 3 readings, excluding readings .. | 3 |
| Figure B-6-8 Comparison of manual and automatically determined corrosion rates | 4 |
| Figure B-6-9 The predicted lognormal distribution curve compared to the Calculated lognormal distributed curve and actual frequency of Corrosin Rates | 5 |
| Figure B-6-10 All Corrosion rates vs Predicted Corrosion Rates where 4 readings have been taken | 5 |

| | |
|---|----|
| Figure B-6-11 All Corrosion rates vs Predicted Corrosion Rates where 4 readings, all readings..... | 6 |
| Figure B-6-12 All Corrosion rates vs Predicted Corrosion Rates where 4 readings, excluding readings 6 | |
| Figure B-6-13 Comparison of manual and automatically determined corrosion rates | 7 |
| Figure B-6-14 All Corrosion rates vs Predicted Corrosion Rates where 5 readings have been taken | 8 |
| Figure B-6-15 All Corrosion rates vs Predicted Corrosion Rates where 5 readings, all readings..... | 8 |
| Figure B-6-16 All Corrosion rates vs Predicted Corrosion Rates where 5 readings, excluding readings 9 | |
| FigureB-6-17 Comparison of manual and automatically determined corrosion rates | 9 |
| Figure B-6-18 All Corrosion rates vs Predicted Corrosion Rates where 6 readings have been taken .. | 10 |
| Figure B-6-19 All Corrosion rates vs Predicted Corrosion Rates where 6 readings all readings..... | 11 |
| Figure B-6-20 Comparison of manual and automatically determined corrosion rates | 11 |

| | |
|--------------------|----|
| Equation 1-1..... | 9 |
| Equation 1-2..... | 9 |
| Equation 2-1..... | 19 |
| Equation 2-2..... | 20 |
| Equation 2-3..... | 21 |
| Equation 2-4..... | 21 |
| Equation 2-5..... | 21 |
| Equation 2-6..... | 23 |
| Equation 2-7..... | 23 |
| Equation 2-8..... | 24 |
| Equation 2-9..... | 24 |
| Equation 2-10..... | 25 |
| Equation 2-11..... | 26 |
| Equation 2-12..... | 26 |
| Equation 2-13..... | 26 |
| Equation 2-14..... | 27 |
| Equation 2-15..... | 28 |
| Equation 2-16..... | 32 |
| Equation 2-17..... | 35 |
| Equation 2-18..... | 35 |
| Equation 2-19..... | 36 |
| Equation 3-1..... | 39 |
| Equation 3-2..... | 39 |
| Equation 3-3..... | 39 |
| Equation 3-4..... | 40 |
| Equation 3-5..... | 40 |
| Equation 3-6..... | 40 |
| Equation 3-7..... | 40 |
| Equation 4-1..... | 45 |
| Equation 4-2..... | 47 |
| Equation 4-3..... | 47 |
| Equation 4-4..... | 48 |
| Equation 4-5..... | 48 |
| Equation 6-1..... | 65 |

List of Symbols and Abbreviations

AP Assessment Period in RBI

API 579 Fitness for Service Methodology

API 581 Risk-based Inspection Methodology

ASME B31.3 The most common design code for Process Piping

ASME B36.10M A construction code for pipes and pipe fittings

CA Corrosion Allowance, additional pipe wall thickness specified at design stage to account for expected corrosion over the design life

CFD Computational Fluid Dynamics

CML Condition Monitoring Location

CoF Consequence of Failure

CR Corrosion Rate of the pipe wall in mm/year

Df(t) Damage Factor as a function of Time

DL Design Life of a piece of equipment

FFS Fitness for Service

FMS Management Systems Factor

GCC Gulf Cooperation Council

Gff Generic Failure Frequency

H₂S Hydrogen Sulphide

HGL Hydrocarbon Gas Liquids

NFPA National Fire Protection Association

PoF(t) Probability of Failure as a function of time

RBI Risk Based Inspection

RCA Remaining Corrosion Allowance

RL Remaining Life of a piece of equipment

| | |
|--------|--|
| RoF(t) | Risk of Failure as a function of time |
| SCC | Stress Corrosion Cracking |
| SCH | Pipe Schedule, an industry classification for the wall thickness |
| Tc | Current Thickness |
| Tmin | Minimum Required Thickness of a pipe wall as per construction code |
| Tnom | Nominal Thickness of a pipe |
| UT | Ultrasonic Thickness Measurement |
| Φ | The standard normal cumulative distribution function |

1. Introduction

1.1. Structural Integrity Overview

1.1.1. Oil and Gas Industry

The oil and gas industry utilises pressure vessels connected by a network of piping systems to extract, separate, and refine crude oil and gas to a level which can be sent for further processing, such as cracking and reforming to ultimately produce common hydrocarbon products such as petrol, diesel, kerosene and tar[1]. In a typical extraction operation, three phase crude oil embedded in rocks below an impermeable layer is brought to the surface and directed downstream where it can be separated into its different constituents (services) and treated before being distributed to consumers.

Regardless of service, all piping and pressure vessels are designed in accordance with design codes. The most common design code for process piping is ASME B31.3. These design codes cover material selection, welding procedure, and component sizing. When sizing components, design conditions are used. Design conditions cover the worst-case scenario of pressure and temperature that can be experienced for a given piece of equipment

When considering the cost or consequence of corrosion within the oil and gas industry, there are several aspects to consider such as environmental, personal injury, financial losses due to repair and financial loss due to loss in production.

The social consequences of failure due to corrosion are most prevalent among facilities operating in areas with less regulation. Dangerous gases such as H₂S can be released into the atmosphere, if the release is ignited, the smoke can contain many dangerous sulphur and nitrogen compounds.

Heavier liquids can seep into the ground water, or, on offshore installations, oil slicks can form which are harmful to aquatic life in the region. Affected species may also enter the food chain[2].

Although there are considerable consequences related to pipes not being up to design, there can be equally large costs associated with overprovisioning pipes sizes.

By increasing each pipe wall thickness by a single size, it is estimated that the 20km of pipework in a typical plant would use an additional 350 tonnes of steel for the pipework alone. Although the heavier pipework is likely to require additional support and the fabrication and transportation costs will be increased, the cost[3] of the steel in the pipe wall alone would cost between \$225k and \$900k using historical steel prices. Therefore, simply specifying the maximum available wall thickness for each given diameter is not the most economically viable solution.

1.1.2. Structural Integrity

These piping systems are constructed to withstand conditions specified during the design stage. Once a plant is operational, the conditions are subject to change, temperatures decrease, pressures decrease, and the composition of the fluids being processed changes, e.g., water-cut increases[4]. Over time, the wall thickness in piping systems and pressure vessels reduces due to internal corrosion and erosion, and if not remedied will eventually lead to failure[5]. To reduce losses due to failure of physical infrastructure operators have structural integrity departments, the purpose of the departments is to triage the condition of all the pressurised equipment in an asset based on an inward flow of information. To parse the different sources of data effectively, a variety of standard actions are undertaken such as RBI, FFS and Repair. Figure 1-1 shows how data flows through to action to comply with the structural integrity policy. The breakdown of failure data in Figure 1-2 shows that corrosion/erosion is a major source of failures. Therefore, developing a methodology which could reduce these types of failures would be hugely beneficial.

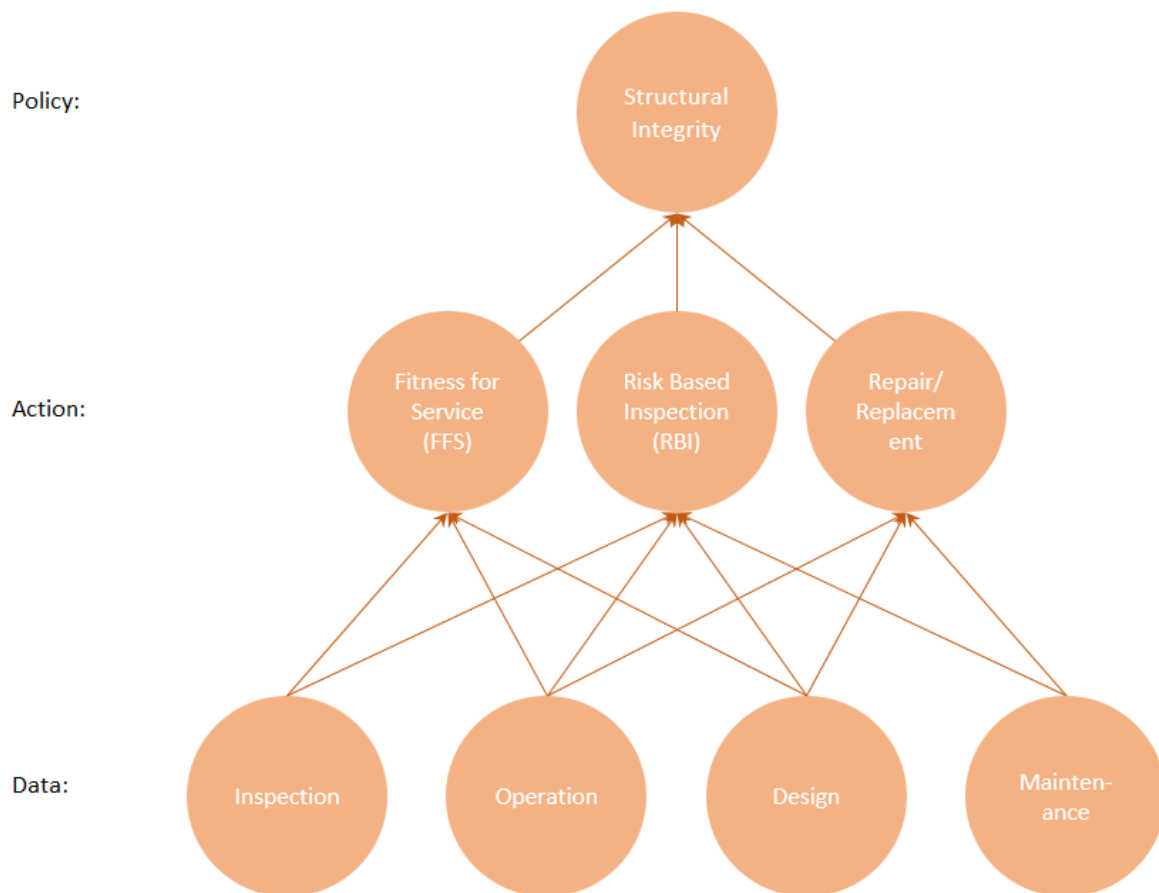


Figure 1-1 Structural integrity will dictate what RBI/FFS/Repair actions are undertaken

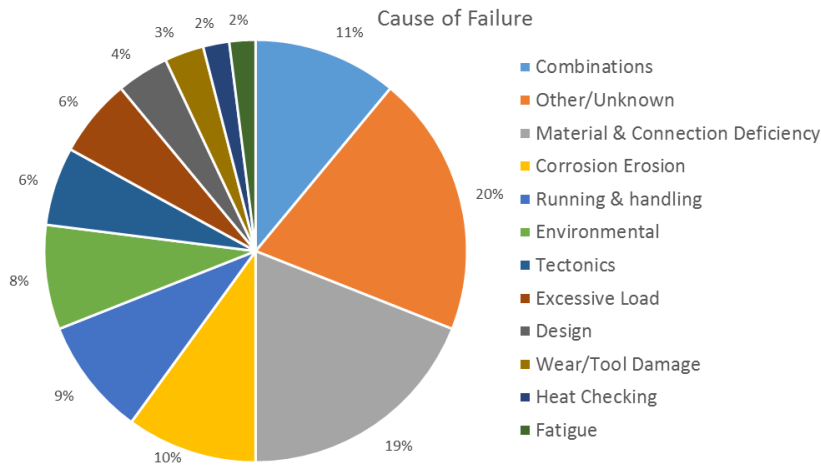


Figure 1-2 Corrosion/Erosion is responsible for between 10% and 50% of piping failures

1.1.3. Damage Mechanisms

In service piping is susceptible to a variety of damage mechanisms[6], shown in Figure 1-3. Thermal and mechanical fatigue can cause failures. This type of damage cannot be inspected for. Stress Corrosion Cracking (SCC) occurs in certain combinations of material and service and is more common at welds. These cracks cannot be seen with the naked eye and must be inspected with dyes or shear wave ultrasonic inspection. Internal wall loss can occur for a variety of reasons such as corrosion, erosion, or cavitation. Localised corrosion may be characterised with radiography. As the thickness reduces, the stress in the remaining pipe wall increases. If left unchecked, loss of containment will occur when the localised stress exceeds the ultimate tensile stress of the material.

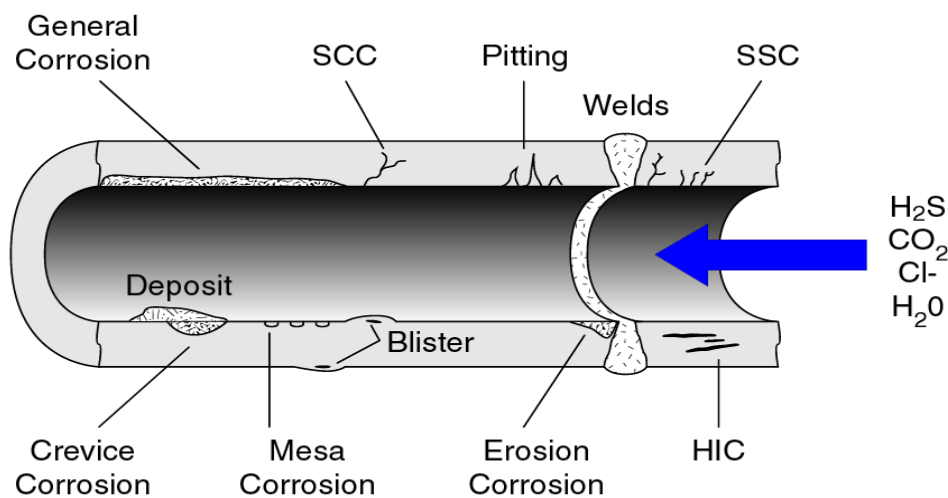


Figure 1-3 Morphology of Damage Mechanisms on pipes[6]

1.1.4. Inspection Methods

1.1.4.1. Ultrasonic Thickness Measurement

There are several methods of inspection that utilise ultrasonic emissions. Generally, a piezoelectric crystal emits a pulse, and the response is recorded. To detect cracks, the signal needs to be emitted at an angle, to detect thinning, the signal must be emitted from the probe parallel to the surface of the pipe. The signal is then received by the probe. The signal appears on the display like in Figure 1-4. A typical readout from a UT probe Figure 1-4 as a series of spikes called the back-wall echo. The wall thickness can then be determined from the distance between the backwall echoes.

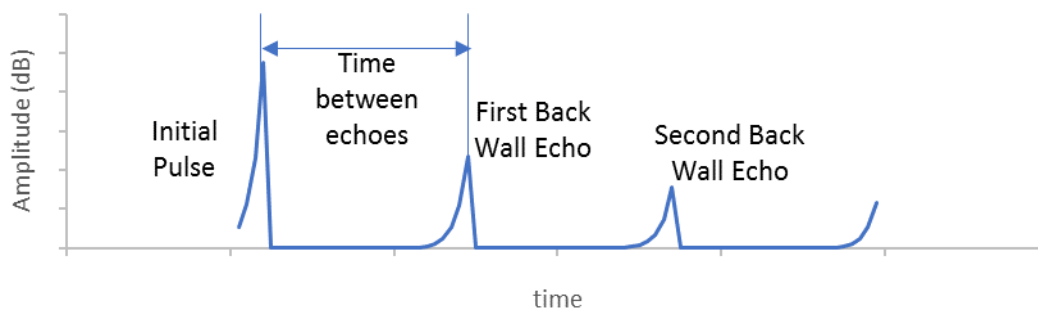


Figure 1-4 A typical readout from a UT probe

As the probe will receive a signal from any part of the pipe that is in direct contact with the probe, the response will only be a single echo when the internal surface is uniform. Where there is pitting as in Figure 1-5 for instance, the probe will receive signals at 2 different times (Figure 1-6) with the amplitude proportional to the amount of surface area with that thickness.

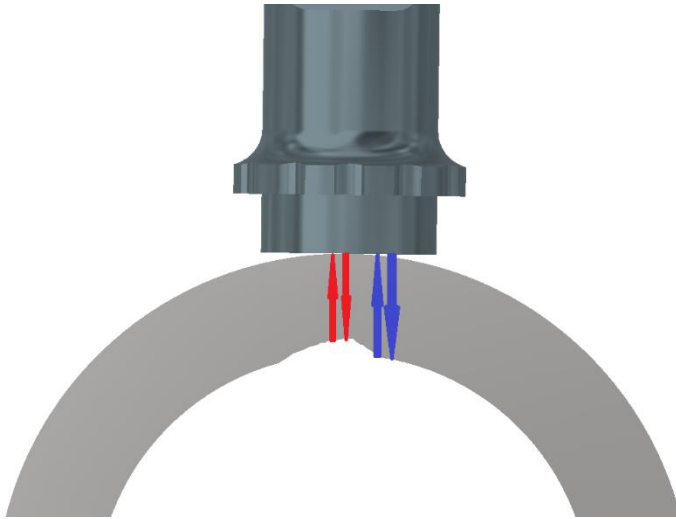


Figure 1-5 Pitting will produce two responses from a single pulse

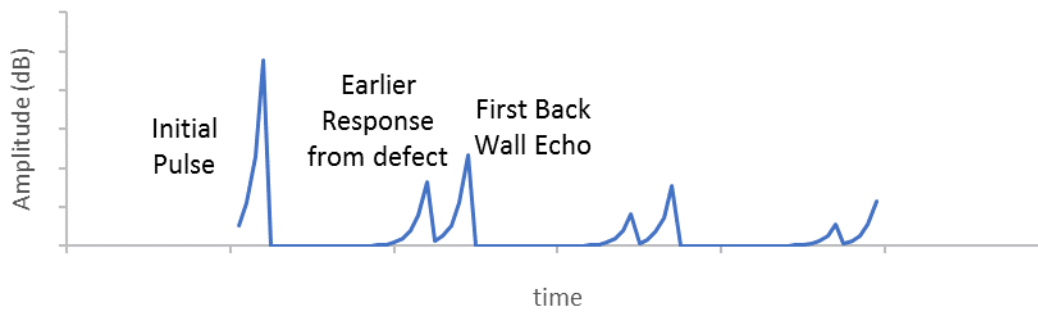


Figure 1-6 A response where the signal is bouncing off a defect and the back wall

1.1.4.2. Shear Wave

Shear wave works on the same principal as 0° UT with the addition of an angled wedge on the probe. The wave is sent at an angle and any defects normal to the probe are sent back as shown in Figure 1-7. This can pick up defects normal to the surface of the pipe[7].

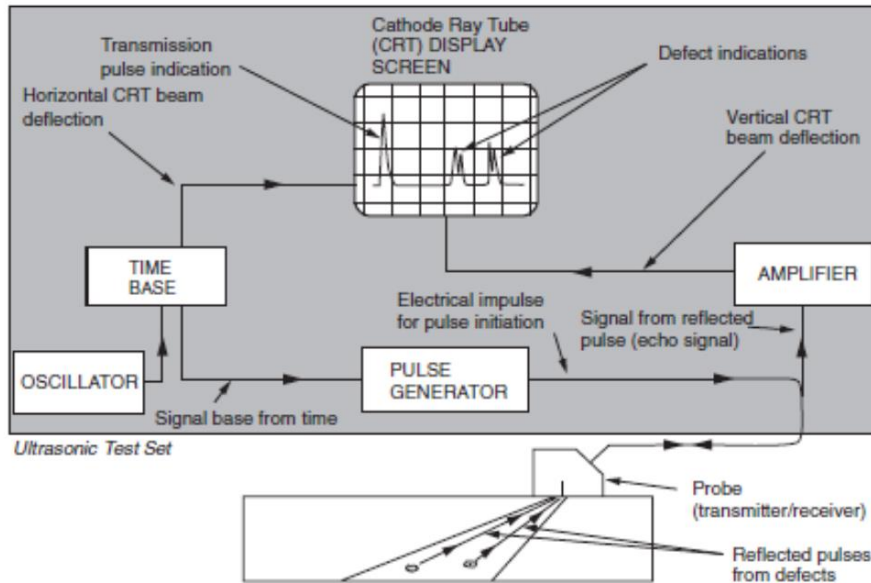


Figure 1-7 Shear Wave inspection[7]

1.1.4.3. Ultrasonic Scanning

A 0° probe can be run along a section of pipe. If the back-wall echo signal appears constant, then the pipe has uniform thickness and there is no localised internal corrosion. If piping is inspected by this methodology, then a single recorded reading is a good representation of an entire section. Thickness readings on small bore piping are less reliable[8]

1.1.4.4. Radiography

Radiography works by placing a radioactive source on one side of a location and a radiographic film on the other side. Any changes in density are then reflected on radio sensitive film as shown in Figure 1-8. A metal sphere of known diameter is placed near the area under inspection. The developed film is then interpreted by a competent inspector, any measurements or defect sizes required for future analysis will be overlaid on the image at this point. Thinning can only be characterised when the source is at right angles to it. If thinning is at 0° or 180° to the source, then the outline of the corrosion will be visible but the depth of it can't be characterised.

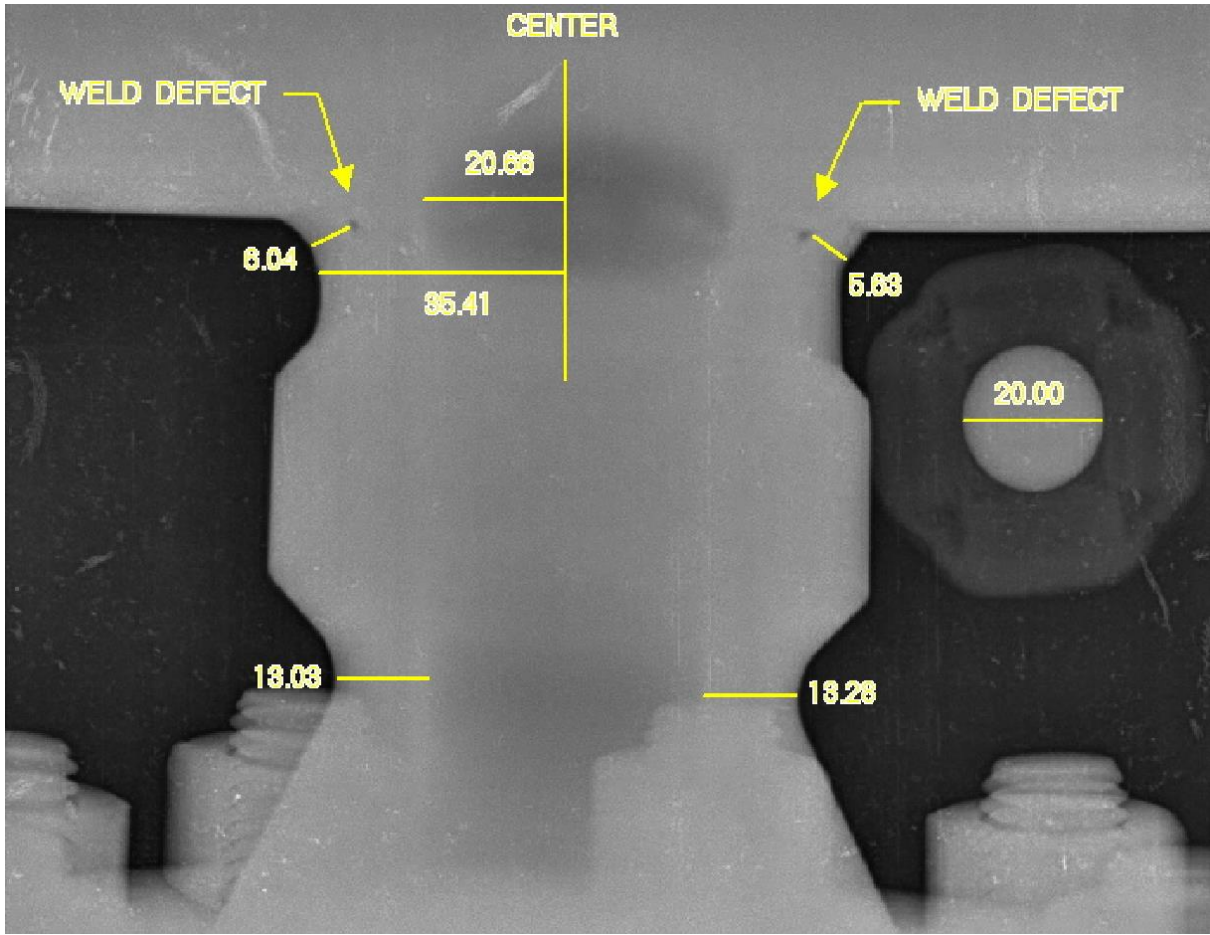


Figure 1-8 A typical radiographic inspection

1.2. Risk Based Inspection

1.2.1. Background

Historically, inspection was carried out at fixed intervals for all known damage mechanisms on all equipment. As inspection resources are limited, risk-based inspection sought to maximise the benefit of inspection resources. The theory of risk-based inspection (RBI) is using confidence gained through historic inspection of equipment to reduce to focus future inspections in areas with the greatest requirement.

This began as a qualitative approach; equipment was prioritised based on its subjective likelihood of failure and consequence of failure. As more data became available, more refined assessment techniques were developed. Table 1-1 tabulates the different levels and factors that risk is comprised of

Table 1-1 Differences between qualitative and quantitative analysis

| Type | Probability (likelihood) of Failure | Consequence of Failure |
|--------------------|---|--|
| Qualitative | Failure history, operating conditions and operational stability. | Equipment is assessed based on business criticality, health and environmental impact |
| Semi Quantitative | Remaining life is considered among other factors such as failure history, operating conditions and operational stability. | The pressure, NFPA rating, equipment location will be factored into the calculation |
| Fully Quantitative | Models developed to encompass temperature, pressure, flowrate, fluid composition, material, inspection history. | Numerical or full scale CFD models which take pressure, temperature, viscosity, pipe diameter into account are used to model the likely release pattern and potential injuries. Flowrates are also considered to determine the loss of inventory and loss of production due to downtime. |

1.2.2. Current Approach

API RBI 581 is the industry standard for fully quantitative risk of failure of pressure systems as a function of time. In this standard, Risk of Failure within a given time (t) RoF(t) is defined in Equation 1-1 as

$$RoF(t) = PoF(t) \times CoF$$

Equation 1-1

Where PoF(t) is the Probability of Failure within a given time and CoF is the Consequence of Failure.

PoF(t) is defined in Equation 1-2 as

$$PoF(t) = FMS \times gff \times Df(t)$$

Equation 1-2

FMS is the management systems factor and it is constant across an installation.

Gff is the generic failure frequency it is constant for piping.

Integrity management systems can be governed by a risk target, i.e. higher CoF items have a lower allowable PoF(t). This is represented graphically in Figure 1-9:

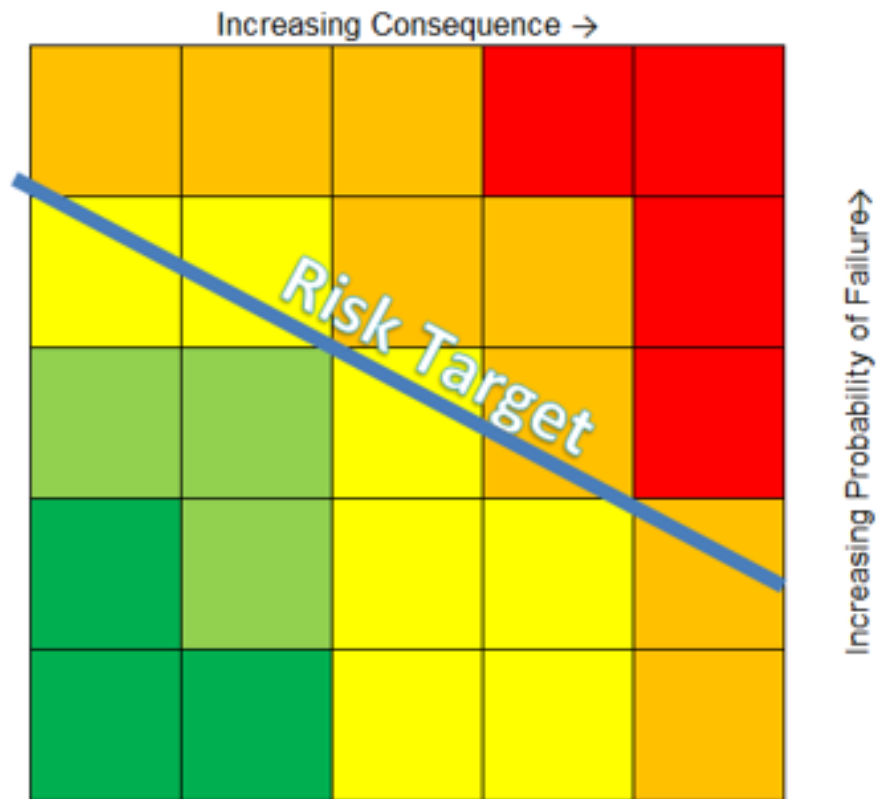


Figure 1-9 Risk Matrix from API 581

By adding time as a third dimension, time (t), the ISO Risk Line becomes a plane. Considering the risk profile as a 3-dimensional solid object, a date where the risk target is reached can be determined from the leading edge of the risk profile shown in Figure 1-10.

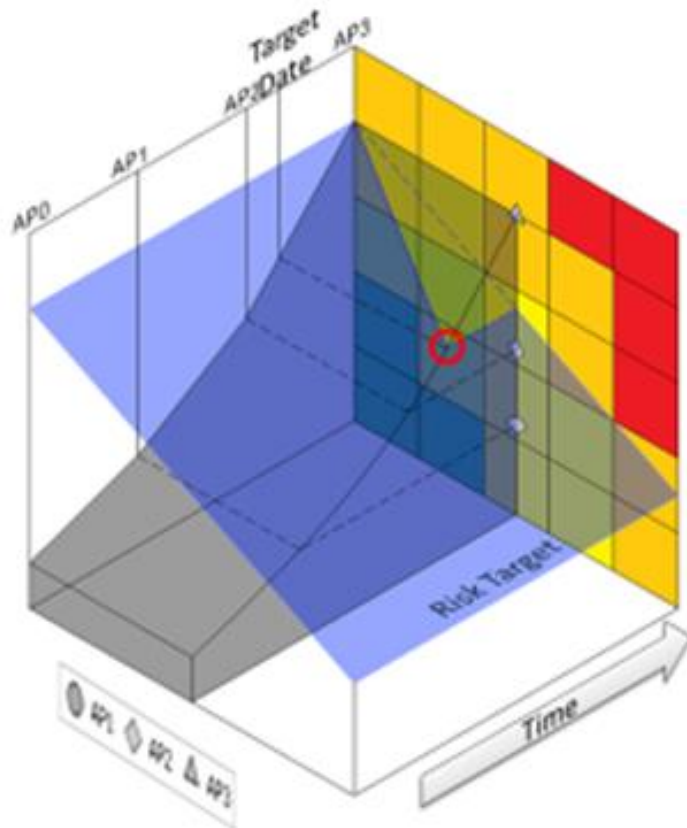


Figure 1-10 Risk profile of a pipe over time projected onto the API 581 Risk Matrix

As $PoF(t)$ increases, a date when the risk target is reached can be determined. Mitigating action such as inspection should be carried out before this date.

Historically, probabilities and consequences of failure were assigned qualitatively by experienced personnel.

Qualitative probabilities of failure ranged from

- 1) No credible failure
- 2) Unlikely failure
- 3) Possible failure
- 4) Single failure likely within 1 year
- 5) Multiple failures likely within 1 year

Qualitative consequences of failure ranged from

- 1) No production loss, small financial losses, no loss of containment
- 2) Small production loss, loss of containment onsite
- 3) Serious personal injury

- 4) Single death, loss of unit within facility, significant environmental impact
- 5) Multiple fatalities, catastrophic environmental impact, loss of facility

1.3. Experience

API 581 provides a framework for incorporating data from design, maintenance, and operation, in practice, the cost of retrieving comprehensive data can be prohibitive. Conservative estimates are often used in these situations which can result in the risk being overstated.

Further to the risk being overstated due to lack of data, the risk calculations will not get updated until the next inspection or at such a time that other information becomes available.

In instances where significant localised corrosion is discovered at a CML, the risk of that equipment may increase, however equipment upstream, downstream or in a parallel train to that CML will not be automatically updated. Upon finding significant corrosion, it is common practice for operators to conduct a special inspection program on any similar equipment susceptible to the same damage. This similarity is drawn from the operator’s experience. The special inspection program can be limited in scope and may not always target the highest priority equipment.

There is a requirement in the industry for being able to prioritise equipment for inspection based on adverse findings in similar equipment.

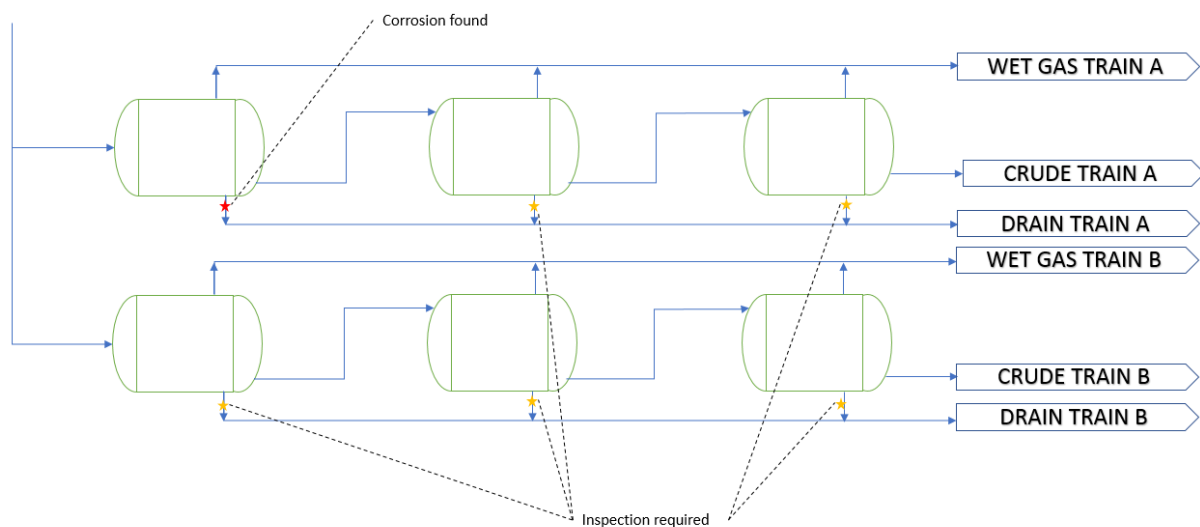


Figure 1-11 Multiple trains fed from the same pipeline

Figure 1-11 shows an example of a single crude pipeline feeding two trains, damage found at one location may warrant inspection at similar points in the same line as well as in the parallel train.

1.4. Aim

The aim of this dissertation was to develop an approach of how accumulated thickness readings can provide a feedback loop to the corrosion rate thereby updating the overall risk of failure within an integrity management system which, in turn could reprioritise the inspection schedule of an entire process or facility.

In a production environment, the corrosion rate is not the only consideration when calculating the overall risk. Figure 1-12 shows that it represents a single input to a single damage factor which in turn drives the probability and ultimately the risk of failure. As there are over 1000 individual inputs required to assess the RoF(t) of each CML, it is not practical to determine these from first principals.

Typically, parameters such as material properties and stream properties are common throughout a system and others such as environmental conditions are common to the entire facility. Even though the corrosion rate is determined at a CML level, it does not necessarily have a bearing on the overall risk result, and in cases where it does, the exact impact is not clear.

By calculating the risk on an individual CML as part of an overall system, it improves on the current method by making more use of inspection data rather than referenced standards

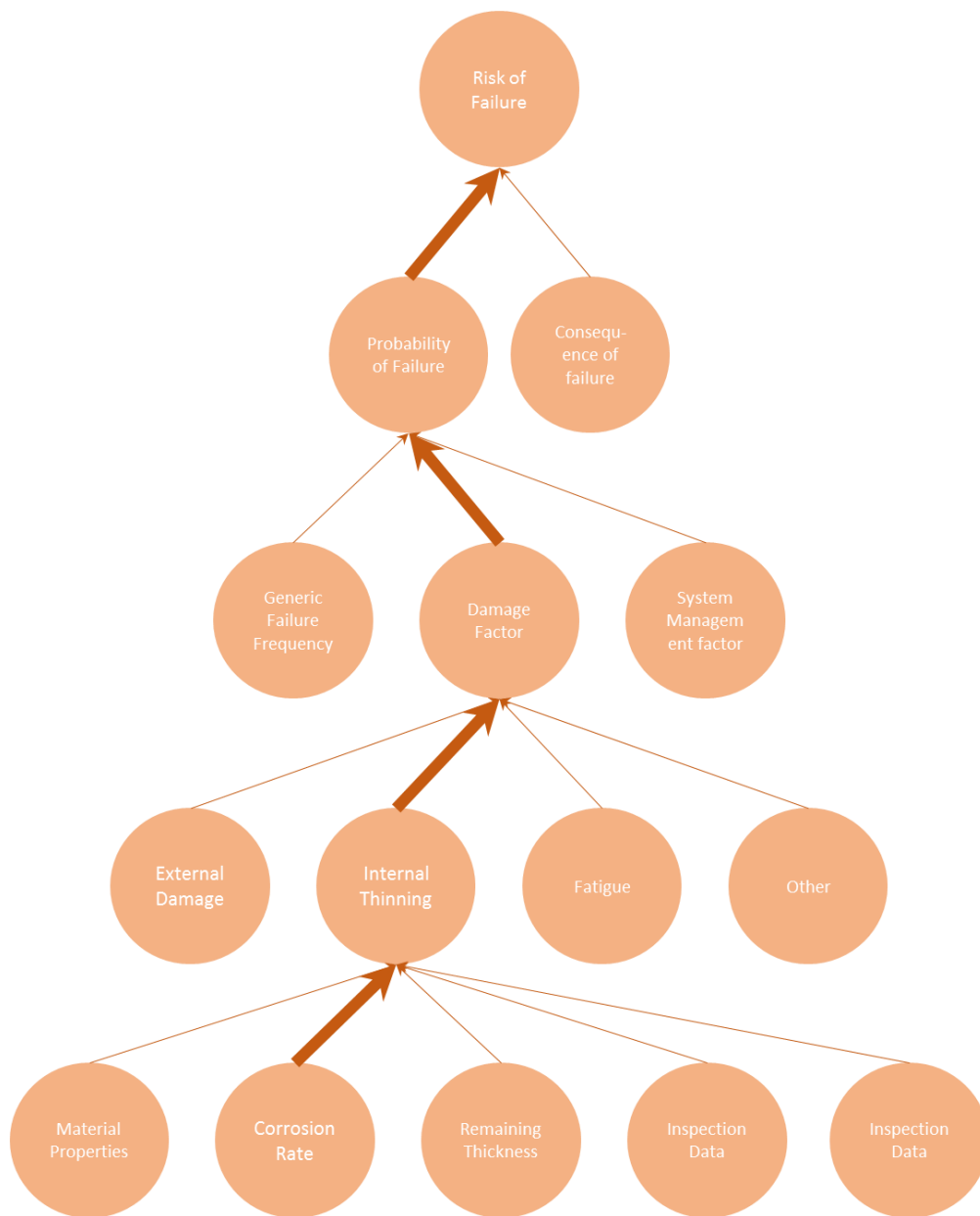


Figure 1-12 How corrosion rate propagates through a risk assessment

1.5. Objective

To achieve the aim of the thesis the first step was to develop a relational database that could accommodate disparate data sources into a universal data structure that could provide consistency and traceability back to its source.

The second objective was to test the data to determine the distribution or corrosion rates present to develop an equation that accommodated feedback from inspection results on similar equipment.

The third objective was to apply machine learning to the thickness readings to determine the corrosion rates which was previously a manual process.

The final objective was to identify any shortcomings in the new methodology by comparing the PoF(t) results from new methodology against the current and identifying any anomalies.

1.6. Contribution of this research

The contribution of this research is to develop an automated dynamic data driven approach to PoF calculation of process piping encompassing a feedback loop to update PoF calculations of uninspected CMLs. The general outline of this process is:

- 1) (Re) Calculate Corrosion rate for global dataset
- 2) Classify CML by criteria
- 3) Determine PoF from subset of CMLs with the same classification
- 4) Prioritise PoF of global dataset
- 5) Add new inspection results
- 6) Return to step 1

1.7. Condition Monitoring Locations (CMLs)

These are referred to as Thickness Monitoring Locations (TMLs) in earlier codes and standards. They are now referred to as CMLs as they used to record other data such as the external condition, this data requires a unique identifier to be stored reliably within an integrity management database.

CMLs are located at areas that give a true representative sample of the section of piping as shown in Figure 1-13, i.e. straight sections may have 1 CML, changes in direction at elbows, changes in diameter, upstream and downstream of valves etc. There are other methods such as long-range UT (LRUT) that can scan for defects but do not give an accurate wall thickness reading.

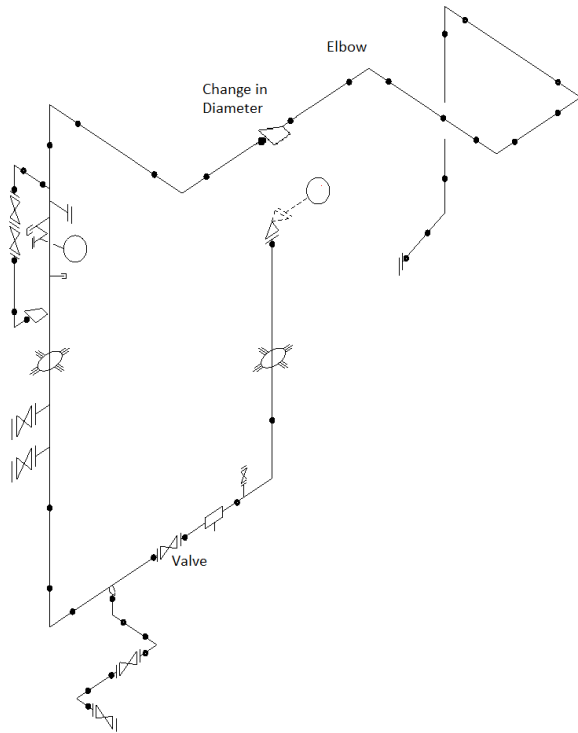


Figure 1-13 Typical piping isometric drawing showing the location of the CMLs

The wall thickness at these CMLs should be greater than T_{min} at the time of inspection and at the recommended future inspection date. The typical inspection frequencies range from six months to six years. The projected corrosion rates (CRs) are estimated by trending the historic data per CML and validating them against prescriptive rates and previously estimated CRs at the same and/or similar locations.

While CRs are not linear in nature[9], if the likely CR at a given CML can be estimated at a given time, then a recommended future inspection date can be prescribed where it is likely that, based on previous experience, knowledge of the process conditions and an understanding of the corrosion mechanisms that the predicted thickness at the CML on the future inspection date will exceed the T_{min} .

2. Literature Review

2.1. Introduction

Applying a new method to Asset Integrity Management to achieve the aims outlined in 1.4, a review of the current methods was required. Figure 1-12 shows the pathway for how corrosion rate as a

single value can feed forward to result in a change in risk, however, to add a feedback loop to this process, the requirements and purpose of each step in the calculation must be understood.

2.2. Theory of Asset Integrity Management

Asset Integrity management (AIM) for pressurised systems is broadly defined as ensuring the asset's fitness for service over its entire life from design phase through installation, design life, possible life extension and eventual decommission.[10]

There are several aspects to asset integrity management covering areas such as process, design, inspection, repair, replacement, mothballing, restarting, and decommissioning.

Oil and gas processing plants are typically run continuously and are only shut down during scheduled maintenance windows. There is considerable cost associated with equipment shutdowns, this is due to several reasons including:

- 1) Equipment shutdowns require considerable planning and monitoring. Depending on the design, the feed stock will have to go through several processes in order to achieve its final state, e.g. untreated gas may need to be dehydrated and have other impurities such as H₂S removed from it. During a shutdown, the equipment within the process will have to depressurise slowly and reach ambient temperature before it can be inspected as rapid depressurisation can cause temperature variation that could undermine the integrity of the system.
- 2) Equipment start-ups need to be reviewed by process engineers. Some equipment needs to reach a critical temperature before it is pressurised as its material may be too brittle at ambient temperature
- 3) Failure due to fatigue. This is commonly detected after start-up[11]. Small bore piping systems (≤ 2 " NPS) are particularly prone to this. This is due both to the magnitude of stress cycles involved and the vibration induced in the system by the pressurisation/ de pressurisation.

Asset integrity management programs are designed to reduce unscheduled downtime by anticipating appropriate remedial action in a timely manner.

2.2.1. Ageing Equipment

"Ageing is not about how old your equipment is; it's about what you know about its condition, and how that's changing over time " [12]

The most common mode of failure of pressurised equipment is internal corrosion. At the design stage, equipment can often have an internal corrosion allowance included in the specification.

Corrosion allowance (CA) is calculated by Equation 2-1

$$CA = DL \times CR$$

Equation 2-1

Where

DL = design life and

CR = Estimated Corrosion Rate

Equipment is inspected regularly during its design life. The measured corrosion rate may differ from the estimated corrosion rate. Where the measured corrosion rate is less than the estimated corrosion rate, the equipment may operate well beyond its original design life. Where it is found, through inspection that the measured corrosion rate exceeds the estimated corrosion rate, a review of the process and material selection may take place. Repair or replacement may also be carried out at this time.

The AIM team must be reactive enough to learn from:

- 1) Inspections
- 2) Process reviews
- 3) Failure investigations and
- 4) Root cause analyses

If an issue is not properly understood and remediated it may have several failures and unscheduled shutdowns attributed to it.

Although failure can be interpreted as loss of pressure containment or inventory, it can also mean operating beyond design limits. Whenever failure leads to an unscheduled shutdown, it is good practice to assess and repair related equipment that may be liable to cause an unscheduled shutdown in the future.

2.2.2. Inspection Frequency

Piping systems were traditionally inspected primarily for internal thinning and external corrosion. As external corrosion is visible to the naked eye it is more readily detected. Initially, the inspection frequency is prescribed based on experience of similar plants in operation. Once operation is stable, the inspection frequency may be adjusted. Construction codes[13] recommend that if the remaining

life of equipment is known then, it should, at least be inspected in half that time. As equipment nears the end of its design life the inspection frequency increases.

2.2.3. The Bathtub Curve

The bathtub curve[14] is used in asset management to describe the failure rate over time. After a plant has been commissioned, it is not uncommon to experience failures during commissioning due to oversight in design or specification. This is then followed by a long period of stable operation until the plant reaches the end of its design life. After this point, increasing remedial action is necessary to allow viable operation until such a point as the plant is decommissioned.

The shape is derived from a combination of multiple two parameter Weibull curves. The general form for Weibull probability is given by Equation 2-2

$$f(T) = \frac{\beta}{\eta} \left(\frac{T}{\eta}\right)^{\beta-1} e^{-\left(\frac{T}{\eta}\right)^\beta}$$

Equation 2-2

Where η is the scale parameter and β is the shape parameter. If $\beta < 1$, the number of failures decreases over time. This is a physical representation of failures that present at start-up. Where $\beta > 1$, the number of failures increases over time. This is a physical representation of failures due to general aging. As both types of failure are present in the system, they can be overlaid resulting in a curve that resembles a bathtub as shown in Figure 2-1.

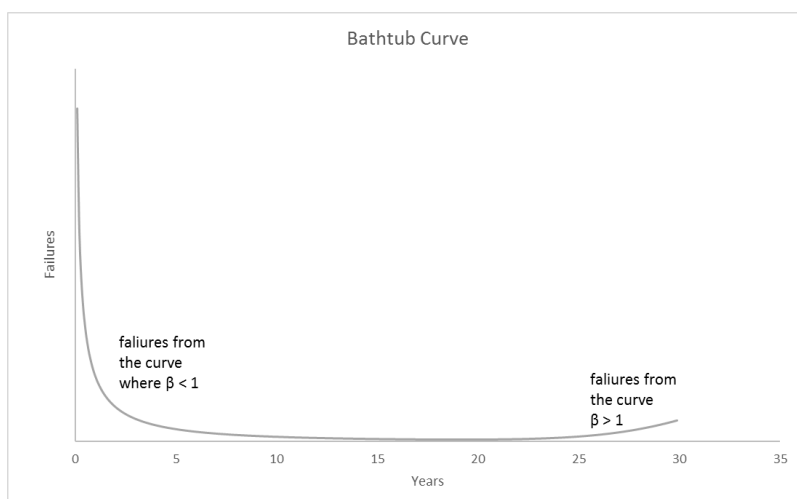


Figure 2-1 Bathtub curve as the sum of 2 Weibull Distributions

2.3. Evolution of Risk Based inspection

It could be argued that inspection was always risk based, however, the earliest attempts at codifying the risk of failure of pressurised equipment by the American Petroleum Institute (API) began in May 1993. Several API member companies including BP, Shell, Exxon, Saudi Aramco, and Texaco[15] organized the API 581 committee to develop and refine a quantitative risk assessment methodology.

The First Edition API 581 was produced in May 2000, following this, there have been a second edition in 2008 and a third in 2016.

2.3.1. Risk

The first edition laid out the basis for the Risk of failure of an item given in Equation 2-3 as

$$Risk_{item} = \sum_s Risk_s$$

Equation 2-3

Where $Risk_s$ is the risk of a scenario happening, a scenario would be the failure of a given type such as thinning, fatigue, stress corrosion cracking etc. The $Risk_s$ for each scenario in Equation 2-4 may have a different failure mode and each failure mode has a different associated consequence, for a thinning failure on piping there are 4-hole sizes considered. ¼ inch, 1 inch, 4 inch and rupture.

$$Risk_s = C_s \times F_s$$

Equation 2-4

Where F_s = Frequency Adjusted calculated by Equation 2-5.

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

Equation 2-5

The $Frequency_{generic}$ varies by hole size, for conditions it is possible for risk to be higher for a smaller hole size due to its higher $Frequency_{generic}$.

The second edition of API 581 introduces the concept of RoF(t) i.e., risk as a function of time.

2.3.2. Probability of Failure

In the first edition, Probability of Failure was calculated at a scenario level, i.e., failure due to internal thinning, fatigue, brittle fracture, and external damage.

2.3.3. Thinning Damage Factor

The thinning damage factor has evolved over the course of the three editions. The common form of equations is a base factor corrected for online monitoring, dead-legs, chemical injection. The chemical injection and dead leg factors are the same, the online monitoring factor largely the same except for the exclusion of oxidation in later editions.

In API 2000, the base thinning damage factor was determined from the A_r/t and inspection history in Table G-7.

In API 581, the base thinning damage factor was determined from an equivalent Inspection history and A_{rt} factor. Table 5.11 is essentially the same as table G-7 (excluding corrections).

The 3rd edition, however, has a completely different approach to the first two. This approach came from a need to address the shortcomings in the first two.

Shortcomings:

When specifying a pipe thickness, there may be excess thickness beyond the corrosion allowance. If this is not included in the calculation it may understate the risk.

2000, CA not directly considered

2008, where $CA = 0$, risk may not immediately take off even though the pipe is at its minimum required thickness

Prescription of $T_{min} + CA$ not clear. This can range from the code based $T_{min} +$ specified corrosion allowance, all the way up to nominal thickness. This causes considerable deviation in results, particularly on small bore, heavy wall piping.

It is found that the recommend inspection frequency can exceed the design life of a piping system, even if it is high risk.[16]

In the 2016, the approach was updated to factor the current condition into the result using Bayesian updating.

2.4. Risk based inspection

Risk based inspection is used as a means of focusing finite inspection resources on the most critical equipment. API RP581 is the most widely used recommended practice in the oil and gas industry for quantitative risk-based inspection.

2.4.1. Risk of Failure

The underlying basis for this methodology is that Risk of Failure as a function of time (RoF(t)) is determined by Probability of Failure as a function of time (PoF(t)) and Consequence of Failure (CoF) this is given by Equation 2-6

$$RoF(t) = PoF(t) \times CoF$$

Equation 2-6

2.4.2. Risk Matrix

Risk is displayed on a risk matrix, risk matrices vary between organisations, however, they will display PoF and CoF on their axes and they are generally represented on a logarithmic scale.

2.4.3. Assessment Periods

To represent the time dimension on a 2D Risk matrix, the PoF(t) is generally assessed at three points of interest.

0AP this is the current date of assessment

1AP this is the date of assessment when the inspection is due

2AP this is the subsequent inspection date

Where the RoF(2AP) is deemed to be acceptable, inspection at 1AP may be skipped depending on the constraints of the integrity management programme.

2.4.4. Risk Target

The concept of a risk target, i.e., a defined maximum level of risk is already established at a qualitative level. This prioritises inspection of equipment depending on its risk category.

For quantitative risk assessment, given that a risk matrix has a logarithmic scale, a risk target, i.e. RoF is constant can be displayed as a straight line on the risk matrix, this is shown by transforming Equation 2-7 to Equation 2-8 and Equation 2-9

$$RoF = PoF \times CoF$$

Equation 2-7

$$\log_a(RoF) = \log_a(PoF) + \log_a(CoF)$$

Equation 2-8

$$\log_a(PoF) = -\log_a(CoF) \times \left(1 + \log_b \frac{a}{b}\right) + \log_a(RoF)$$

Equation 2-9

Where b is difference between the PoF categories and c is the difference between CoF categories

As this equation takes the form $y = mx + C$ it will have a slope of -1 provided the logs have the same base.

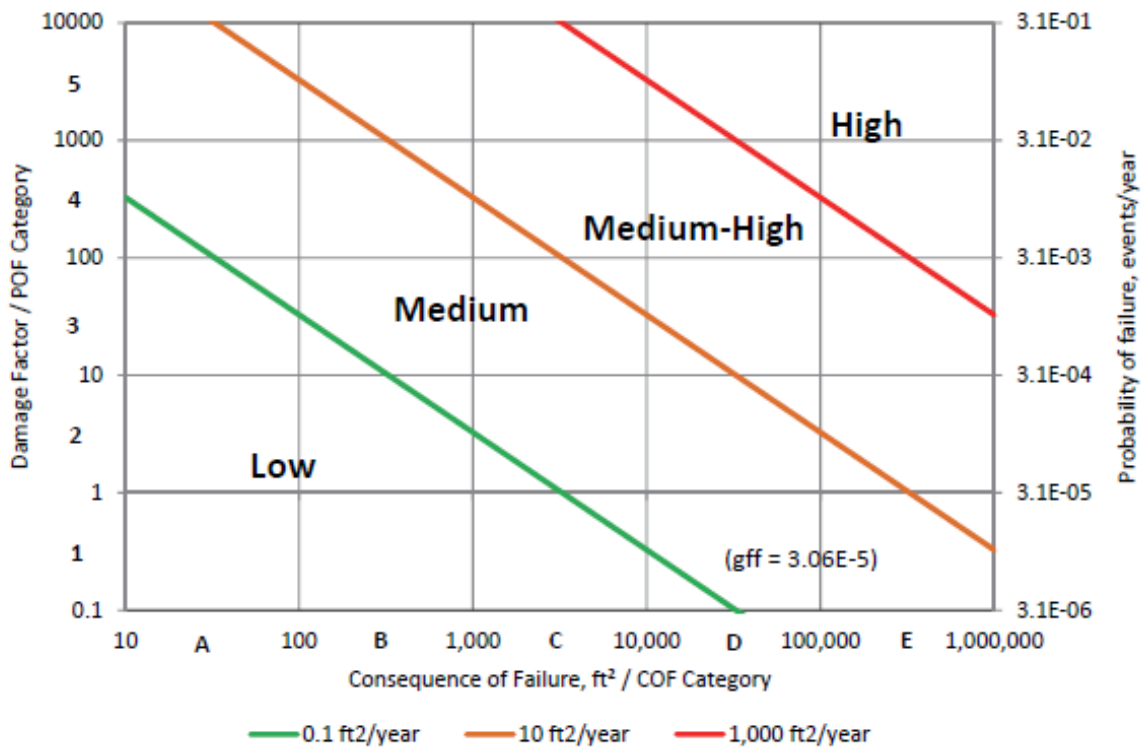


Figure 2-2 Risk Targets API 581 2016, each scale is logarithmic with base 10

Where the boundaries do not increase logarithmically, e.g. API 581 2008 the iso risk line becomes an iso risk curve as shown in Figure 2-3.

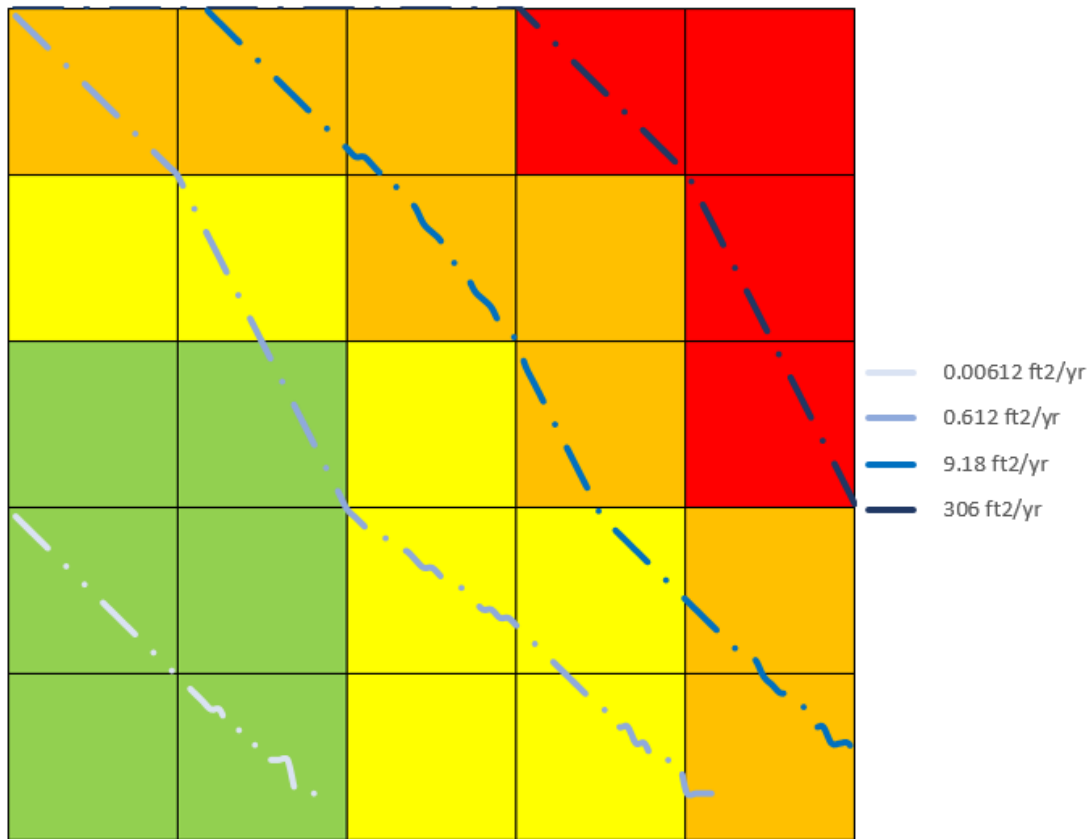


Figure 2-3 Risk curves in API 581 2008 as the categories are not consistently spaced

As CoF is constant with respect to time for a given CML, the maximum allowable probability is determined from Equation 2-10

$$PoF_{target} = \frac{Risk_{target}}{CoF}$$

Equation 2-10

2.5. API 581 Approach to internal thinning

API 581 is the leading quantitative methodology for determining the probability of failure within the oil and gas industry.

- 1) PoF(t) must be calculated for discrete values of t, this can be readily computed with an appropriate software, however, if input data is incomplete it may cause large deviations in the result.
- 2) The calculation methodology considers failure as loss of containment. Many integrity management programs define failure as operation outside of design limits[17]. Furthermore, data management systems may just provide the assessor with a single minimum required thickness to compare against.
- 3) There are several factors that are applied throughout the methodology that may be based on conservative assumptions. Assessments usually apply the worst-case conditions to the equipment being assessed.

Probability of failure as a function of time is determined from Equation 2-11

$$PoF(t) = gff \times FMS \times Df(t)$$

Equation 2-11

Where Gff is the generic failure frequency based on the equipment type, for all pipes it equals 3.06×10^{-5} /yr.

F_{MS} is the management system factor; this is calculated on a plant wide level and ranges in value from 0.1 to 10.

D_f(t) is the damage factor as a function of time. For standard unlined pipe, API 581 uses Bayesian updating and posterior probabilities, determined by referencing previous inspection results in standardised tables to determine the future base damage factor given by Equation 2-12.

$$D_f^{Thin}(t) = \max \left[\left(\frac{D_{fB}^{Thin}(t) \cdot F_{IP} \cdot F_{DL}}{F_{OM}} \right), 0.1 \right]$$

Equation 2-12

Where the base factor D_{fb}^{thin}(t) is given by equation Equation 2-13

$$D_{fb}^{Thin}(t) = \sum_{i=1}^3 \frac{PO_{pi}^{Thin} \times \Phi(-\beta_i^{Thin})}{1.56 \times 10^{-4}}$$

Equation 2-13

Where Φ is the standard normal cumulative distribution function.

Po^{Thin} is determined from conditional probability of existing thinning, β_i is calculated using Equation 2-14:

$$\beta_i^{Thin} = \frac{1 - D_s \cdot A_{rt} - SR_P^{Thin}}{\sqrt{D_{s_i}^2 \cdot A_{rt} \cdot COV_{\Delta t}^2 + (1 - D_{s_i} \cdot A_{rt})^2 \cdot COV_{s_f}^2 + (SR_P^{Thin})^2 \cdot COV_p^2}}$$

Equation 2-14

The following values are recommended although they can be determined for a given system:

$$COV_{\Delta t} = 0.2$$

$$COV_{s_f} = 0.2$$

$$COV_p = 0.05$$

The equation is sensitive to $COV_{\Delta t}$ variation.

Given that both F_{ip} and F_{dl} are constant the final damage factor is a function of the base damage factor. The weighting, in a worst-case scenario would increase the PoF by 1 level. This may not be suitable for piping systems that don't experience predictable linear corrosion. Localised corrosion may not affect the COV sufficiently in the correct area.

The inspection effectiveness was held constant as plants typically have standardised thickness inspections. Where supplementary inspections take place, they are generally inspecting a damage mechanism such as stress corrosion cracking that would not normally be detected by conventional thickness measurement.

As the $FMS \times gff$ remains constant, the variability of the corrosion rate is the only significant factor affecting the $PoF(t)$.

The PoF according to API581 is ultimately determined by the corrosion rate, however with the absence of empirical thickness data, this can be determined through published rates or lab measurements. It is stated that general corrosion rarely exceeds four times the expected amount, however, localised corrosion may be significantly higher.

If an integrity management programme of recording ultrasonic thickness readings has been established, equal confidence can be applied to historic readings. The methodology prescribes assigning high confidence to the prior probabilities, lower confidences are reserved for published rates and comparative lab results.

As part of the calculation, three damage states are considered.

As expected: although API 581 does not quantitatively prescribe what constitutes each damage state, for the purpose of this comparative assessment, the damage state has been categorised based on the actual corrosion rate CR_a .

Given that the corrosion allowance (CA) is based on the design life (DL) (typically 20 years) the expected corrosion rate CR_e can be determined from Equation 2-15

$$CR_e = \frac{CA}{DL}$$

Equation 2-15

Damage state1: $CR_a \leq CR_e$

Damage state2: $CR_e \leq CR_a \leq 4 \times CR_e$

Damage state3: $4 \times CR_e \leq CR_a$

Three confidence levels are considered:

Low confidence applies to published data

Medium confidence applies to lab data

High confidence applies to direct measurement

To apply this methodology in a manner that updates the risk at an uninspected cml (CML_u) given that elevated damage was found at a similar location (CML_d), the following assumptions must be made:

- 1) the previous damage state of CML_u was p1, i.e., undamaged, otherwise, it is likely to have been repaired under an integrity management program.
- 2) The elevated damage at CML_d is either p2 or p3.
- 3) The confidence that this damage could apply to CML_u is medium as comparable in-service conditions approximate lab conditions.
- 4) There are no other low or medium confidence data applied to the assessment.

2.6. ANOVA in the Nuclear Industry

The nuclear industry, and more generally, the power generation industry has developed their own procedures for dealing with internal corrosion on secondary systems. These are more transferrable as the consequence of failure is comparable. Primary systems in the nuclear industry have a far more intense inspection regime than would be practicable in oil and gas.

H. Yun et al[18] discusses the methodology of determination of thinning as either present or not. In this paper, the null hypothesis is that there has been no internal thinning. In general, given a known error in measurement, a Monte Carlo simulation was applied to the results of grid pattern. The paper then compares the results and threshold for several different methods of analysing the readings including

- 1) analysis of variables (ANOVA)
- 2) Total Point Method (TPM)
- 3) Locational Deviation Method (LDM)
- 4) Binomial variables uniformity test (BVUT)
- 5) Continuous variable uniformity test (CVUT)

Where the grid pattern was on an elbow where the thickness could not be considered uniform, the measurements were taken along a line with the same relative polar coordinates.

For each subset of data, the mean, standard deviation, and variance could be calculated. Factoring in the known error in measurement could then prove to a degree of confidence if thinning had occurred.

For a section, if the null hypothesis was rejected, i.e., where thinning had occurred between two inspections, corrosion rate was taken as the slope of the readings, for greater than two, a least squares approach was used to find the slope.

In practice, this could be implemented in the oil and gas industry, however, inspectors are usually able to determine if there is localised corrosion present from the backwall pattern of the A scan results.

2.7. Norsok Models v API 581 models

There were several published methodologies developed by NORSOK for calculating corrosion rates. An example is NORSOK M-506 for calculating corrosion rates based on internal pressure, temperature, pH, CO₂ fugacity and partial pressures of H₂S in carbon steel pipes.

There have been several studies performed comparing the predicted corrosion rates with the observed.

The paper Coupling Norsok Co₂ Corrosion Prediction Model with Pipelines Thermal/Hydraulic Models to Simulate Co₂ Corrosion Along Pipelines[19] dealt exclusively with pipelines; therefore, the results would only be applicable to straight sections of piping. There are several factors that have considerable effects on corrosion rate that are not found in straight sections of piping such as:

Microbial action found in dead legs.

Excessive corrosion found on the extrados of elbows.

The flow state can change at fittings, the difference in corrosion rate between turbulent and laminar flow is discussed in this paper[20].

Corrosion due to cavitation can often be found on eccentric reducers on the suction side of the pump.

In the above study, temperature, pressure, and viscosity were modelled as a gradient.

The temperature gradient alone could account for double the difference in corrosion rate.

The results of the study found:

Both methodologies are commonly used in the specification of corrosion allowance during the design stages of piping systems. As they are a guide to general corrosion of a piping system, they do not give enough insight into local areas where variations in temperature, pressure and water content exceed the stated amounts in the process flow diagram. Packages such as CAESAR can calculate local temperatures in a piping system. Normal operating conditions can range from ambient temperature and pressure up to design pressure. In a separator, which is typically the first pressure vessel feedstock encounters in processing, the feed line will be operating close to the design temperature and pressure. The drain piping will operate intermittently at a lower temperature and pressure.

Localised corrosion does not invalidate these scientific models for corrosion. If the models are retrospectively applied to areas where consideration is given to localised conditions such as the

presence of CO₂ in water in a dead legs[21] in gas piping systems, then the actual corrosion rate can be reconciled with these models. However, short of measuring these conditions on site they can be missed.

2.8. Failure investigations

In most industries where a critical engineering component fails in service, a failure investigation is carried out. Depending on the type of failure e.g., fatigue, creep, brittle fracture, corrosion or stress corrosion cracking, there are an array of tests that can be performed on the failure specimen.

A comprehensive internal thinning inspection programme will often catch corrosion before loss of containment, the affected area will then be replaced. Even when loss of containment does not occur, it is common to perform failure analysis on the replaced section (failure specimen) to gain a greater understanding of the damage mechanism active in the system.

Failures can occur from either a single mechanism, or through a combination of mechanisms. Where internal corrosion is deemed to be a factor, the microscopy or environmental analysis may be performed on the failure specimen.

The outcome of the failure analysis will be mitigating action to prevent future corrosion of this nature. These mitigating actions include:

Modification of operating parameters[22]

Increased monitoring at susceptible areas

Addition of corrosion inhibitors.

2.9. Distribution of Corrosion Rates

There are many articles surrounding the probability of detection[23] [24] of corrosion given its presence and the probability of failure given a linear corrosion rate. There is, by comparison, relatively few papers that discuss the distribution of corrosion rates of a system.

The global dataset was tested for normality and lognormality. The results contained in APPENDIX C show that Weibull distribution is the closest approximation. Normal distribution tends to underestimate the corrosion rate and lognormal tends to provide a reasonably conservative estimate. Ultimately lognormal distribution was ultimately chosen as it provided the most suitable conservative values with a minimal amount of data.

This chapter will outline how this methodology could be used to update risk by applying the knowledge that damage was found at a similar location. However, from a top level there are fundamental differences that make this methodology difficult to apply in practice.

The suitability of a distribution can be determined by its ability to closely model the frequency of occurrence of a corrosion rate one standard deviation beyond its expected value.

As determining the probability of failure relies on the distribution of corrosion rate. A variety of distribution types were examined in this study. Ultimately Log Normal distribution was selected.

There are several sources[25] that cite internal corrosion of piping systems as a Weibull distribution. A modified Kolmogorov/Smirnov test was applied to each piping system individually and to the global data set. It was found that they did not fit a normal or lognormal distribution. The results of the quantile/quantile (Q-Q) plot are available in APPENDIX C. It was notable that the projected probability of a corrosion rate tended to be under/overstated above the mean.

Other distribution types were considered but not used for various reasons. Gumbel distribution could be used to predict the maximum expected corrosion rate given data inputs.[26] As this itself is dependent on a confidence interval, it does not give probability of failure in exact terms.

Extreme value theory is used on a global set. A tail weighted maximum likelihood is calculated by [27] this method. This distribution was found to model the tail of the distribution.

Weibull distribution is applied in reliability engineering at predicting failures above a certain threshold. Binomial distribution[28] can be used to calculate the overall PoF of a piping system if there is a known CR limit in place. Poisson Distribution can be used to determine the number of expected failures within a given time frame. The general form is given in Equation 2-16

$$P_{x_t}(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

Equation 2-16

However, in practical terms, Poisson distribution reverts to the mean when applied to a single failure in a single year. It can, however, be used to measure current failure data over the historic failure data to determine if the generic failure rate is increasing.

For a given piping system, previous observations of corrosion rates at that location and other representative locations can be used to determine a probabilistic distribution of the future corrosion rate at that point.

2.9.1. Distribution Selection

CDF graphs for each individual piping system were prepared and Weibull and lognormal curves were fitted over the data. These plots are available in APPENDIX C . The overall results are tabulated below. Overall, the risk target date tended to be consistent between two parameter Weibull and log normal. The results diverged as the corrosion rates increased, however, they were beyond the practical limits of what corrosion rates are expected in the industry overall.

2.10. Determining PoF(t) of internal corrosion

2.10.1. Dataset

The chosen dataset for the corrosion rate classification was a sour service gas unit. Data was generated from the global dataset using the data model shown in Figure 2-4:

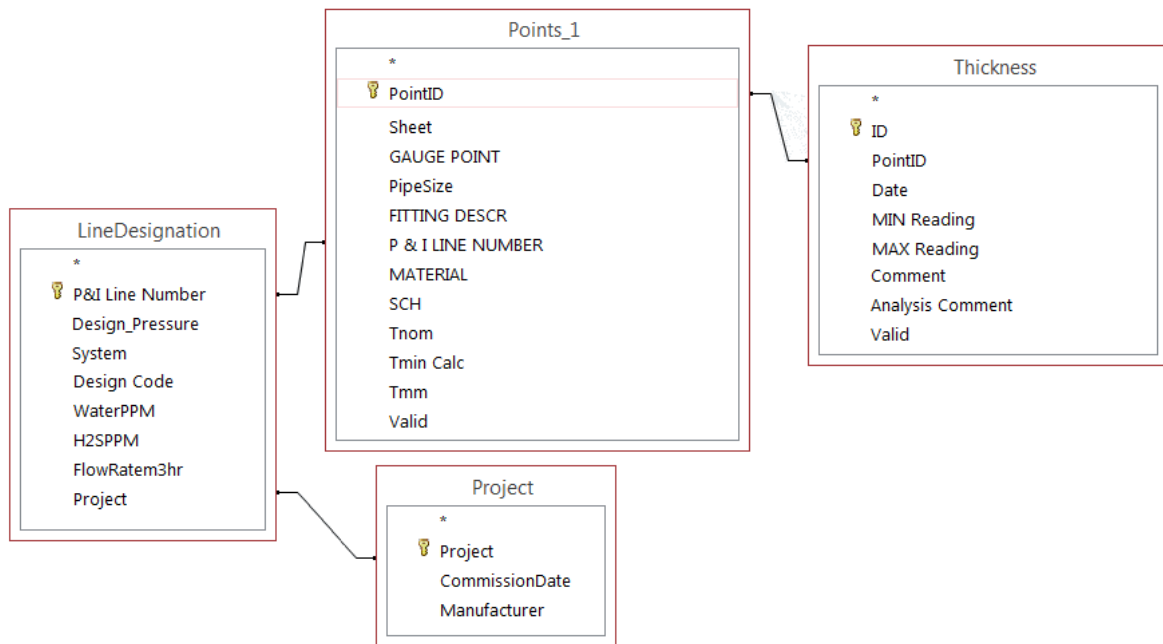


Figure 2-4 CML Integrity Management Data Model

A typical summary of a piping system would look like Figure 2-5. T_{nom} , T_{min} and T_c are all variable over the entire system. This is due to the variation in nominal diameter and design pressure. It is not uncommon for T_c to exceed T_{nom} as pipes have a mill tolerance of 12.5%.

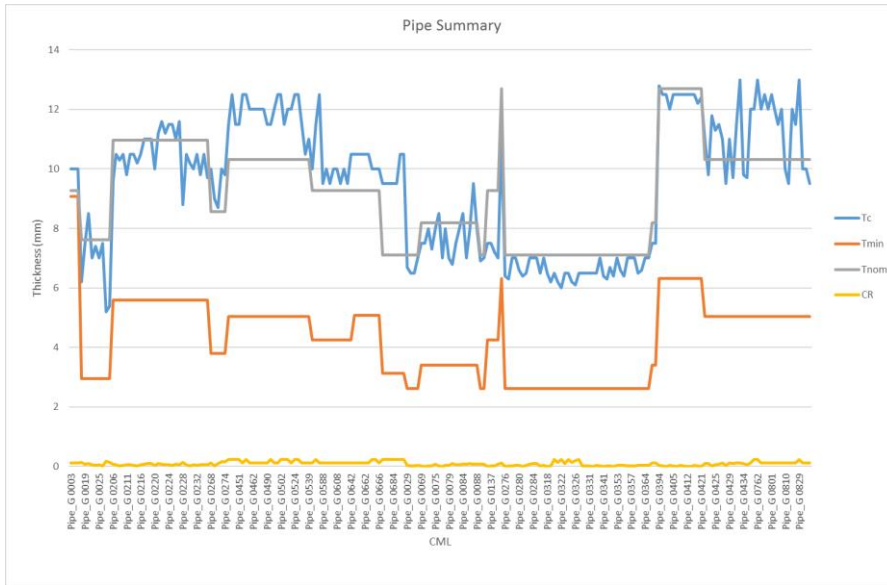


Figure 2-5 Total Piping System thickness profile

By focusing on a single section of pipe with a constant T_{nom} and T_{min} any abnormality can be seen more clearly as shown in Figure 2-6. This section can be treated as a set

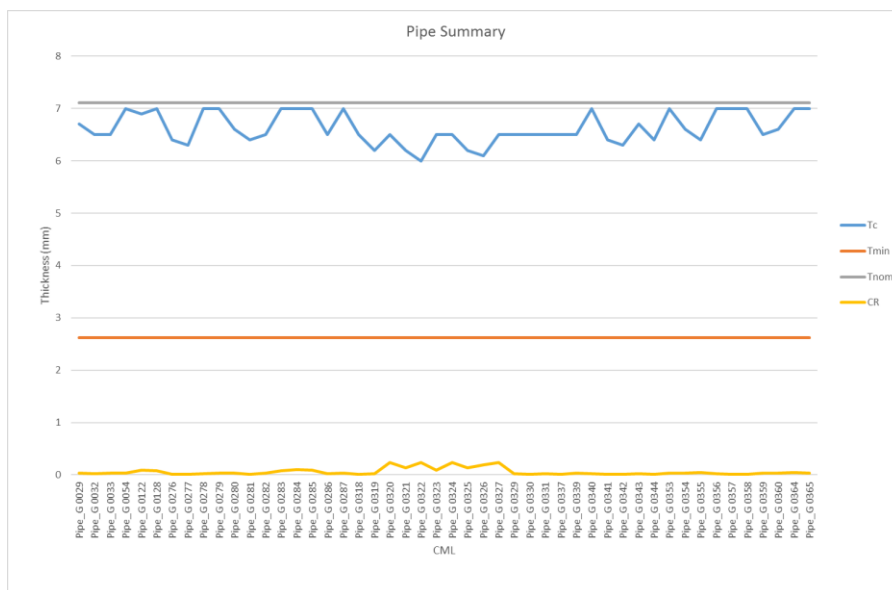


Figure 2-6 Single Piping System thickness profile

2.10.2. Remaining Life

The concept of the remaining life of a CML appears in many early versions of integrity management practices[29] and as a footnote in construction codes. The general form of remaining life equations is given by Equation 2-17

$$RL = \frac{RCA}{CR}$$

Equation 2-17

Where:

RL = Remaining Life in years

CR = Corrosion Rate in mm/year

RCA = Remaining Corrosion Allowance in mm

Minimum Required Thickness

Tmin is determined by the integrity management system in place. A common value for Tmin is the design code pressure requirement Tmin. These equations are generally a variation of Equation 2-18

$$t_{min} = \frac{P \times D}{S \times E}$$

Equation 2-18

Where

P = Design Pressure

D = Diameter

S = Allowable Stress of the material for a given design code

E = Joint Efficiency

Integrity management systems treat failure as operating beyond the permitted envelope, even though an actual loss of containment may not occur. When equipment is found to be operating in this condition, there are approaches, most notably API579 which make use of redundancy in pipework design, it is possible to operate where the RCA is < 0, however, where these conditions are found, mitigating action are generally carried out as soon as practicably possible.

There are other probabilistic approaches where the Tmin is treated as a distribution, due to the variations of the Ultimate Tensile Strength (UTS) of the steel. [30]

For the purposes of this study however, the minimum required thickness will be treated as a predetermined scalar value calculated from ASME B31.3.

Remaining Corrosion Allowance (RCA)

The RCA at a given inspection point can be determined by Equation 2-19 from the current thickness (T_c) and the minimum required thickness (T_{min})

$$RCA = T_c - T_{min}$$

Equation 2-19

There are many ways to increase the RCA at a given CML such as:

- 1) Reinspect, this may increase T_c if the previous inspection result was incorrect
- 2) Repair, this increases the value of T_c
- 3) Rerate, this reduces P thus reducing the T_{min}
- 4) Replace the section this increases the T_c

Although the RCA can be increased, for the purposes of this study, it will be taken as the difference between T_c and the design code T_{min} .

2.11. Machine Learning

Machine learning covers a broad range of algorithms that predict results based on certain inputs. The difference between machine learning and standard predictive models is that predicted results from machine learning can change over time as without manual reassessment. This process is known as training. The predicted results may be in the form of numeric output, e.g. corrosion rate in mm/year or a classification e.g. corrosion type, which is discussed in chapter 4.2.

The research conducted as part of this thesis was to demonstrate that machine learning could be used to supplement manual methods of PoF calculation. Due to the quality of data available, it was not possible to explore more sophisticated models than those outlined in this section.

To determine the most suitable type of machine learning for determining corrosion rates and corrosion type classification, the data acquisition model and required outputs must be considered. Artificial Neural Networks (ANN) provide a methodology for obtaining a numeric output from a set of known inputs.

Supervision is a term used in machine learning that trains an algorithm to give a known output. Manually calculated corrosion rates based on a linear least squares regression model has known inputs and known numeric outputs and can be trained using supervised ANN.

Unsupervised learning is useful for pattern classification and already has many applications in the oil and gas industry such as anomaly detection in welds[31], an image is scanned for features which are then classified together.

Despite developing a universal data model for the CMLs, the data may be incomplete for a variety of reasons. Because of this, multiple classification algorithms may be run on the data. Where more than one classification algorithm may be available, a Random Forest may be used to determine the overall classification of a feature.

2.12. Summary

From looking at the flow of corrosion rate estimation through to risk of failure in Figure 1-12, each step appears to be contained, i.e. each level passes a scalar number to the next level up with no feedback mechanism. By combining the uncertainties into a single equation, using a supervised ANN to calculate the corrosion rates and an unsupervised classification algorithm to group similar CMLs together the aim of creating a POF feedback loop can be achieved.

3. Data Structure

3.1. Introduction

The data structure has borrowed common elements from Pipeline Open Data Standard database (PODS) and asset management tools such as Asset Condition Evaluation Tool (ACET). There were a number of critical factors that were not available which are known to have an effect on corrosion rates such as flowrate, water content, pipe orientation and operational continuity. Were this data readily available, it could be incorporated into the data model.

3.2. Risk Based Inspection

In API581, PoF(t) is driven by several factors including fatigue, cracking and fracture. In these calculations, the corrosion rate is a scalar input and is independent of the entire system, therefore updating the PoF(t) calculations are done manually and typically not until the next inspection is carried out.

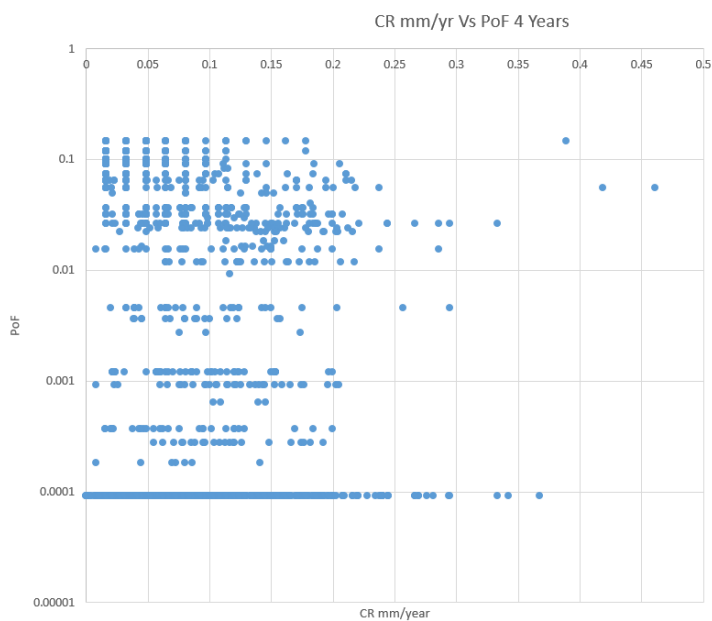


Figure 3-1 PoF profile of a facility after 4 years vs corrosion rate

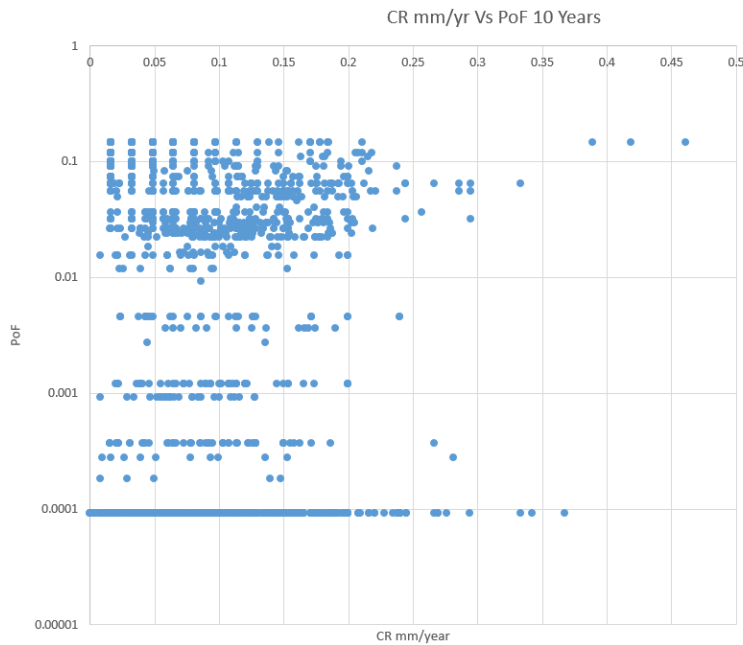


Figure 3-2 PoF profile of a facility after 10 years vs corrosion rate

Figure 3-1 and Figure 3-2 show the corrosion rate (Cr) vs PoF for the facility, note that the highest CR does not immediately translate to a higher PoF but over time the general trend is true.

3.3. PoF Calculation

Probability of Failure

The limit state equation for failure given in Equation 3-1 is by definition

$$RL \leq 0$$

Equation 3-1

Therefore, the PoF at a time t in the future as shown in Equation 3-2 as

$$PoF(t) = P(RL - t \leq 0)$$

Equation 3-2

Substituting Equation 2-19 into Equation 3-2 gives Equation 3-3

$$PoF(t) = P\left(CR \geq \frac{RCA}{t}\right)$$

Equation 3-3

Taking {C} a set of CMLs {CML₁, CML₂... CML_n} that have been classified according to similarity in their expected corrosion rates then be defined in terms of a mean \bar{x} (Equation 3-4) and standard deviation σ (Equation 3-5)

$$\bar{x} = \sum_{i=1}^n \frac{\ln (CR(CML_i))}{n}$$

Equation 3-4

And

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - \ln (CR(CML_i)))^2}{n}}$$

Equation 3-5

Then the likelihood of a corrosion rate being exceeded can be determined by Equation 3-6

$$P(CR \geq x) = 1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

Equation 3-6

Where Φ is the cumulative distribution function of the standard normal distribution.

Combining all the equations together, for a thickness t_c , within the set the PoF(t) can be determined by Equation 3-7

$$PoF(t) = 1 - \Phi\left(\frac{\ln\left(\frac{t_c - t_{min}}{t}\right) - \mu}{\sigma}\right)$$

Equation 3-7

3.3.1. Effect of inspection on PoF

From the time of inspection, for a constant RCA and a known corrosion rate distribution, a continuous PoF curve can be drawn.

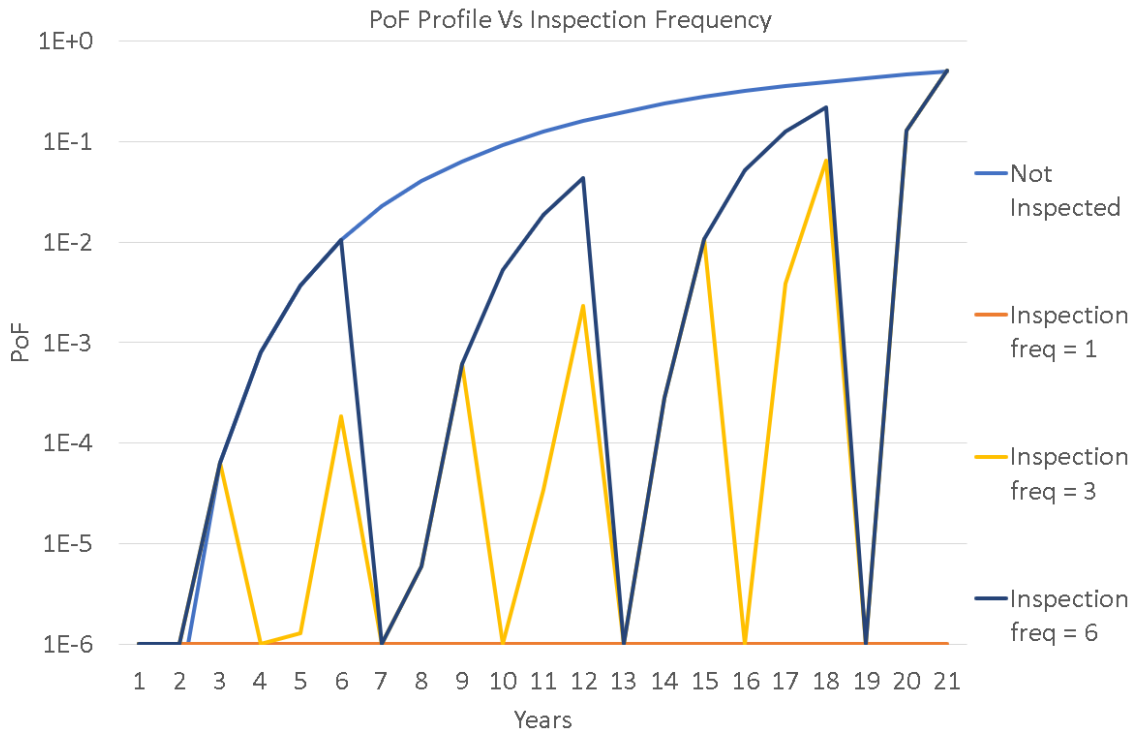


Figure 3-3 PoF profile of 1,3 and 6-year inspection frequency

Figure 3-3 is the idealised PoF profile of a pipe where the initial corrosion allowance was 3mm and the design life was 20 years. The standard deviation for the corrosion rate was taken from Appendix A. The inspected curve assumes a corrosion rate of 0.15 mm/year with the wall loss being confirmed at each inspection. The reduced RCA at each inspection means that the PoF increases quicker and exceeds the previous maximum after each inspection period.

3.4. Summary

By defining boundary conditions of a CML, a continuous PoF(t) curve can be drawn. This curve can be updated as inspection data gets added. Immediately after inspection, the PoF(t) is reduced to 0 and as equipment approaches its minimum required thickness, the PoF(t) reaches its threshold in a shorter timeframe. Equation 3-7 can be used with any representative dataset, however, as it includes a term for standard deviation, excessive scatter or corrosion rates within the dataset may prematurely increase the risk.

4. Machine Learning Approach to Corrosion Rates

4.1. Introduction

This chapter outlines a novel approach to corrosion rate calculations by the application of machine learning. To validate the results against the current industry approach, the corrosion rate for the dataset was first calculated using the current approach. It was then attempted to calculate corrosion rate through classification and finally, corrosion rate was calculated a neural network.

4.2. Current Approach

Currently, where the thickness history of a CML is available, the future corrosion rate is classified visually based on the perceived trend

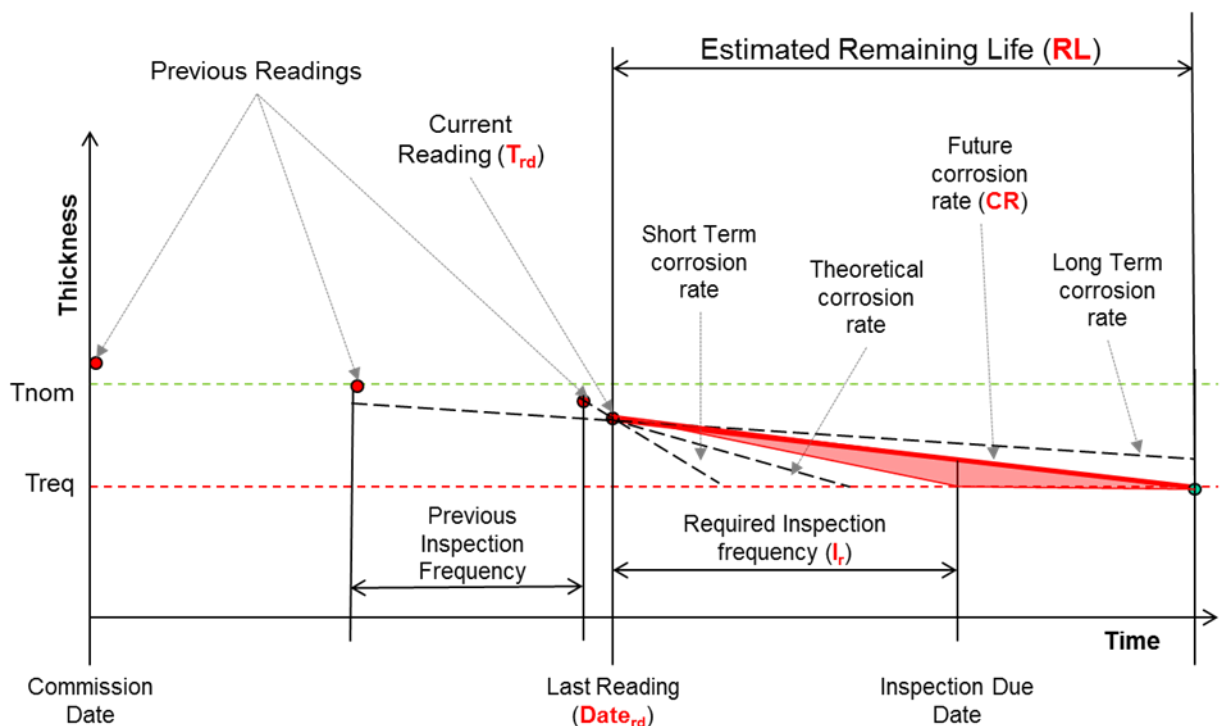


Figure 4-1 Visual Assessment of a Single CML

Each CML is reviewed in isolation, where the data does not conform to a straight line e.g. in Figure 4-1, the reading which is least likely to help predict an accurate future corrosion rate is removed from consideration. There are several reasons why a reading may be removed from consideration.

- 1) Replaced Section: When a section has been replaced with the exact same material, the previous reading may produce a trend that is useful for forecasting, however including the new reading would distort this calculation. This is information that may not fit the standard inspection reporting framework. A review of the piping systems repair history would have to be known in this case.

- 2) Short term corrosion: Short term corrosion is the linear corrosion rate taken between the last two readings. Where the short-term corrosion is higher than expected, the assessor will look to assess the credibility of this rate.

There are many reasons that corrosion rate would have increased since the previous inspection. A commonly documented case is increasing water cut[32] in offshore installations. At the time of assessment, if there have already been replacements made on the system, and a failure investigation has indicated that the corrosion was due to the current operating conditions then the assessor may judge the calculated short-term corrosion to be credible.

A 12.7 mm pipe may have an installation thickness anywhere between 12 and 14 mm. An inspection taken one year apart could easily show a difference of 1mm giving a short-term corrosion of 1mm/year, which likely exceeds the design corrosion rate. If the latest reading is greater than 12mm the assessor may conclude that there is no active corrosion at that location.

- 3) Readings taken on fittings: Fittings such as elbows formed by bending may have thickness deviations of up to 20%[33]. The repeatability of thickness measurements on fittings such as Tees, Elbows and reducers is a well-documented issue.[34]

All these issues are considered qualitatively in the corrosion rate assessment of each CML. Due to the qualitative nature of the assessment, the disparate sources of data and the sheer volume of CMLs to be assessed, it is too complex to consistently apply all the pertinent data at hand for each assessment.

4.2.1. Corrosion rate characteristics

By classifying the corrosion type using the available criteria for the first 10% of points in a system, the remaining CMLs can be compared to see if they conform to the same rules. 10% was selected as a practical amount to cover all likely scenarios of corrosion.

While design and operation data related to the piping systems can be queried quite easily, there is no method of searching a database for the likeness of the shape of the corrosion. A neural network was set up to rank all other points in the system in terms of similar corrosion history.

If there is some commonality between the top search results and the CML being assessed, e.g. a common process condition, material, fitting type etc, further confidence is gained in the determined corrosion rate. This selection can then be grouped or further split out.

A single CML with a reading 0.2mm less than the previous inspection may have anywhere between 0 mm and 0.4 mm wall loss. However, if a group of points experience an average drop of 0.2mm, then each CML should have an increased risk profile between then and the next inspection.

4.2.2. Linear Corrosion

Where T_c from the latest inspection result appears to follow on from the previous inspections, then the corrosion is classified as linear. The corrosion rate is then determined using a least squares approach on all data points.

4.2.3. Error in measurement

The paper Development of Wall-Thinning Evaluation Procedure for Nuclear Power Plant Piping[35] discusses the error in measurements taken by UT. Data collected by 12 qualified inspectors on five pipes and 14 fittings under laboratory conditions showed that a standard deviation of up to 0.25mm on fittings and 0.05mm on straight pipe. In this study, a factor of 2 was applied to these deviations to allow for the fact that the quality of the data was unknown, and that the resolution of the data was 0.1mm.

Given that inherent error in measurement is equal in magnitude to the expected change in wall thickness between inspections, the corrosion rate may appear to vary considerably during the early inspections. This is further compounded by the 12.5% mill tolerance allowed on pipes in most construction codes.

To test how the PoF(t) handles errors in measurement, a Monte Carlo simulation was performed on the data set. It was assumed that the errors in measurement followed normal distribution.

Each thickness reading was updated using Equation 4-1

$$t_{new} = t_c + (-1)^n \times \sigma \sqrt{\ln\left(\frac{1}{2\pi p^2 \sigma^2}\right)}$$

Equation 4-1

Where

n is a random integer between 1 and 2

p is randomly distributed between 0 and 1

σ is the standard deviation in error measurement, a conservative value of 0.1mm was used for straight pipe and 0.5mm was used in this Monte Carlo simulation.

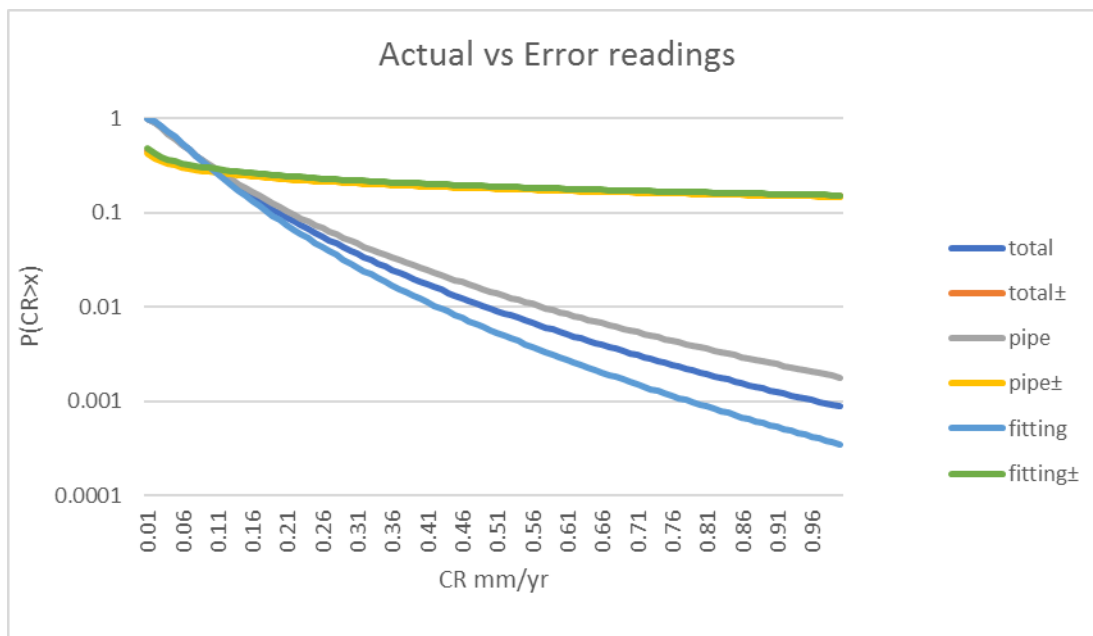


Figure 4-2 CR Distribution of Monte Carlo Simulation

The results of the simulation given in Figure 4-2 show that error in measurement causes greater likelihood that a given corrosion rate is exceeded. This is due to the loss in useful data. Almost 50% of the points had to be excluded as the CR < 0. This was coupled with the CRs on the high-end of the scale doubling in value.

4.2.4. Short Term Corrosion

Where T_c from the latest inspection appears lower than predicted from previous inspections shown in Figure 4-3, it is safer to forecast the corrosion rate from the latest two inspection points.

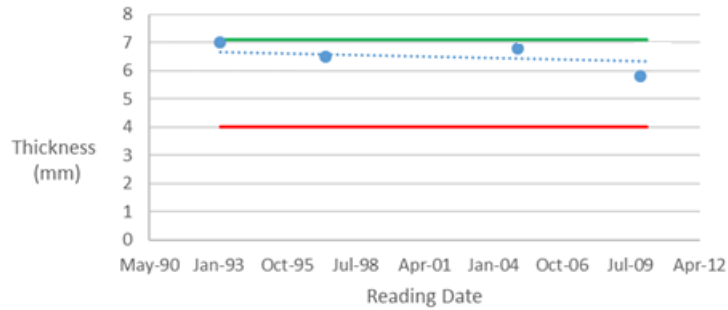


Figure 4-3 T_c is lower than expected based on all previous readings

4.2.5. Long Term Corrosion

Where T_c from the latest inspection appears higher than predicted from previous inspections shown in Figure 4-4, it is safer to forecast the corrosion rate from the overall time.

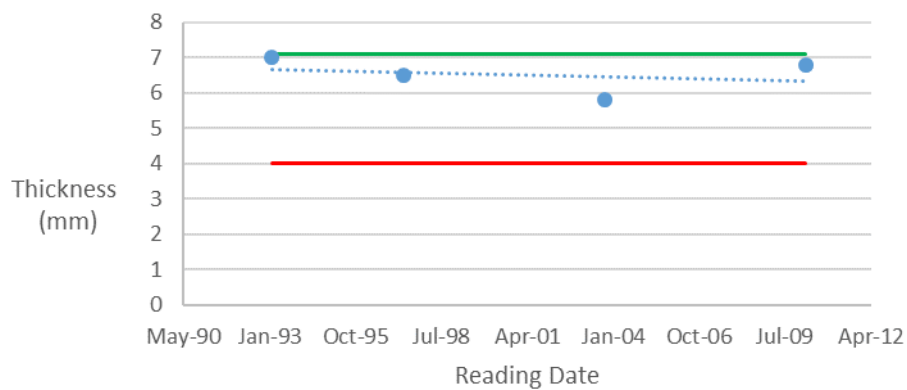


Figure 4-4 Thickness Readings Over Time

For each individual CML, the Min Reading and Date can be plotted. Corrosion rate is determined by taking a linear trend of the previous inspection readings.

4.3. Data Normalisation

In order to apply machine learning to the data set, the data first needed to be normalised. The results of machine learning can be skewed where one parameter is significantly larger than the other. In this case, thicknesses in mm range from 0 to 50 whereas the dates in years ranged from 1970 to 2017. The data was normalised at a CML level by subtracting the reading at the first year from all subsequent readings. The script to produce this data can be found in APPENDIX D.

The raw data used in this study was formatted Point, Date, T_c in mm.

| Date | T_c | ΔY | ΔT_c |
|--------|-------|------------|--------------|
| May-14 | 7.3 | 0.00 | 0 |
| Oct-12 | 7.6 | 1.58 | 0.3 |
| Sep-11 | 8 | 2.67 | 0.7 |
| Feb-03 | 8.5 | 11.25 | 1.2 |
| Jan-00 | 9 | 14.33 | 1.7 |

Table 4-1 Raw values get normalised for machine learning

4.4. Corrosion classification

As the number of readings per cml was variable, a neural network was constructed for each number of readings from three up to 10. Where there were more than 10 readings at a CML, the latest 10 readings were taken.

This diagram is represented mathematically in Equation 4-2 by multiplying the inputs by at least one weight matrix, converting it through a sigmoid function to generate outputs. The sigmoid function is used to train the matrices as it is monotonic and differentiable for all values.

$$\{I\} x \sigma[W1] x \sigma[W2] \dots \sigma [Wn] = \{O\}$$

Equation 4-2

Where the sigmoid function $\sigma(x)$ is given by Equation 4-3 as:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Equation 4-3

The values within the weight matrices are trained by a process known as backpropagation, a selection of data with known outputs called a train set is run through the model. Where the output is used for classification, in this instance, corrosion type, the values in the output vector will equal 0 or 1.

The specific type of backpropagation used to train this data was stochastic gradient descent which recursively apportions the error back through the network in proportion to the rate of change or gradient of the neuron at a given value. When the sigmoid function is used as an activation function, its differential is given by Equation 4-4

$$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$

Equation 4-4

Initially, the weight matrices e.g. in Equation 4-5 are populated with random variables. The results are then compared to the expected results. The error is then run backwards through the weight matrices, apportioning the largest change to the weight where the rate of change of error is the largest.

$$\sigma \left(\begin{matrix} Y_1 \\ Tc_1 \\ Y_2 \\ Tc_2 \end{matrix} \right) \times \begin{bmatrix} w_{11} w_{21} w_{31} w_{41} \\ w_{12} w_{22} w_{32} w_{42} \\ w_{13} w_{23} w_{33} w_{43} \\ w_{14} w_{24} w_{34} w_{44} \end{bmatrix} = \begin{matrix} O_1 \\ O_2 \\ O_3 \\ O_4 \end{matrix}$$

Equation 4-5

This can be represented graphically as shown in Figure 4-5:

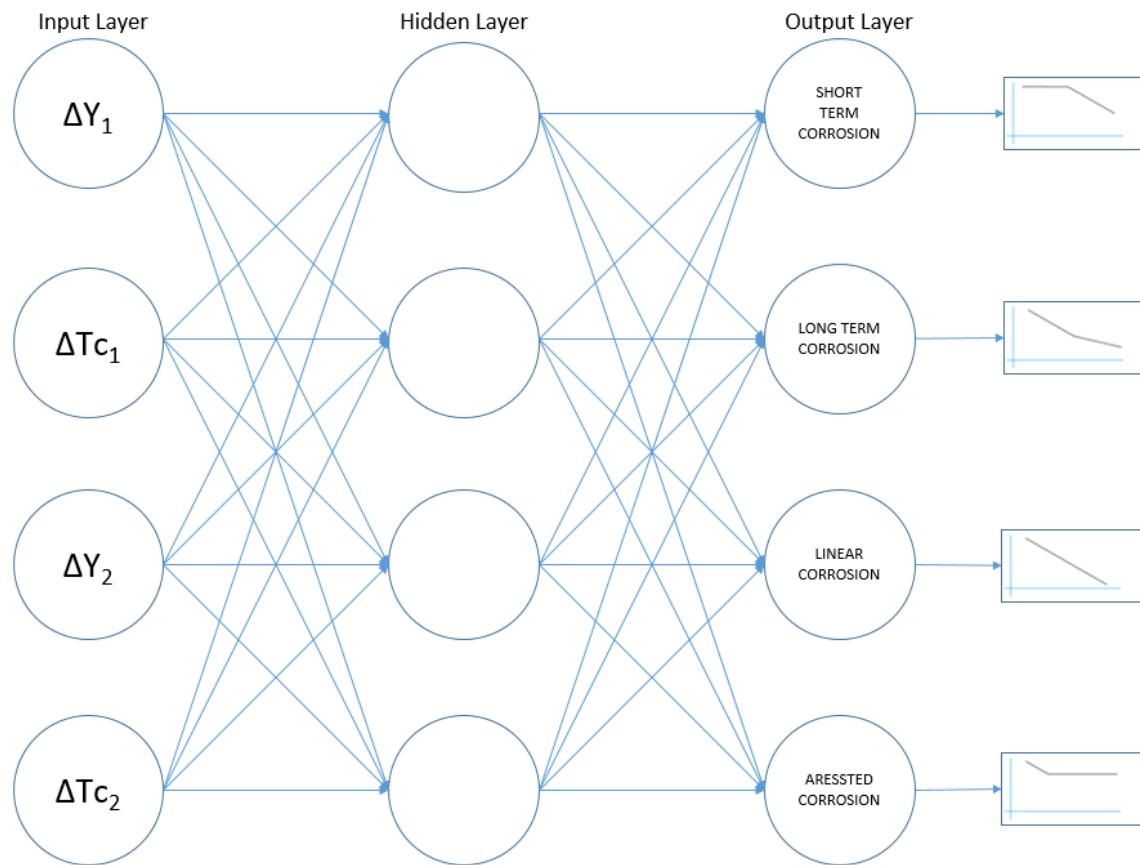


Figure 4-5 Neural Network to determine corrosion mode

During inspection, a reading may be taken at a location that shows significant corrosion, over and above what was expected. In this instance, an operator will want inspection carried out on CMLs that have a similar susceptibility.

This susceptibility was determined by searching for similar location where there may historically have been a comparable set of readings.

In practice, where a single CML shows a deviation from the expected thickness within 0.5 mm it is often accepted as being an anomaly, however, if there were many points that showed a similar deviation, further consideration could be given to these anomalous readings. Currently, there are no reliable methods of classifying or grouping points in this manner except for the recall of the assessor. This is a flawed method as there can be multiple assessors working on a single piping system.

The cost function varies by application. A standard cost function applies equal cost to false negative predictions and false positive predictions, however, in a corrosion classification algorithm where the majority of cases should return false, the cost of a false negative i.e., a result that appears negative when it should not, should be higher than a false positive i.e., a result that appears positive when it

should not. In this instance, a false positive will require unnecessary inspection whereas a false negative may lead to the loss of integrity of a system.

| | | |
|----------------|---------------------|--------------------|
| | No Predicted Damage | Predicted Damage |
| No Real Damage | No Action | Unnecessary Repair |
| Real Damage | System Failure | Necessary Repair |

The cost of scheduled repair on critical piping systems can be orders of magnitude lower than the losses in production alone. There are also environmental, safety and legislative issues associated with uncontrolled releases.

Training:

When behaviour had been observed on the training data, in this instance short term corrosion, it was assigned an output of 1 to the short-term corrosion output node.

The data was collated in excel and SQL before being analysed using the package neuralnet in RStudio.

Train set:

In machine learning, data is split into training and validation sets. In the data set, where readings were outside the expected value, they were supplemented by an inspector's comment. These were then split into 70% training and 30% validation for each known type of corrosion.

This code is available in Appendix D.

4.5. Corrosion Rate Calculation

All the CMLs had a corrosion rate calculated by manual assessment of the thickness history. Each point had its entire history assessed based on the assumption that corrosion rate was linear. Any readings that could be considered outliers by overlaying supplemental data were excluded from assessment.

The pattern of exclusion was repeatable in many cases, e.g. a line that had been replaced may have a thickness history that reduced consistently and was then restored to its nominal thickness. In such cases, the restored thickness could be ignored for the purposes of corrosion rate calculation.

Given that all the CMLs had this manual assessment performed, they had a validated corrosion rate. By assuming that the same thickness reading produced the same corrosion rate, a neural network was implemented to fit these inputs of thickness and time against a defined output of corrosion rate.

The total number of readings available for each point varied from two to nine. The total number of readings was skewed towards the lower amount. 90% of the CMLs had less than 6 readings. An individual network was set up for each amount of thickness readings in order to prevent the results of CMLs with lower reading amount skewing the CMLs with a higher number of inspections.

The results of these neural networks are available in APPENDIX B. in Figure B-0-1 and subsequent figures, the Neural network calculated results (y-axis) were plotted against the manually calculated results (x-axis).

The network appeared to predict corrosion rates with a higher degree of accuracy where the readings were lower, and overall, the resulting predictions produced lognormal distribution curves that could be used for meaningful predictions.

4.6. Summary

The methodology used in this assessment is suitable for analysing large datasets to give general PoF predictions however, the classification model did not identify any useful patterns to give more refined predictions.

5. Discussion of results

5.1. Results Overview

The current neural network does not classify corrosion by type but is able to retrieve points from the global data set in order of decreasing similarity of the CMLs corrosion history. This data allows us to identify CMLs with similar corrosion history, thus providing a useful tool for manual classification of corrosion type.

An attempt was made to get the neural network to automatically classify the different corrosion types by random forest, however the network could not be trained sufficiently to provide consistent output. This is due to the number of potential classifications, the variety of data points within each classification type and the varying number of data points within each CML which created too many variabilities for the tool to learn and process correctly. Where the number of potential corrosion types is reduced and the data points within each CML is consistent it will be possible to train the network to identify the corrosion types using this methodology.

One of the most significant issues with this methodology is teaching the neural network to ignore scatter within the data sets. A significant potential application of this tool is for it to identify where a process change has occurred, and the rate of corrosion increases. This is difficult to code within the tool given the dual needs of overlooking scatter but also identifying significant and dramatic shifts within the pattern of readings where they are significant.

A significant potential application for this tool is re-classifying data points that will have previously been manually classified as scatter, now as leading indicators of increased corrosion. When readings are taken and a more significant decline than expected is recorded, the tool will be able to identify previous readings that may have been leading indicators of the increased corrosion and re-classify them across CMLs.

5.2. Accuracy

The corrosion rates manually calculated from the least squares approach excluding outlying data was compared to the calculated corrosion rates from the neural network.

5.2.1. Overall Accuracy

Comparing the real and calculated CRs of CMLs where not all the thicknesses were used in the assessment is less reliable. Figure 5-7 shows that some predicted rates differed significantly from the manually calculated value however, as these errors balance out over a large dataset, the resulting lognormal distribution curve in Figure 5-1 appears to give a close enough approximation to provide a meaningful corrosion rate distribution curve to use in a PoF calculation.

It should be noted that directionally, most predictions made by the neural network are better than or equal to calculating CR using all the available points.

The data appeared to be scattered or distributed on both sides, i.e. the Cr on some CMLs were calculated to be above what they were and some below.

Comparing the real (x-axis) and calculated (y-axis) CRs of CMLs where all the thicknesses were used in the assessment Figure 5-2 show the Neural Network calculations to be reliably within 0.1 mm/year. The results did not show a particular bias, however, Figure B-0-5 shows that the log normal distribution curve generated from the predicted results is closely aligned with the calculated results.

The large spikes shown in Figure 5-7 are the absolute error between the corrosion rate as calculated by the neural network and the corrosion rate as calculated from all available thickness readings.

This disconnect would be a useful methodology for screening and highlighting CMLs that would require manual interpretation. Figure 5-2 shows the accuracy of the entire dataset, Neural Network Results shows the dataset broken down by type and number of readings available. Figure 5-6 shows that the accuracy is reduced when compared to the entire data set but the majority of CMLs still fall within ± 0.1 mm/year.

5.2.2. Accuracy of Neural Network Entire Dataset

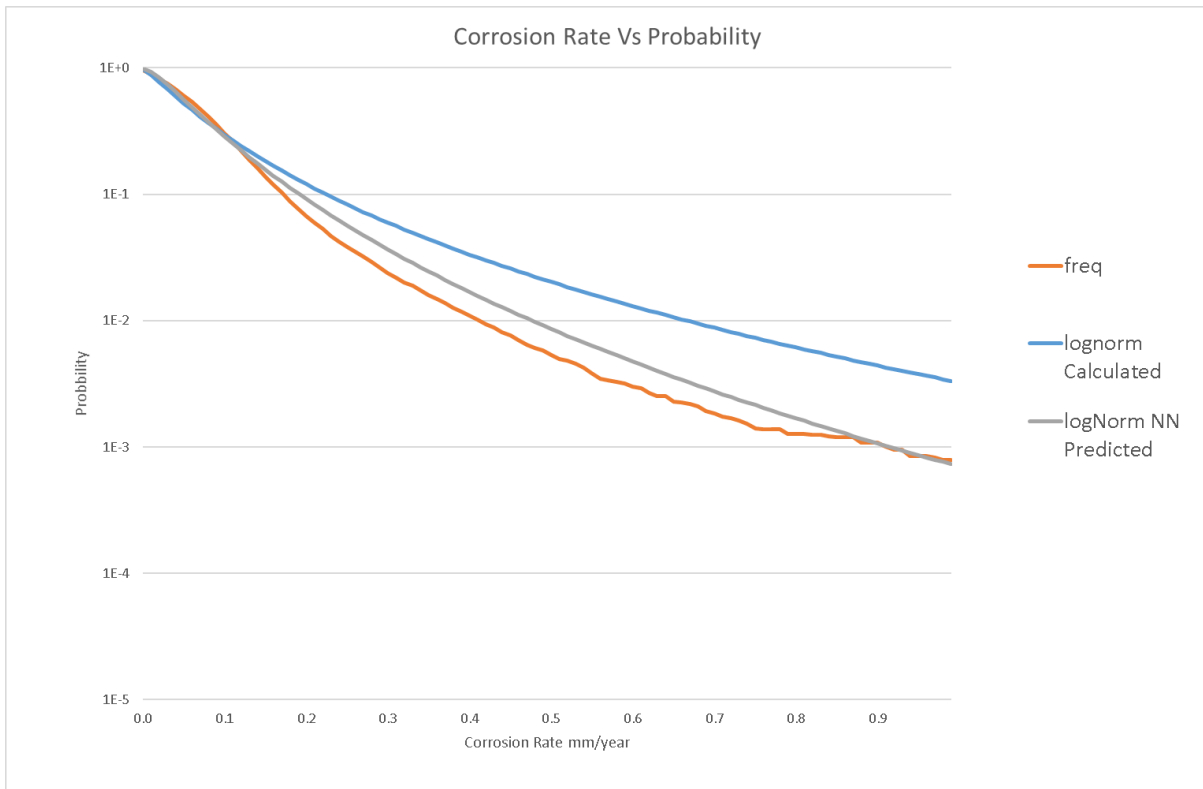


Figure 5-1 The lognormal distribution curves for the entire dataset from the calculated and predicted corrosion rates are relatively close

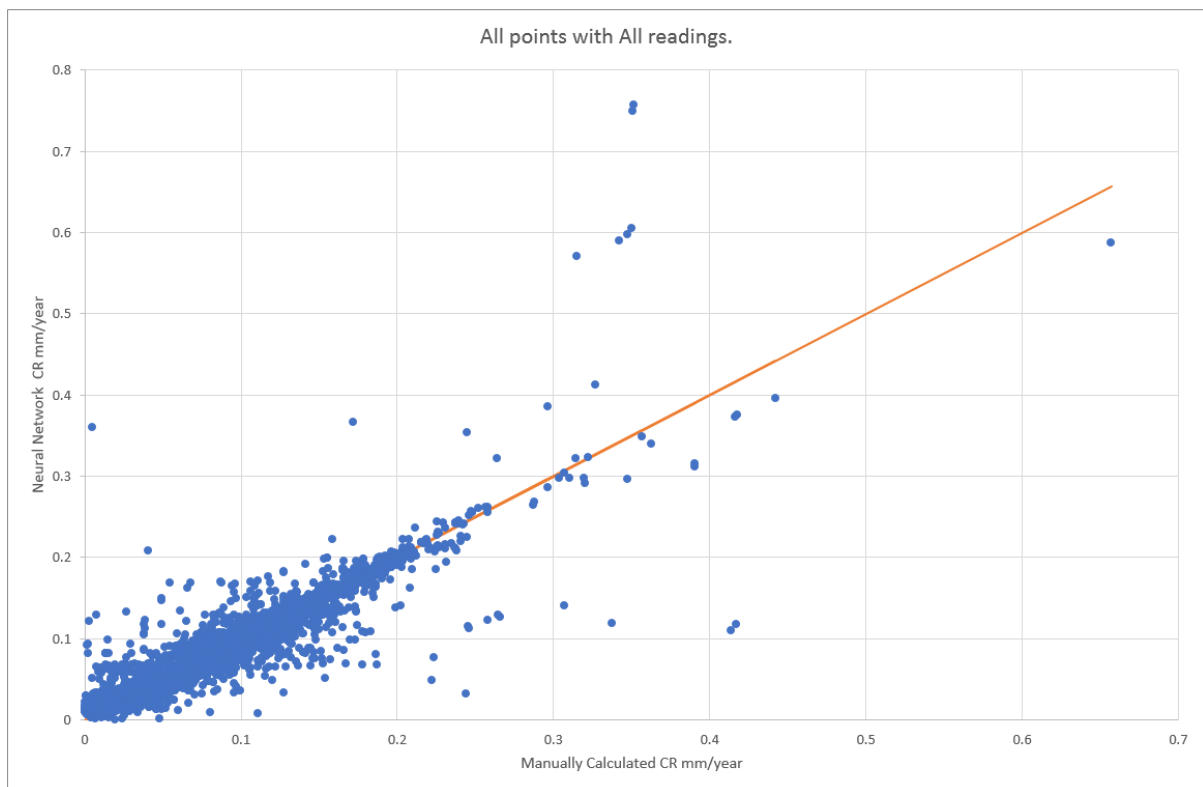


Figure 5-2 All Corrosion rates vs Predicted Corrosion Rates

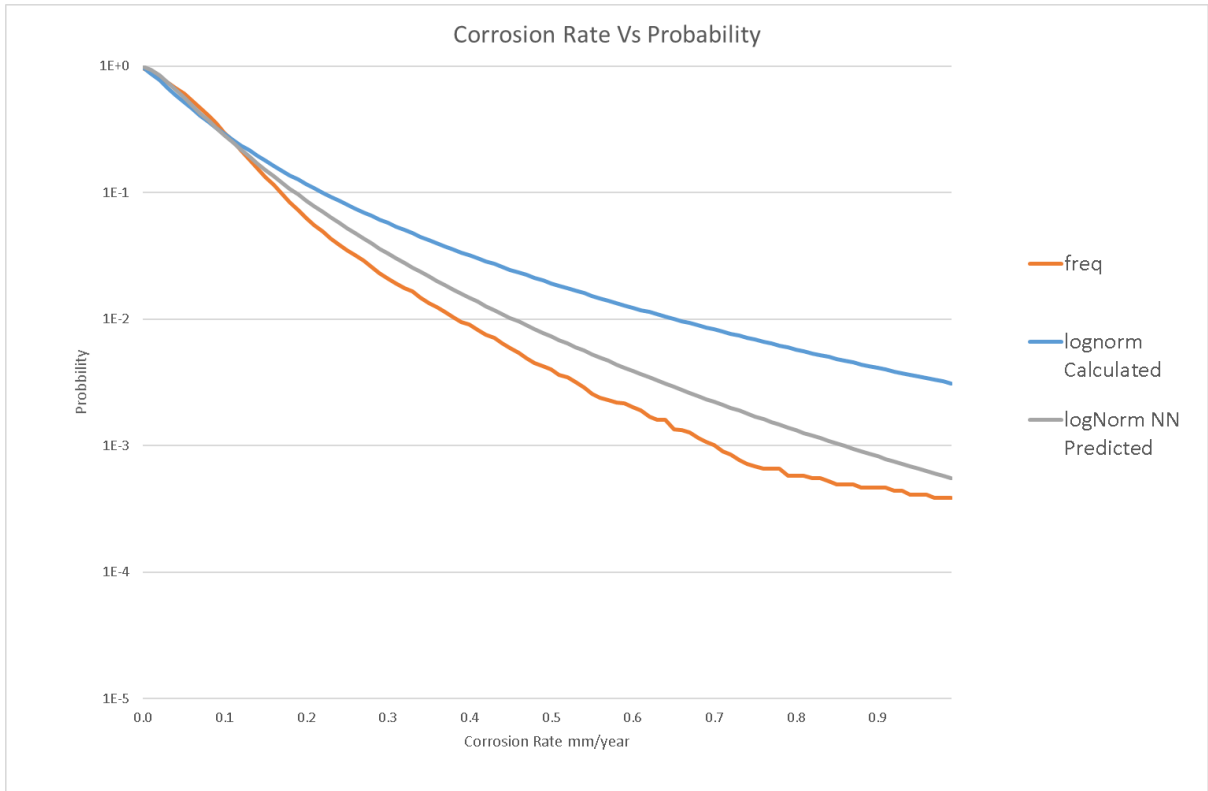


Figure 5-3 The lognormal distribution curves for the CMLs with no invalid readings

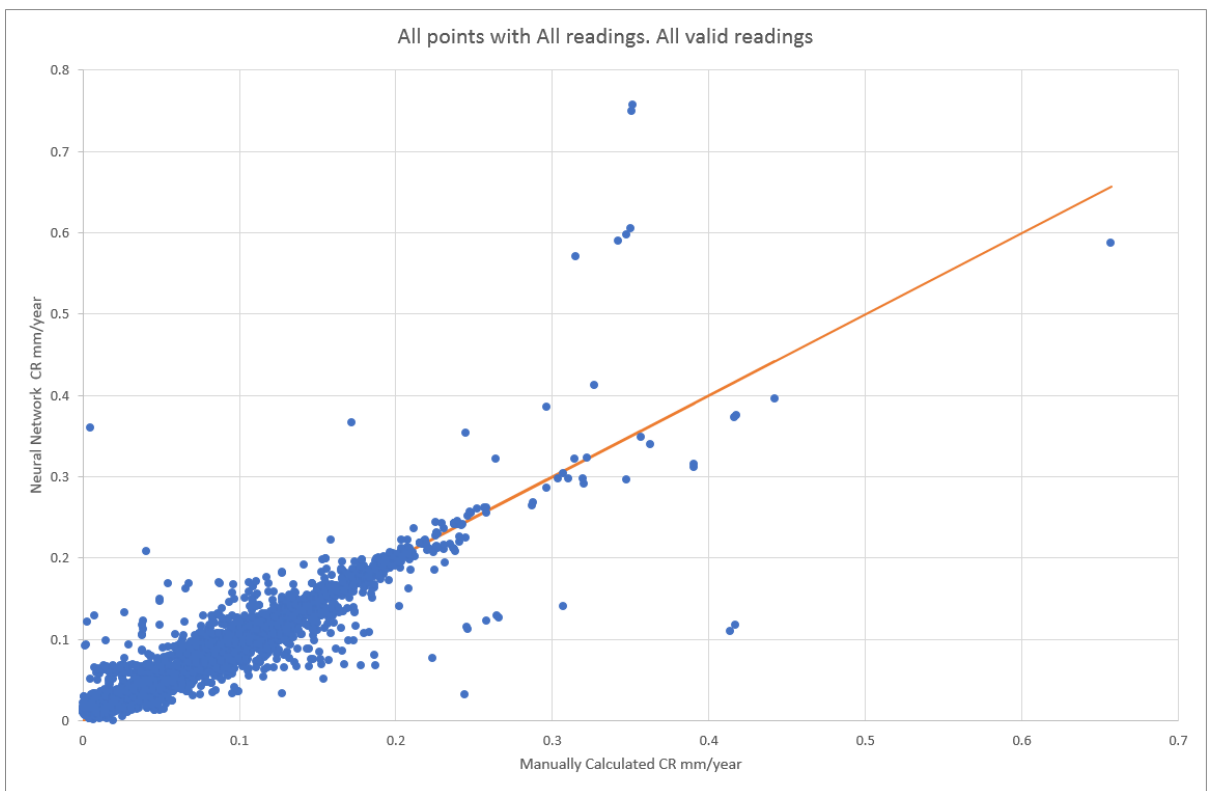


Figure 5-4 All Corrosion rates vs Predicted Corrosion Rates all valid readings

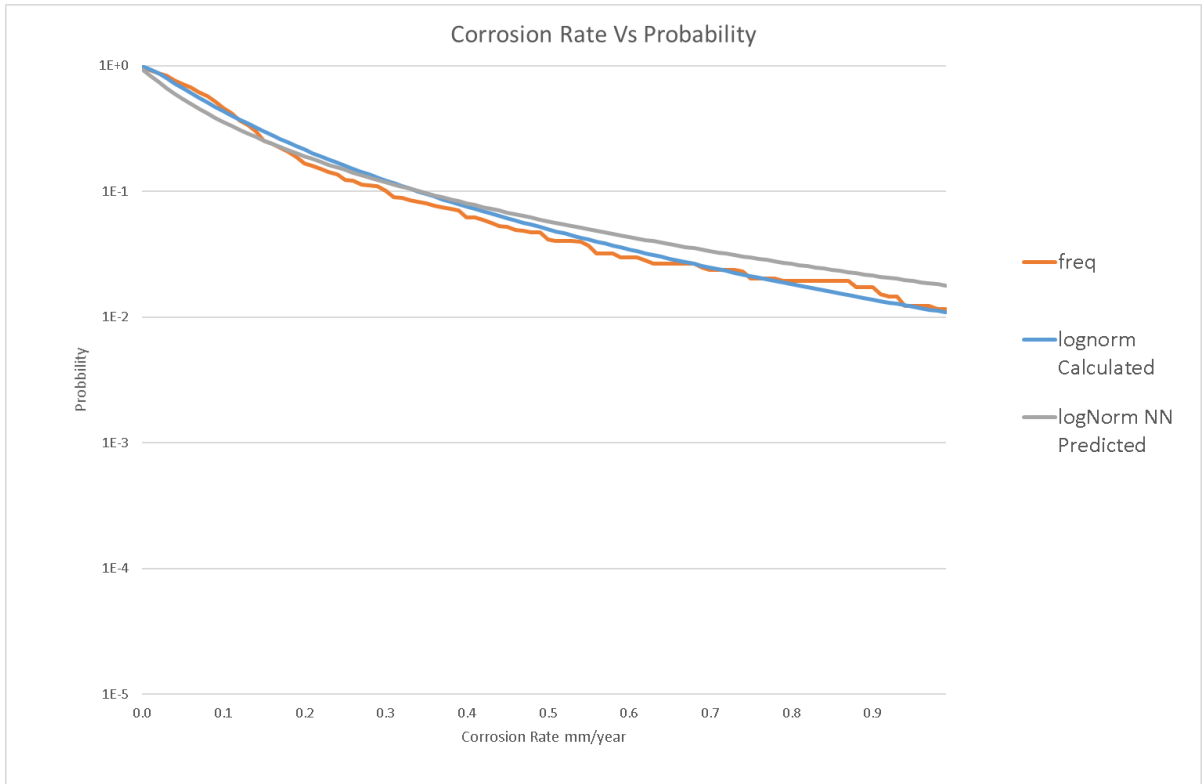


Figure 5-5 The lognormal distribution curves for the CMLs with no invalid readings

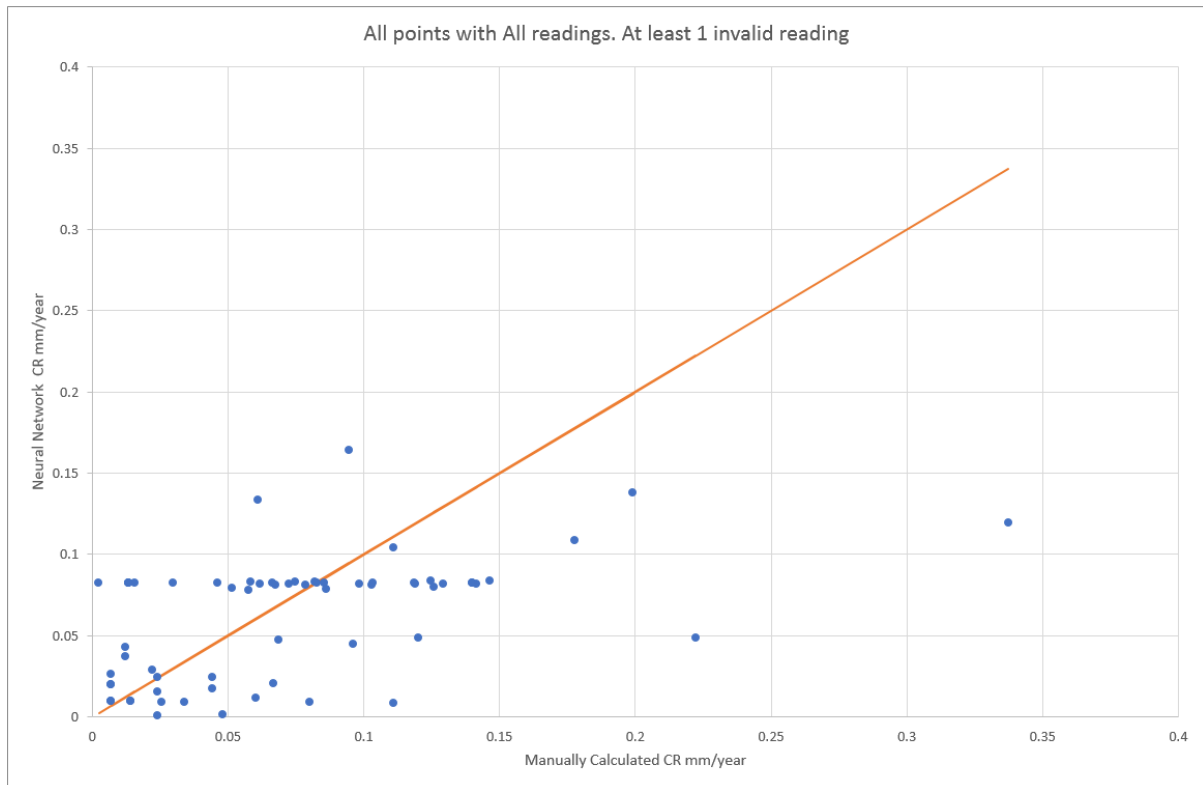


Figure 5-6 All Corrosion rates vs Predicted Corrosion Rates excluding readings



Figure 5-7 Predicted CR vs CR Including All Readings

6. Conclusions and Future Work

6.1. Conclusions

The conclusions that can be drawn from this study are:

6.1.1. Current Approach

The current approach to determining the PoF(t) of a piping system has been difficult to implement in practice due to a lack of available data. The calculation is not transparent enough to determine the effect of supplemental data on the future forecast.

The previous versions of API 581 were also highly subjective and did not provide a mechanism of applying local corrosion data to a piece of equipment.

Since the 2008 version of API 581, the 2016 version has standardised the PoF and CoF to form straight line “ISO Risk” lines. Given that the CoF has remained constant, this requires that the Internal Thinning module be more time sensitive, however, the actual result is driven by an understanding of historical damage that has not yet been fixed.

6.1.2. Feedback

Feedback was shown to be updated by adding a simulated inspection.

Log normal distribution was shown to overestimate the corrosion rate slightly but more accurately predicted the extreme values; therefore, it is more suitable for risk assessment.

Weibull distribution was a better fit, but it tended to underestimate the higher end corrosion rates. Given that the distribution follows this pattern it is more useful for forecasting the remaining life of an aging asset.

The feedback loop is possible to implement across multiple piping systems and on multiple classifications such as diameter, fitting type service, material that are available through the integrity management system. The following scripts have been included in Appendix D.

Script 1: Sql Code forData normalisation

Script 2 calculates the corrosion rate based on the trained network.

Script 3 trains the neural network based on a database of readings Based on the CML piping integrity model.

6.1.3. CR Calculation

The CR could be calculated effectively and accurately by a neural network. On review, it did show a higher degree of effectiveness on CMLs that had their entire thickness history included in the corrosion rate calculations.

It was less effective on CMLs with excluded readings however it was generally somewhere between the all-inclusive corrosion rate and the manually assessed corrosion rate. This was inclusive of corrosion rates that tended negatively, i.e. when a line was replaced the network was still capable of assigning the previously determined corrosion rate to that CML.

The purpose of any of these studies is to recommend the next appropriate inspection date. The sensitivity of the corrosion rate as applied in an API 581 study or otherwise generally does not vary if the corrosion rate is ± 0.05 mm/year.

The neural network was able to calculate over 99% of CMLs to within that tolerance. The CMLs outside of this threshold were generally unusual cases with similar representative data not applied to the same network.

6.1.4. Classification

The accuracy of the neural network was not high enough to use as a classification, i.e. it would not be enough to determine if a system is seeing elevated corrosion. There were several factors in this study that likely reduced the effectiveness of the classification:

- 1) There was not enough known elevated corrosion data present. If the data used was on a live system with elevated corrosion experienced at known locations, then the exact pattern of elevated corrosion could be incorporated into the neural network.
- 2) The majority of CMLs had a low number of readings. There are several scenarios that may start off with the initial 3-4 readings being identical. Because there was no clear way to delineate these CMLs in the study they were all considered to be the same.
- 3) The exact operating conditions were unknown.

In the original data set, comments were left by both the inspector and the analyst. The inspectors' comments often provided context to the analyst and future analysts could make use of both the inspectors and the previous analysts' comments.

The inspectors' comments often contained easily searchable terms such as pitting, scab corrosion, internal corrosion, confirmed reading, arrested corrosion. These all provided valuable context to an assessor but were not included in the neural network training data.

A future assessment could gain accuracy if an algorithm was able to recognise these data points.

6.1.5. Validation

This model was not validated against the failure data as this was not available.

6.2. Comparison of PoF vs API 581

The PoF as determined by API 581 is influenced strongly by the inspection effectiveness. The guidance in API 581 for assigning this effectiveness is subjective. A comparison of PoF was performed against the feedback method. There was no clear correlation between the 2 methods

6.2.1. Effectiveness A

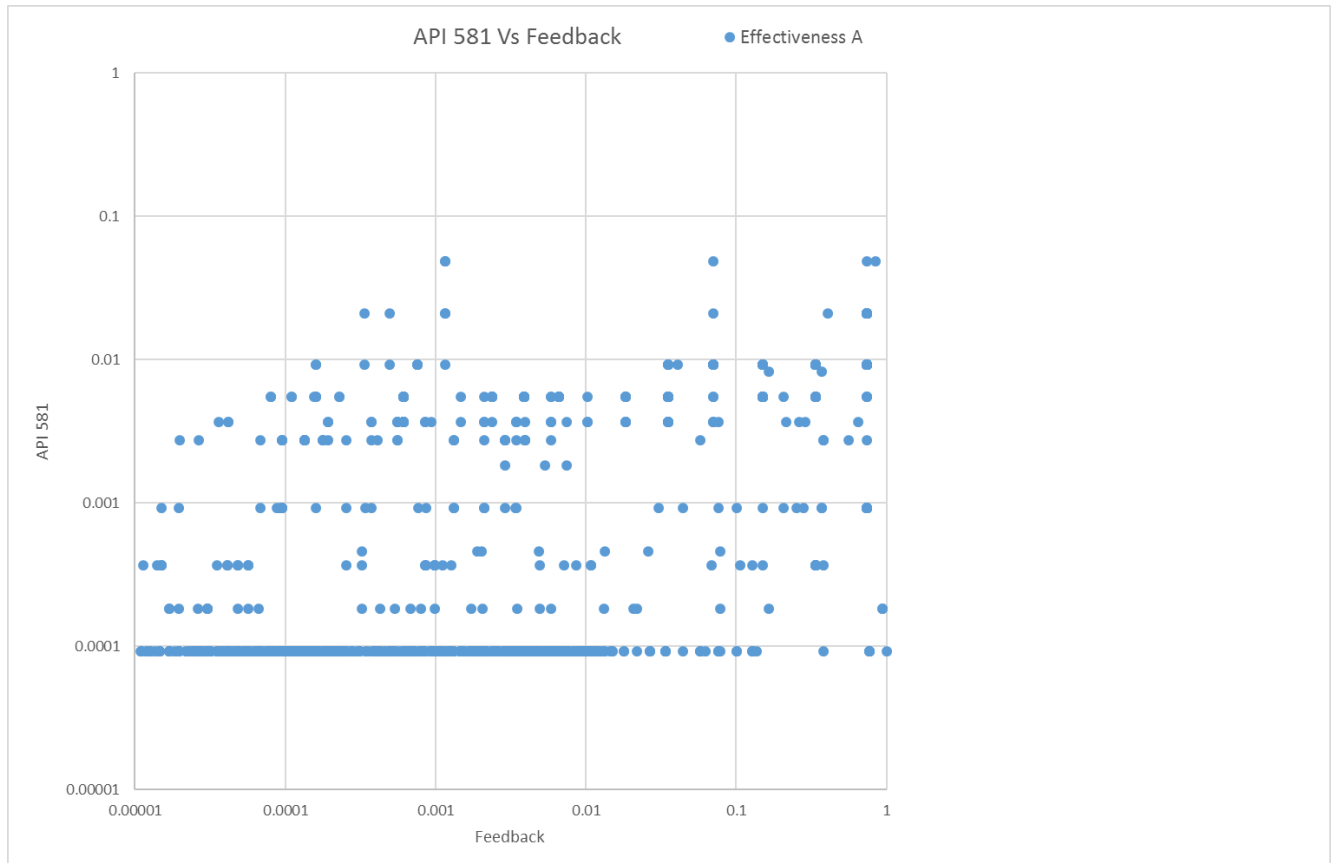


Figure 6-1 Forecasted Risk with inspection effectiveness A

6.2.2. Effectiveness B

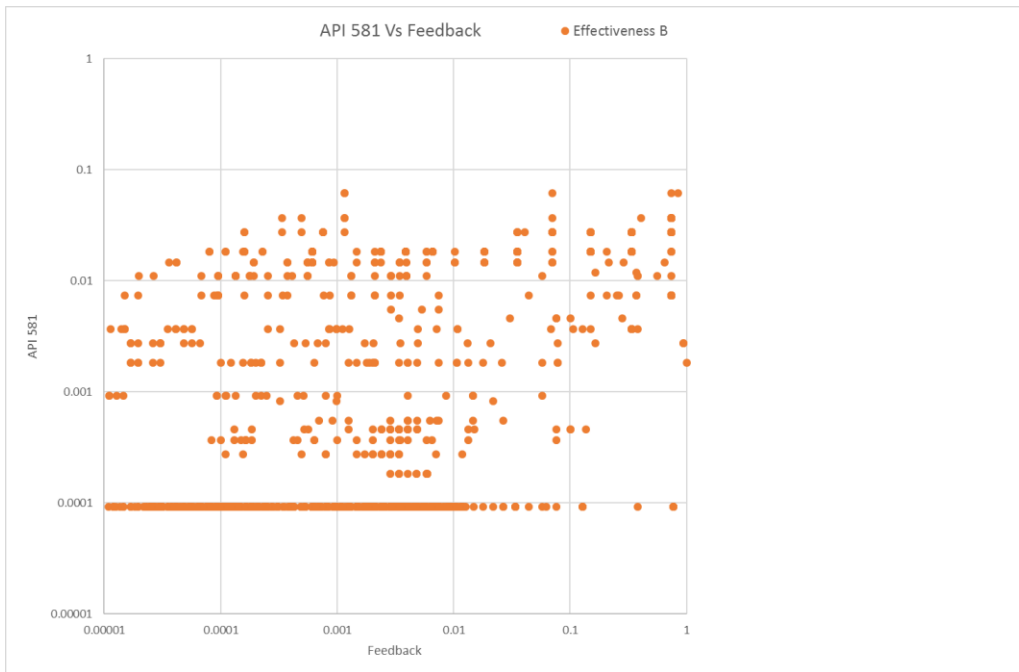


Figure 6-2 Forecasted Risk with inspection effectiveness B

6.2.3. Effectiveness C

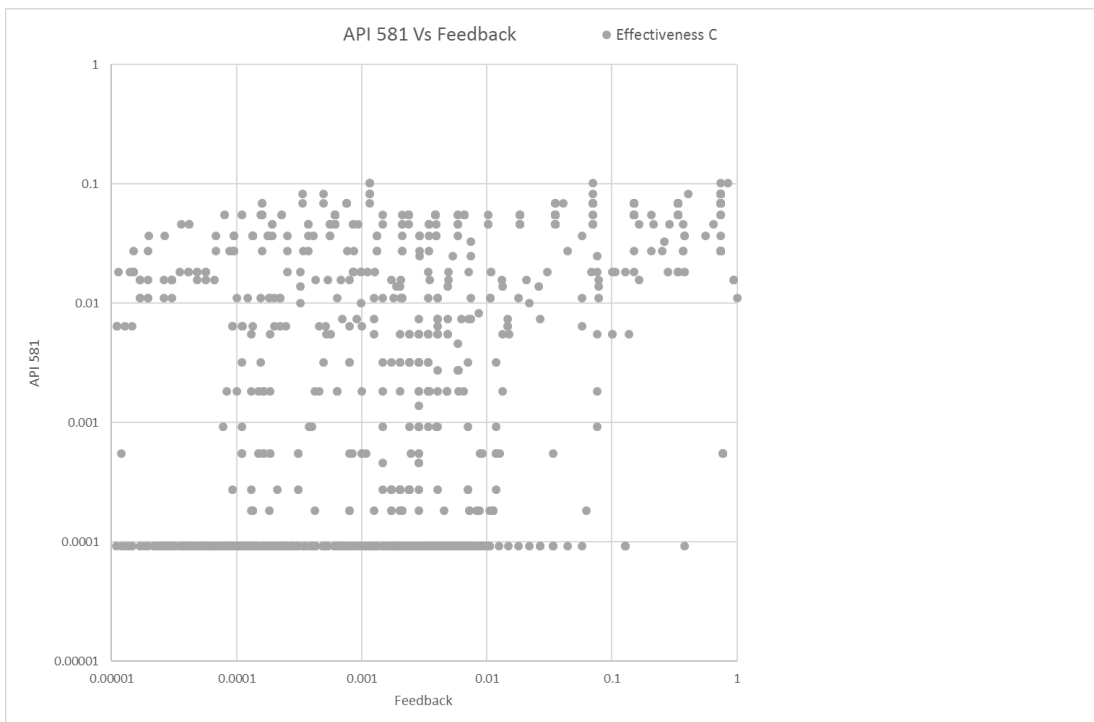


Figure 6-3 Forecasted Risk with inspection effectiveness C

6.2.4. Effectiveness D

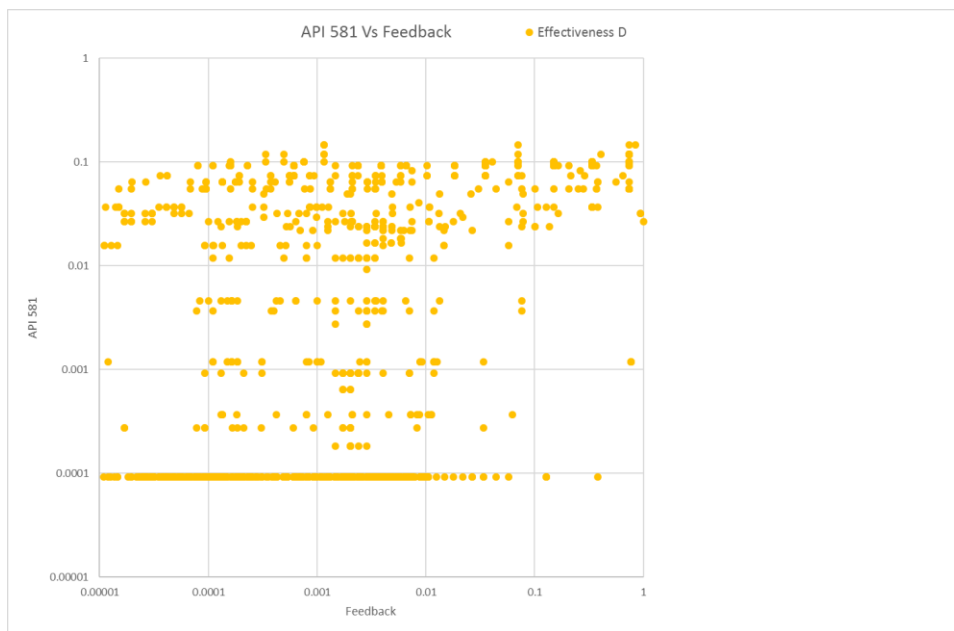


Figure 6-4 Forecasted Risk with inspection effectiveness D

6.3. Future Work

Future development and refinements of these methods could potentially be pursued in the following areas.

While it did show a correlation, the PoF tended to be understated by API 581, particularly when there were repeated inspections at the same location confirming the corrosion rate.

For the neural network to be implemented in practice, there would likely be resistance from the asset owners as it was equally likely to underestimate the corrosion rates. To address this issue, the cost function could be modified. The package used in this study did not have readily available way in which to modify the cost function. The package use mean squared error (MSE) which does not differentiate between overestimation and underestimation. Other cost options which could be trialled on the data set include mean error (ME) and mean absolute error (MAE).

Asymmetric cost functions[36] can be developed to produce values that err on the side of conservatism. A simple asymmetric cost back propagates multiplied by a variable which depends on whether the cost is positive or negative.

More bespoke non-linear asymmetric cost functions could be developed for this purpose that restrict outlying i.e. negative results from being calculated.

Random forest classification could be gradually implemented over time as inspection anomalies could be parameterised. This could then be used to identify the main drivers of risk within the system.

6.3.1. Supplementary operation and construction data

It is common that a process will have the same design pressure and temperature throughout, however, the temperature and pressure vary continuously throughout the system. Furthermore, the orientation and flowrate at a CML level may vary significantly. A study of corrosion in dead legs[37] shows the significance of orientation on CMLs at localised areas of reduced flow.

None of the CMLs in the available dataset had the year of construction considered. NACE MR0175[38] is a standard developed for metallic materials exposed to H₂S in the oil and gas industry. Although the primary requirements laid out in this standard were to prevent stress corrosion cracking, they have proven to be beneficial to reducing corrosion of pressurised equipment in sour service. Some of the piping in this study may not have been built to NACE requirements and its inclusion may have introduced scatter.

6.3.2. Combining supplementary non-UT inspection results

API 581 determines risk based on the confidence of inspection results. In this study, the manually determined corrosion rates were calculated by assigning TRUE or FALSE to each reading. The resultant corrosion rate was then used in training the neural network to produce the same corrosion rate with the same inputs. This could lead to a source of error if the sampling methods were different. It is possible that the first UT reading taken at a cml was overstated as the location may just be scanned for presence of corrosion. Another location may have a similar reading determined by radiography. These would both produce the same output from the network, whereas a manual review may exclude the initial UT reading.

Adding a third input neuron per inspection which denotes confidence in the reading would produce more reliable results.

6.3.3. Classification Potential

The neural network does not identify which points are outliers, i.e., given a thickness history at a given CML, a CR is determined. It may be possible by analysing the data, on a case by case basis, to determine which reading was excluded. This is a path that could potentially lead to a successful classification network.

Reliable output from classification could not be developed from the thickness readings and dates alone. In the available dataset, the majority of CMLs had 5 readings or less. This didn't appear to be enough to get the models to converge. Future attempts at classification could make better use of the supplemental data available at a CML level such as process conditions, fitting type, material etc.

6.3.4. Corrosion Classification Potential

A corollary of the classification would be identifying the actual corrosion type e.g. CO₂, H₂S etc.

In practice, an experienced corrosion engineer would be required to manually review the thickness history of CMLs experiencing a known type of corrosion in order to generate a training set of data for the system.

This would be beneficial when a specific corrosion type gains prevalence in a system after lying dormant since commissioning.

6.3.5. Validation of Results

For the feedback loop to be validated, the total number of failures each year would have to be compared to the expected number of failures. The probability of an expected number of failures (f) occurring within a system of CMLs(N) where each CML(i) has its own PoF is calculated with a modified binomial distribution formula[39]

$$P(f) = \frac{N!}{f!(N-f)!} \prod_{i=1}^N (PoF_i)^f (1 - PoF_i)^{N-f}$$

Equation 6-1

6.4. Recommendations

Feedback Loop

The feedback loop can be readily applied to life extension feasibility studies where the failure of a single location within the next assessment period is not as crucial as forecasting the likely remaining life of the total asset.

The original objective of the study was to calculate a localised PoF(t) for a CML which could in turn drive the PoF(t) of a different CML via a feedback loop. It's potential here is more limited because the corrosion type can't be classified. I.e. it is possible to have a result change the PoF(t) however it may not always necessarily reflect the correct changes.

To improve the accuracy of these calculations, failure and replacement history would need to be factored in to the PoF calculations as a means of normalising the results.

Neural Network

If the accuracy of the corrosion rate classification can be improved sufficiently, regression analysis could be performed across the data set which would be able to group the CMLs automatically.

This study took data from a broad range of processes and attempted to unify them where possible into a single data model in order to provide a distribution that would react as more data became available.

For this model to be applied in practice, a deeper knowledge of a single process should be sought. It is common in industry to have parallel streams or trains within a single process. Starting with a single corrosion loop, all available data should be collated for both trains, then the corrosion rates should be determined manually for 1 train. These results can then be fed into the neural network.

Once the network has been trained, it should then be applied to the inspection readings on the second train. A manual evaluation of the points in the second train will show the effectiveness of the network

References

- [1] Energy Information Administration, (last updated 14-April-2021), Use of oil - U.S. Energy Information Administration (EIA), Available at: <https://www.eia.gov/energyexplained/oil-and-petroleum-products/use-of-oil.php> (Access: 11-May-2021).
- [2] Popular Science, (last updated 01-March-2012), Deepwater Horizon Oil Did Enter the Food Chain, Available at: <https://www.popsci.com/science/article/2012-03/deepwater-horizon-oil-entered-ocean-food-chain-through-tiny-plankton/> (Access: 25-May-2020).
- [3] World Steel Prices, (last updated 25-May-2020), Free Data for Global Carbon and Stainless Steel Prices, Available at: <https://worldsteelprices.com/> (Access: 25-May-2020).
- [4] Al-Hashem, A.H. et al, The Effect of Water-Cut on the Corrosion Behaviour of L80 Carbon Steel Under Downhole Conditions, CORROSION 2013 (CD-ROM), Orlando, Florida, March 2013
- [5] Viking Engineering, (last updated 14-September-2018), Common Failures in Onshore Oil & Gas Wells, Available at: <https://www.vikingeng.net/news-resources/gat2004-gkp-2018-09> (Access: 25-May-2020).
- [6] Richardson, A., (last updated 01-March-2018), Enbridge's Mess, Minnesotans for Pipeline Clean-up, Available at: <https://healingmnstories.files.wordpress.com/2018/03/line-3-abandonment-report.pdf> (Access 25-May-2020).
- [7] Mathers, G., (last updated 11-December-2013), Ultrasonic Examination Part 1 - TWI, Available at: <https://www.twi-global.com/technical-knowledge/job-knowledge/ultrasonic-examination-part-1-127> (Access: 25-May-2020).
- [8] Lebowitz C.A., Ultrasonic Measurement of Pipe Thickness, Review of Progress in Quantitative Non-destructive Evaluation, 1993, doi.org/10.1007/978-1-4615-2848-7_255
- [9] Nasser, A.M.M., A Review on Oil and Gas Pipelines Corrosion Growth Rate Modelling Incorporating Artificial Intelligence Approach, Proceedings of the IOP Conference Series: Earth and Environmental Science, Vol. 476, pg no. 4, Paper ID no. 012024, June 2020
- [10] Inspectioneering, (last updated 25-May-2020), Overview of Asset Integrity Management (AIM), Available at: <https://inspectioneering.com/tag/asset+integrity+management> (Access: 25-May-2020).
- [11] Jungbauer, D., Acoustic Fatigue Involving Large Turbocompressors and Pressure Reduction Systems, Southwest Research Institute ,Vol. 27013, pg no. 111-222, 1998, doi.org/10.21423/
- [12] HSE Books (Health and Safety Executive) RR509 - Plant Ageing: Management of Equipment Containing Hazardous Fluids or Pressure, 2006
- [13] API Standards (American Petroleum Institute) API 510, Pressure Vessel Inspection Code: In-Service Inspection, Rating, Repair, and Alteration, 2014
- [14] API Standards (American Petroleum Institute) API RP 581, Risk-Based Inspection Technology, 2008
- [15] Kayley, L., Examining the Latest Changes to API RP 581 Risk-Based Inspection Methodology - Thinning and the Probability of Failure Calculations, Inspectioneering, Vol. 22, Issue 6, November/December 2016, Article 7
- [16] Noori, S. A., Case Study of the Use of API 581 on HK and HP Material Furnace Tubes, Journal of Pressure Vessel Technology, Vol. 127, pg no. 49-54, 2005, doi.org/10.1115/1.1845475

-
- [17] API Standards (American Petroleum Institute) API RP 580, Risk Based Inspection, First Edition, May 2002
- [18] H. Yun et al., Development of Wall-Thinning Evaluation Procedure for Nuclear Power Plant Piping - Part 2: Local Wall Thinning Estimation Method, Nuclear Engineering and Technology, Vol. 17, pg no.420-431, 2020, doi.org/10.1016/j.net.2020.03.001
- [19] Mohyaldin, M.E., Coupling Norsok CO₂ Corrosion Prediction Model with Pipelines Thermal/Hydraulic Models to Simulate CO₂ Corrosion Along Pipelines, Journal of Engineering Science and Technology, Vol. 6, pg no. 709 - 719, 2011
- [20] Keating, A. et al., Prediction of Two-Phase Erosion-Corrosion in Bends, CORROSION Vol. 57 (7), pg no. 621–633, 2001, doi.org/10.5006/1.3290389
- [21] Al Sayed, T. et al. Proposed Guideline for Identification and Assessment of Dead-Legs in Process Piping, Abu Dhabi International Petroleum Exhibition & Conference, 2016, doi.org/10.2118/183512-MS
- [22] Schmeal, W.R. et al. Process Corrosion Control: Corrosion in Amine/Sour Gas Treating Contactors, Chemical Engineering Progress, Vol. 74, Paper ID no. 6077910, 1978
- [23] Fahr, A., A Comparison of Probability of Detection (POD) Data Determined Using Different Statistical Method, Scientific and Technical Aerospace Reports, Volume 32, Issue 7, pg no. 2834, 1993.
- [24] Gosselin, S.R., Performance Demonstration Based Probability of Detection (POD) Curves for Fatigue Cracks in Piping, Proceedings of ASME 2005 Pressure Vessels and Piping Conference, pg no. 53-64, Denver, Colorado, USA, July 2005, doi.org/10.1115/PVP2005-71027
- [25] De Silva, D., et al. Condition Assessment and Probabilistic Analysis to Estimate Failure Rates in Buried Pipelines, 5th National Conference of the Australasian Society for Trenchless Technology, pg no. 15, Sydney, Australia, 2002
- [26] Caleyó, F., et al. Probability Distribution of Pitting Corrosion Depth and Rate in Underground Pipelines: A Monte Carlo study, Corrosion Science Volume 51, Issue 9, pg no. 1925-1934, 2009, doi.org/10.1016/j.corsci.2009.05.019
- [27] Kemp, M., Tail Fitting Probability Distributions for Risk Management Purposes (Lecture Notes), Institute and Faculty of Actuaries, (Last updated June 2016), Available at: <https://www.actuaries.org.uk/system/files/field/document/A4RiskKEmp.pdf> (Access: 25-May-2020).
- [28] Dann M.R., Stochastic Corrosion Growth Modelling for Pipelines Using Mass Inspection Data, Reliability Engineering & System Safety Volume 180, 2018, Pages 245-254 doi.org/10.1016/j.ress.2018.07.012
- [29] API Standards (American Petroleum Institute) API 510 , Pressure Vessel Inspection Code: In-service Inspection, Rating, Repair, and Alteration, Tenth Edition, May 2014
- [30] Witek, M., Steel Pipeline Failure Probability Evaluation Based on In-Line Inspection Results, Proceeding's 13th Pipeline Technology Conference, pg no. 5, Berlin, 2018
- [31] Nacereddine, N., et al. Unsupervised Weld Defect Classification in Radiographic Images Using Multivariate Generalized Gaussian Mixture Model with Exact Computation of Mean and Shape Parameters, Computers in Industry, Vol. 108, pg no. 132-149, June 2019, doi.org/10.1016/j.compind.2019.02.010

-
- [32] Al-Hashem, et al., The Effect of Water-Cut on the Corrosion Behavior of L80 Carbon Steel Under Downhole Conditions, CORROSION 2000 (CD-ROM), Orlando, Florida, March 2000
- [33] H. Yun et al., Development of Wall-Thinning Evaluation Procedure for Nuclear Power Plant Piping - Part 2: Local Wall Thinning Estimation Method, Nuclear Engineering and Technology, Vol. 17, pg no. 420-431, March 2020, doi.org/10.1016/j.net.2020.03.001
- [34] Yi, W.G., A Study on the Ultrasonic Thickness Measurement of Wall Thinned Pipe in Nuclear Power Plants, Proceedings of the 12th Asia-Pacific Conference on Non-Destructive Testing, pg no. 5, Auckland, New Zealand, November 2006
- [35] Hun Yun, Seung-Jae Moon et al., Development of Wall-Thinning Evaluation Procedure for Nuclear Power Plant Piping, Nuclear Engineering and Technology Volume 48, Issue 3, pg no. 820-830, 2016, doi.org/10.1016/j.net.2016.01.020
- [36] Crone, S.F., Training Artificial Neural Networks for Time Series Prediction Using Asymmetric Cost Functions, Proceedings of the 9th International Conference on Neural Information Processing, Paper ID No. 8030558, Singapore, November 2002, doi.org/10.1109/ICONIP.2002.1201919
- [37] Al Sayed, T. et al. Proposed Guideline for Identification and Assessment of Dead-Legs in Process Piping, Abu Dhabi International Petroleum Exhibition & Conference, Paper ID No. SPE-183512-MS, 2016 doi.org/10.2118/183512-MS
- [38] NACE Standards (National Association of Corrosion Engineers) NACE MR0175, Petroleum and Natural Gas Industries — Materials for Use in H₂S-Containing Environments in Oil and Gas Production, 1975
- [39] Sorensen, J.D., Probabilistic Analysis - Uncertainty Modelling - Statistics, Probability, Stochastic Variables, Lecture Notes, Training Course at TWI, pg no. 38, 2015.

APPENDIX A Table of Sample Piping System Results

| Name | AverageCR | InAverageCR | SDCR | SDInCR | count |
|-------------|------------------|--------------------|-------------|---------------|--------------|
| Pipe_A | 0.065468 | -2.31373 | 0.051876 | 0.722574 | 171 |
| Pipe_B | 0.017589 | -2.78851 | 0.005591 | 0.782237 | 8 |
| Pipe_C | 0.070554 | -2.38106 | 0.029472 | 0.623202 | 227 |
| Pipe_D | 0.009037 | -1.61217 | 0.014483 | 6.67E-06 | 14 |
| Pipe_E | 0.029317 | -2.7795 | 0.009797 | 0.617529 | 12 |
| Pipe_F | 0.114406 | -2.22434 | 0.02648 | 0.611624 | 35 |
| Pipe_G | 0.052442 | -2.82515 | 0.049416 | 1.200756 | 212 |
| Pipe_H | 0.151934 | -1.92195 | 0.01757 | 0.294212 | 22 |
| Pipe_I | 0.048717 | -2.54807 | 0.044553 | 0.736954 | 395 |
| Pipe_J | 0.041035 | -2.80745 | 0.03082 | 0.699943 | 36 |
| Pipe_M | 0.03894 | -3.27321 | 0.025175 | 0.659749 | 586 |
| Pipe_N | 0.02507 | -2.21332 | 0.012216 | 0.447083 | 12 |
| Pipe_O | 0.083964 | -2.35704 | 0.029869 | 0.699034 | 59 |
| Pipe_P | 0.079491 | -2.56622 | 0.057232 | 1.255287 | 64 |
| Pipe_Q | 0.033605 | -2.2264 | 0.021096 | 0.494004 | 17 |

Table A-1 Aggregated data used in Appendix B and Appendix C

APPENDIX B Neural Network Results

Accuracy of Neural Network With 3 Available Readings

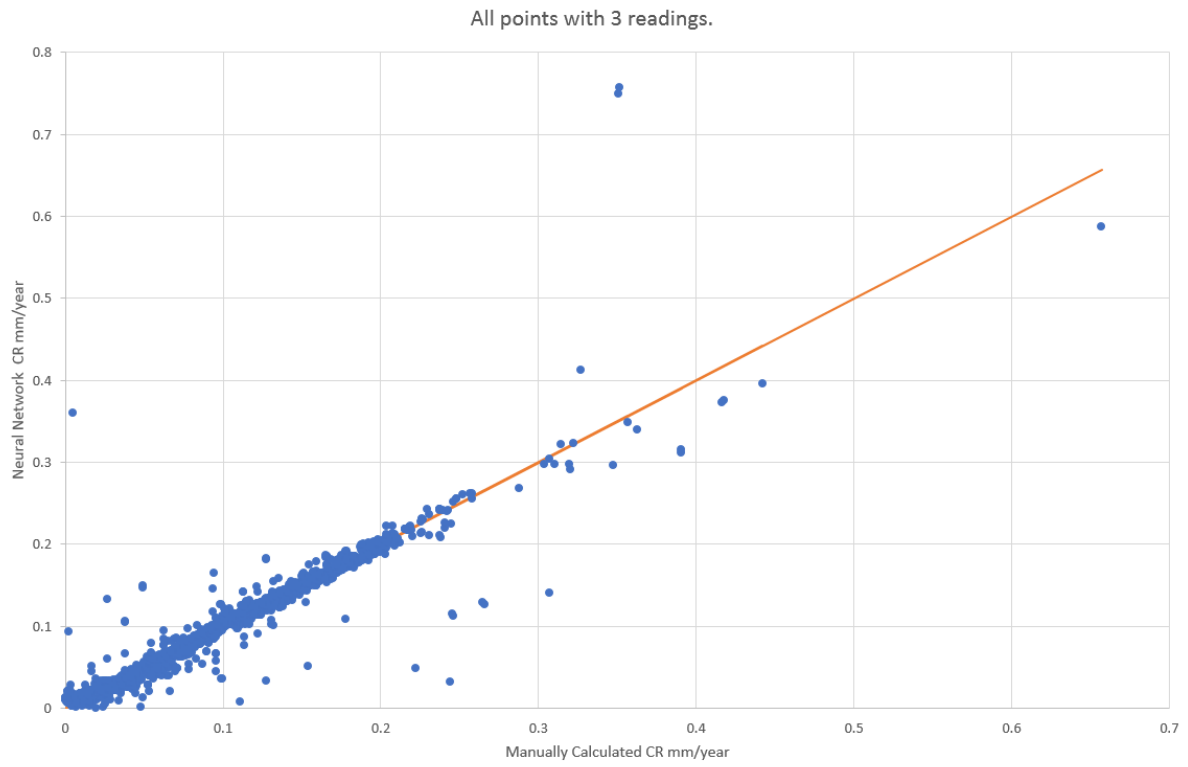


Figure B-0-1 All Corrosion rates vs Predicted Corrosion Rates where 3 readings have been taken

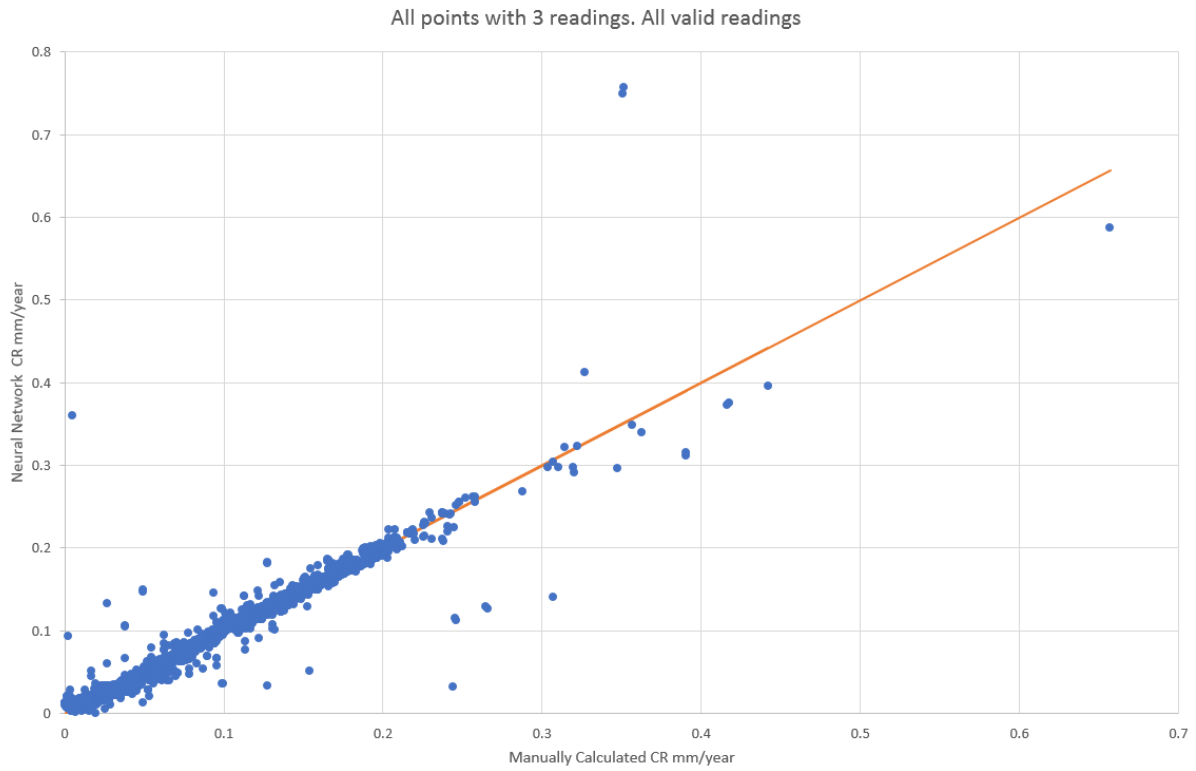


Figure B-0-2 All Corrosion rates vs Predicted Corrosion Rates where 3 readings, all readings

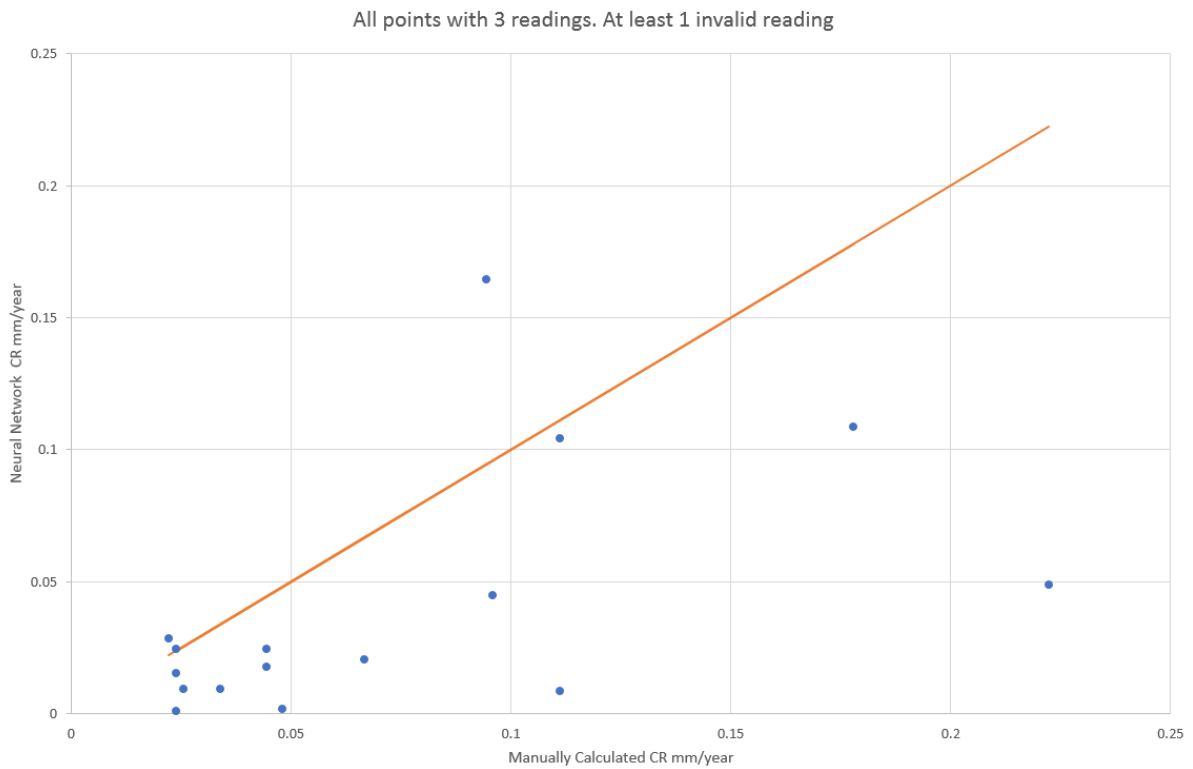


Figure B-0-3 All Corrosion rates vs Predicted Corrosion Rates where 3 readings, excluding readings

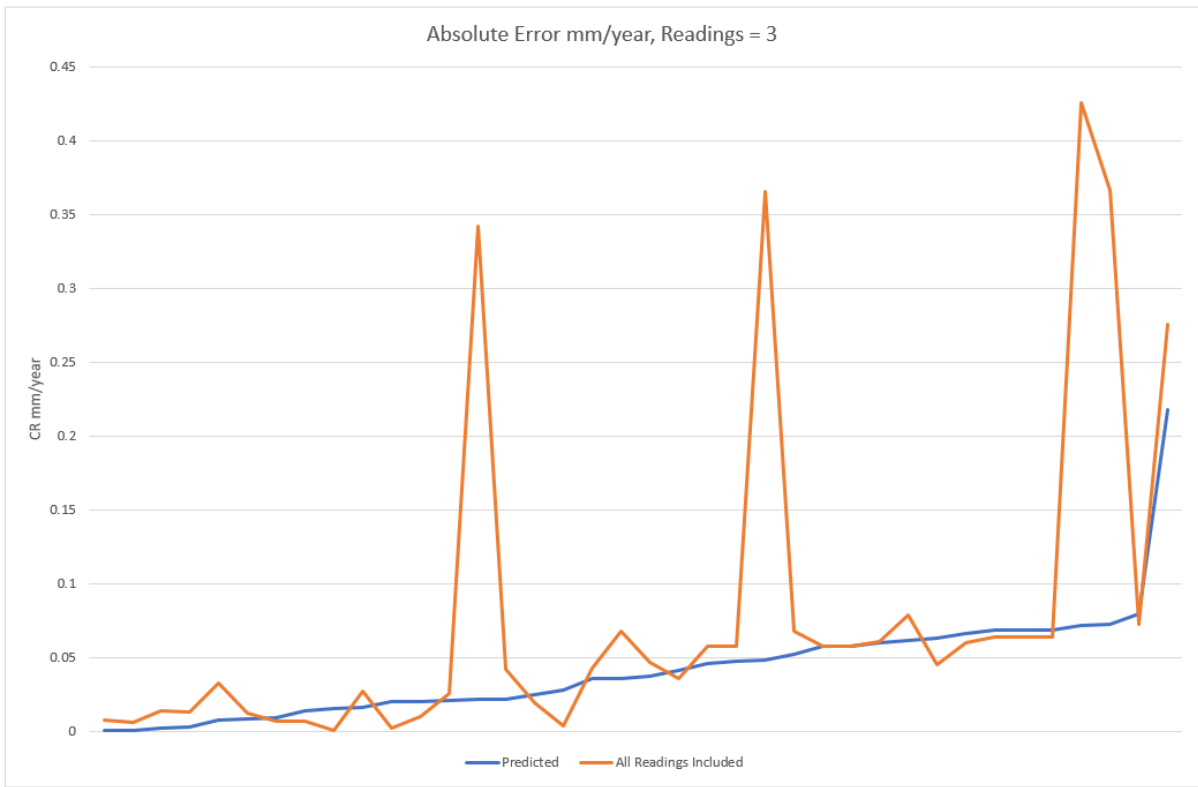


Figure B-0-4 Comparison of manual and automatically determined corrosion rates

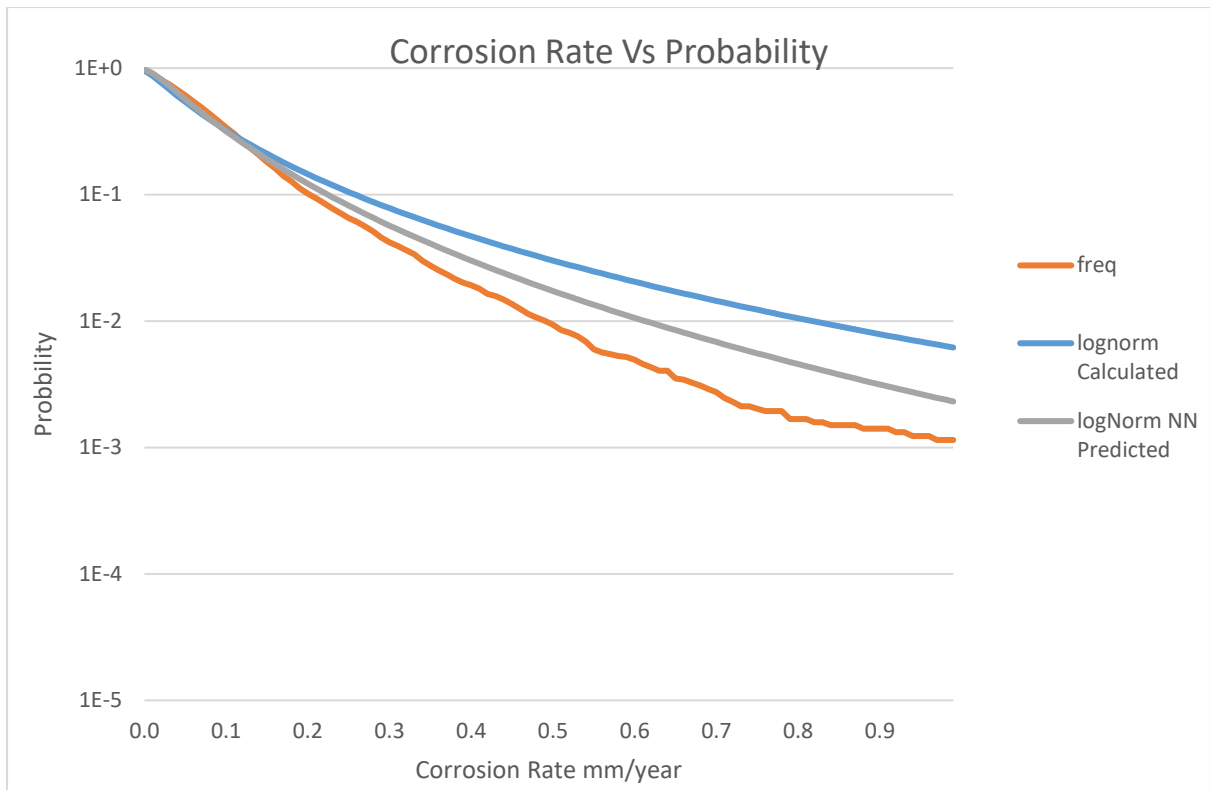


Figure B-0-5 The predicted lognormal distribution curve compared to the Calculated lognormal distributed curve and actual frequency of Corrosin Rates

Accuracy of Neural Network With 4 Available Readings

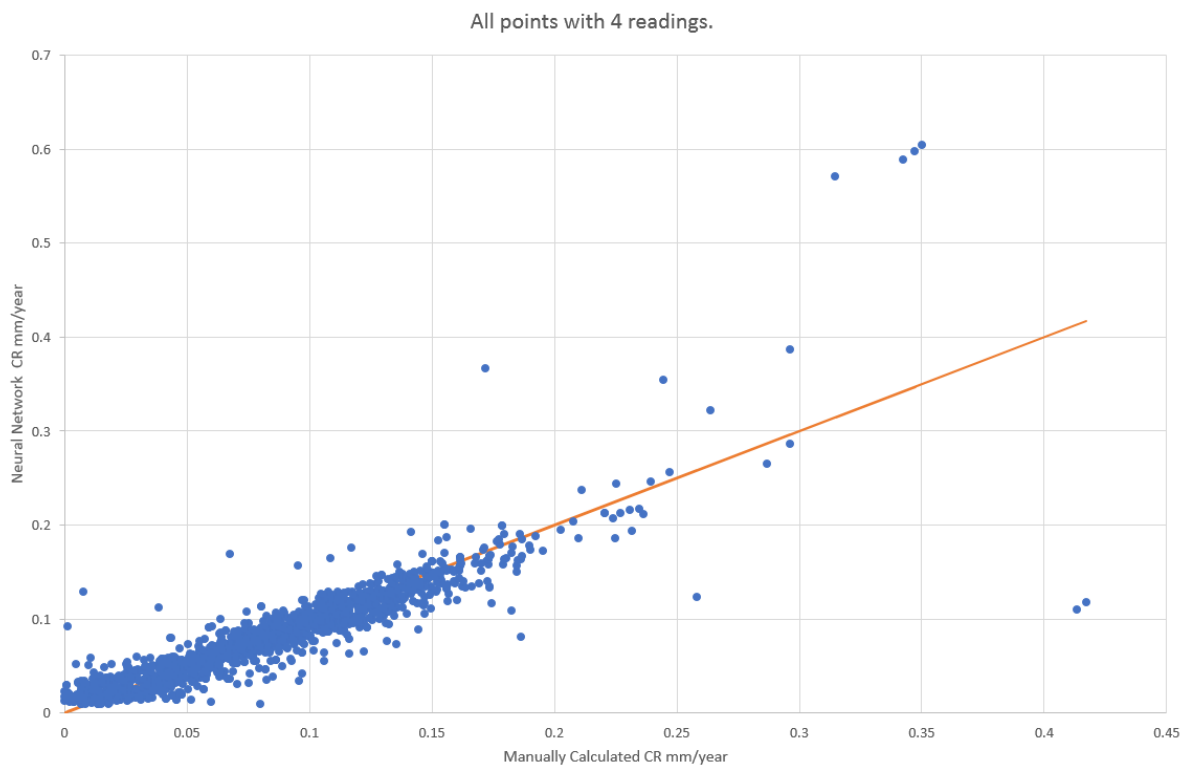


Figure B-0-6 All Corrosion rates vs Predicted Corrosion Rates where 4 readings have been taken

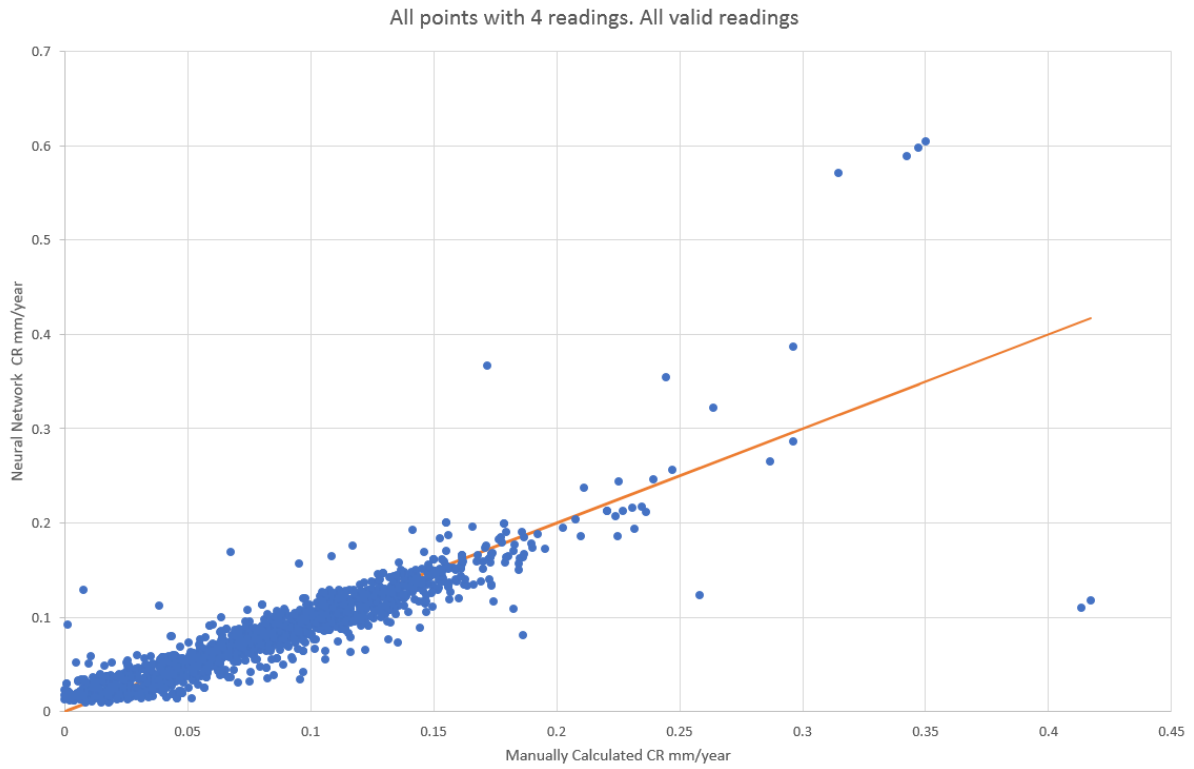


Figure B-0-7 All Corrosion rates vs Predicted Corrosion Rates where 4 readings, all readings

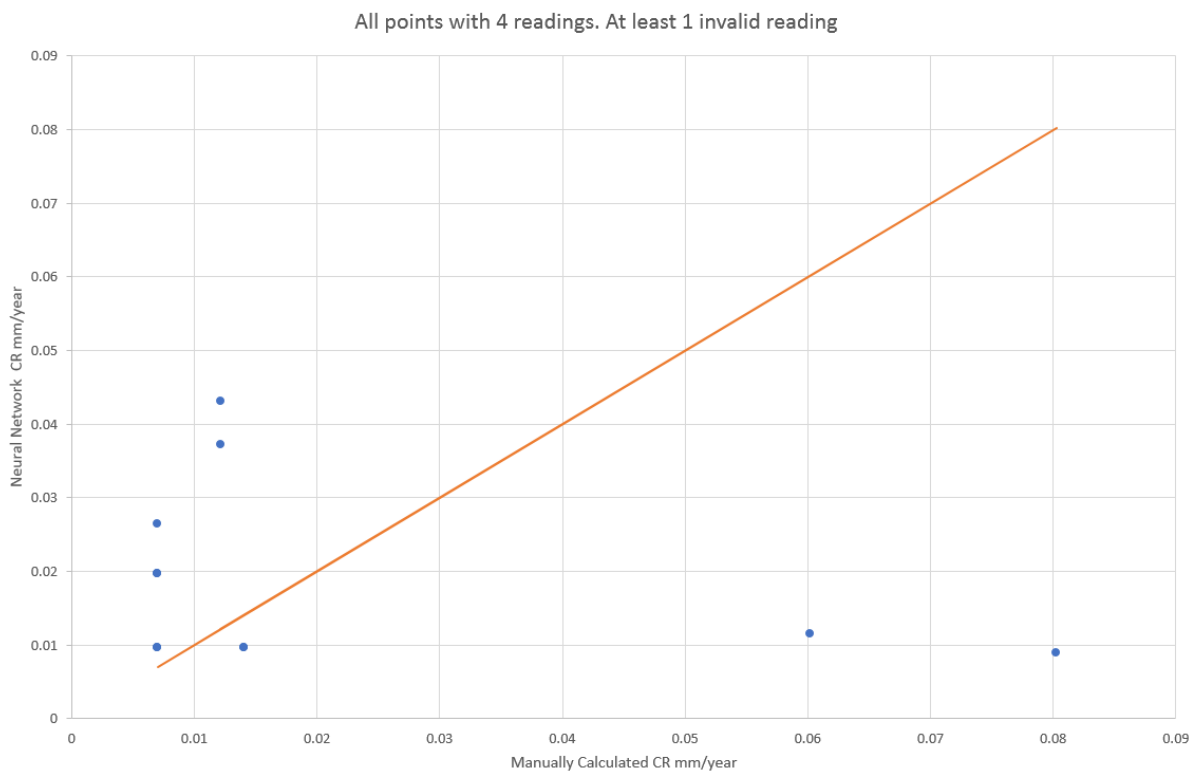


Figure B-0-8 All Corrosion rates vs Predicted Corrosion Rates where 4 readings, excluding readings

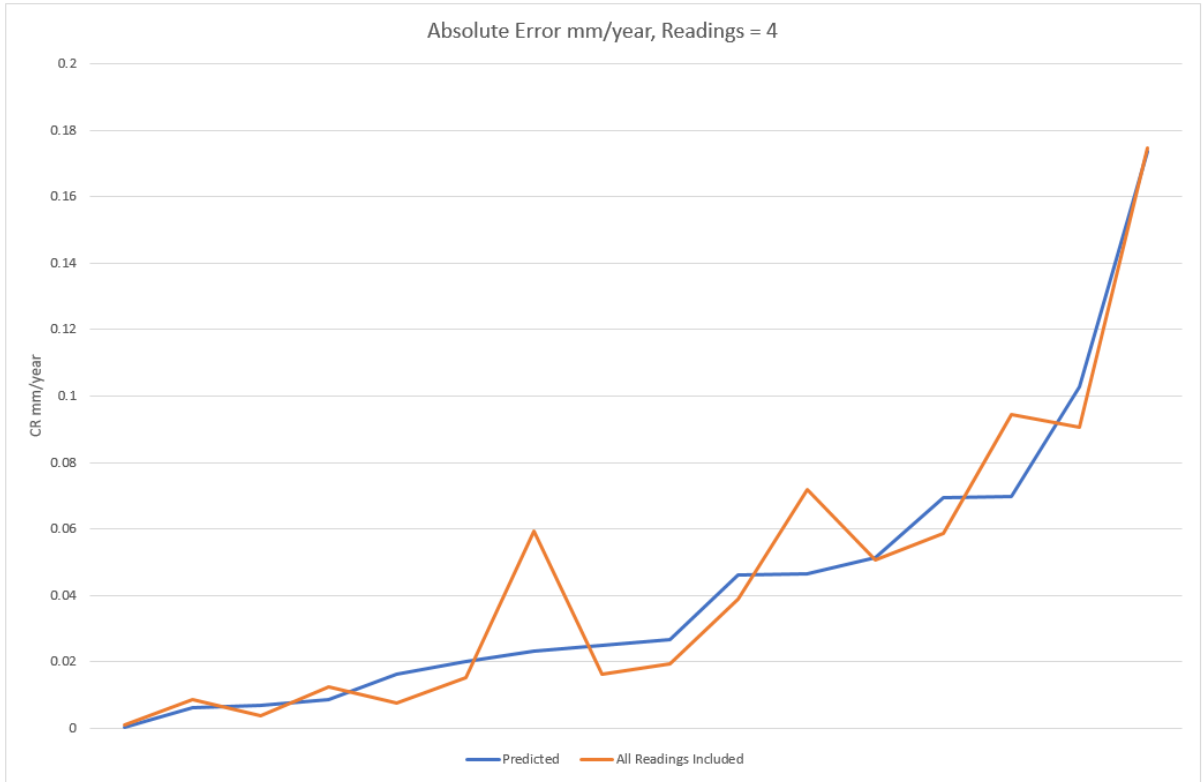


Figure B-0-9 Comparison of manual and automatically determined corrosion rates

Accuracy of Neural Network With 5 Available Readings

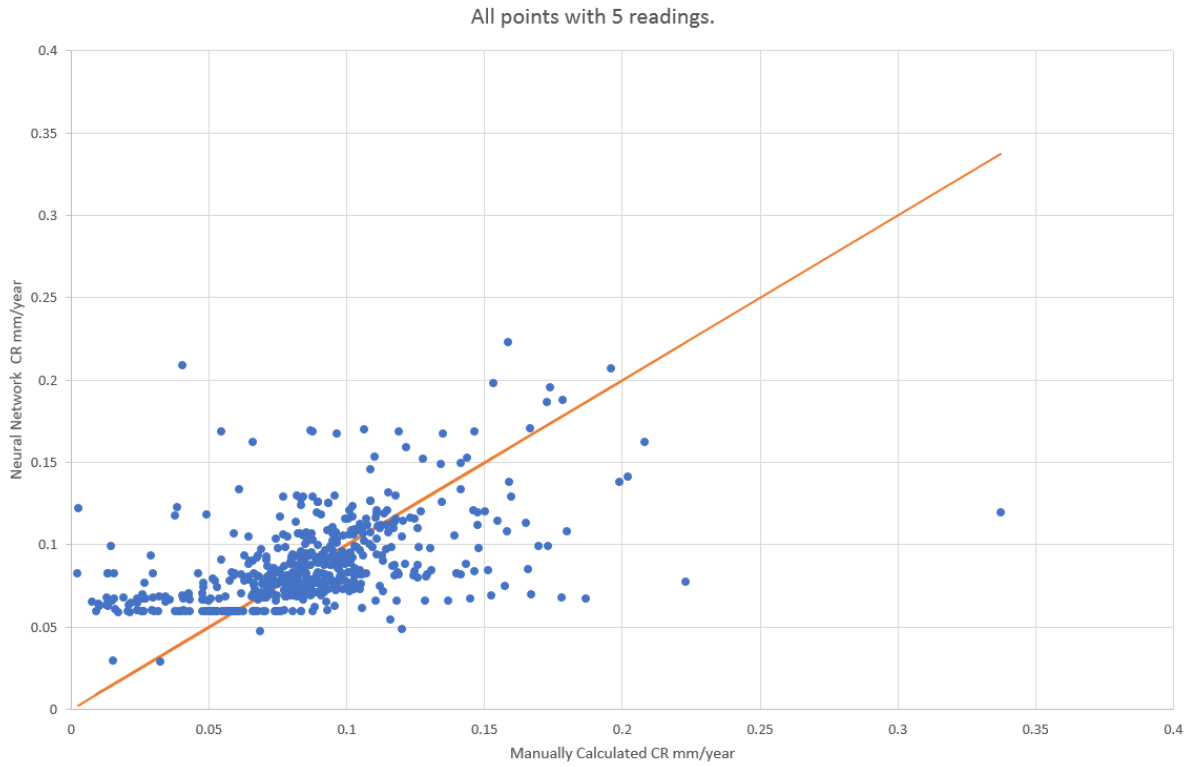


Figure B-0-10 All Corrosion rates vs Predicted Corrosion Rates where 5 readings have been taken

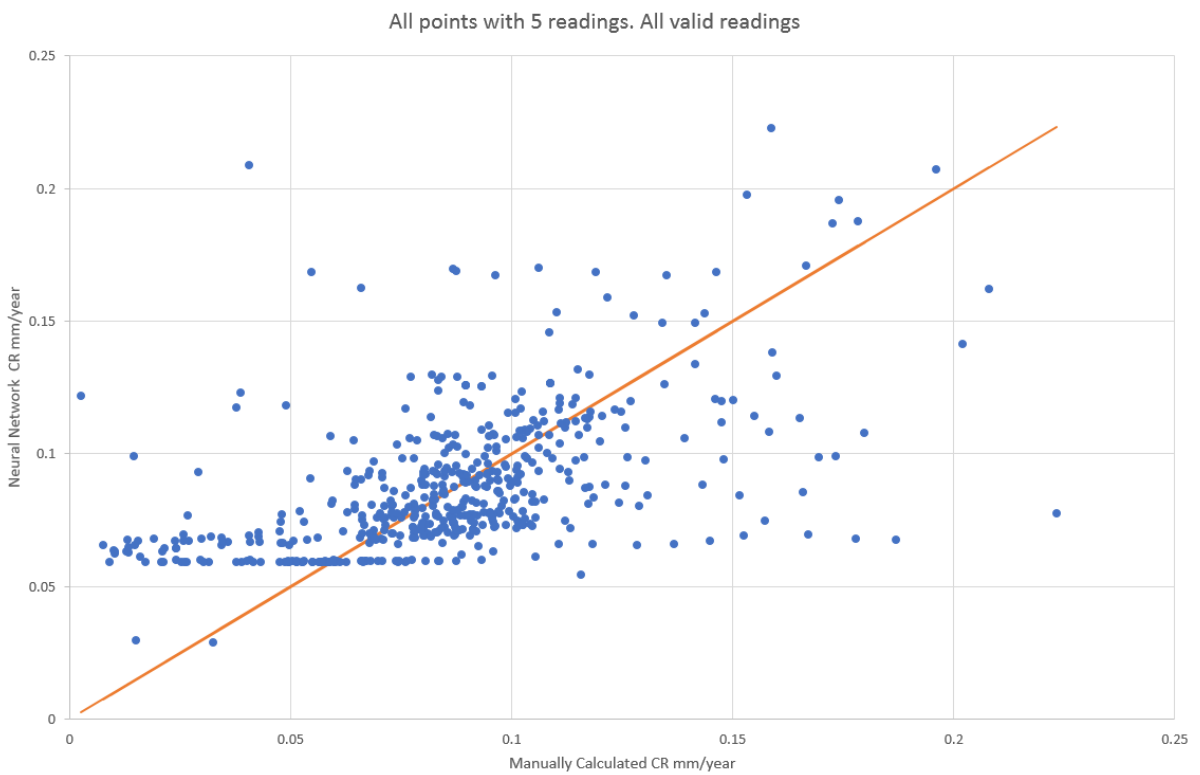


Figure B-0-11 All Corrosion rates vs Predicted Corrosion Rates where 5 readings, all readings

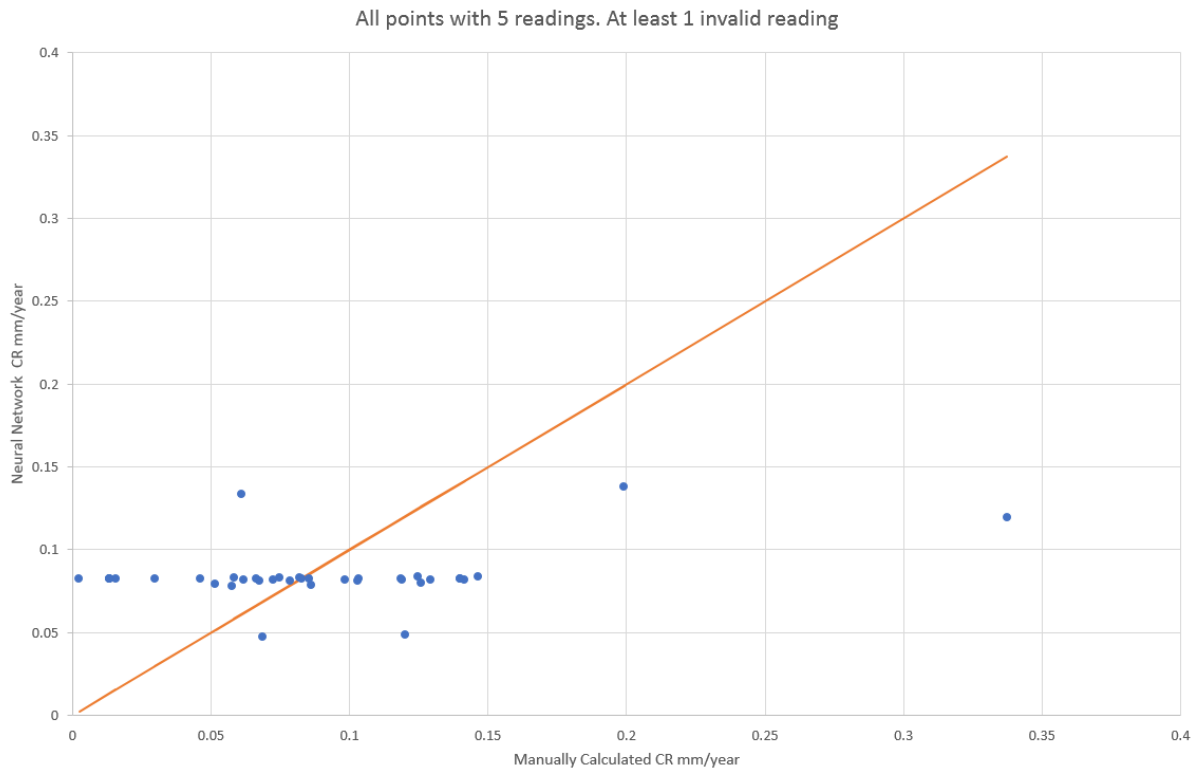
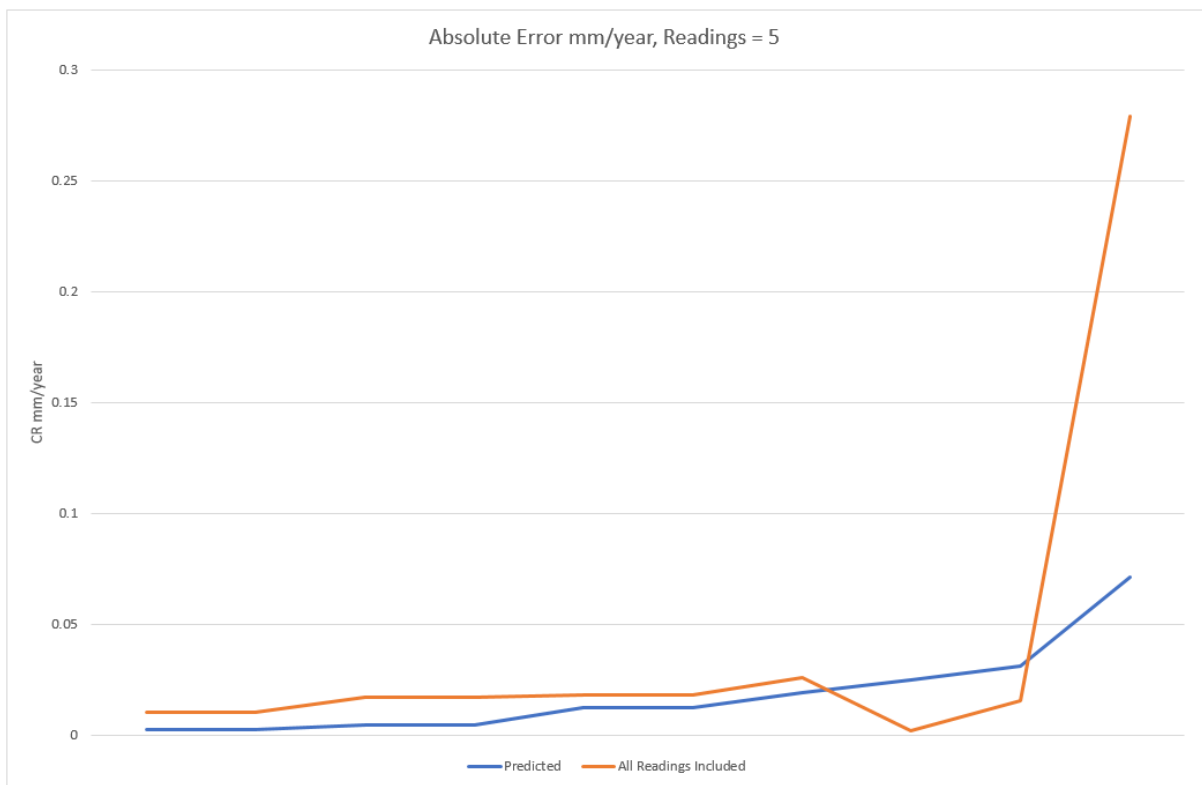


Figure B-0-12 All Corrosion rates vs Predicted Corrosion Rates where 5 readings, excluding readings



FigureB-0-13 Comparison of manual and automatically determined corrosion rates

Accuracy of Neural Network With 6 Available Readings

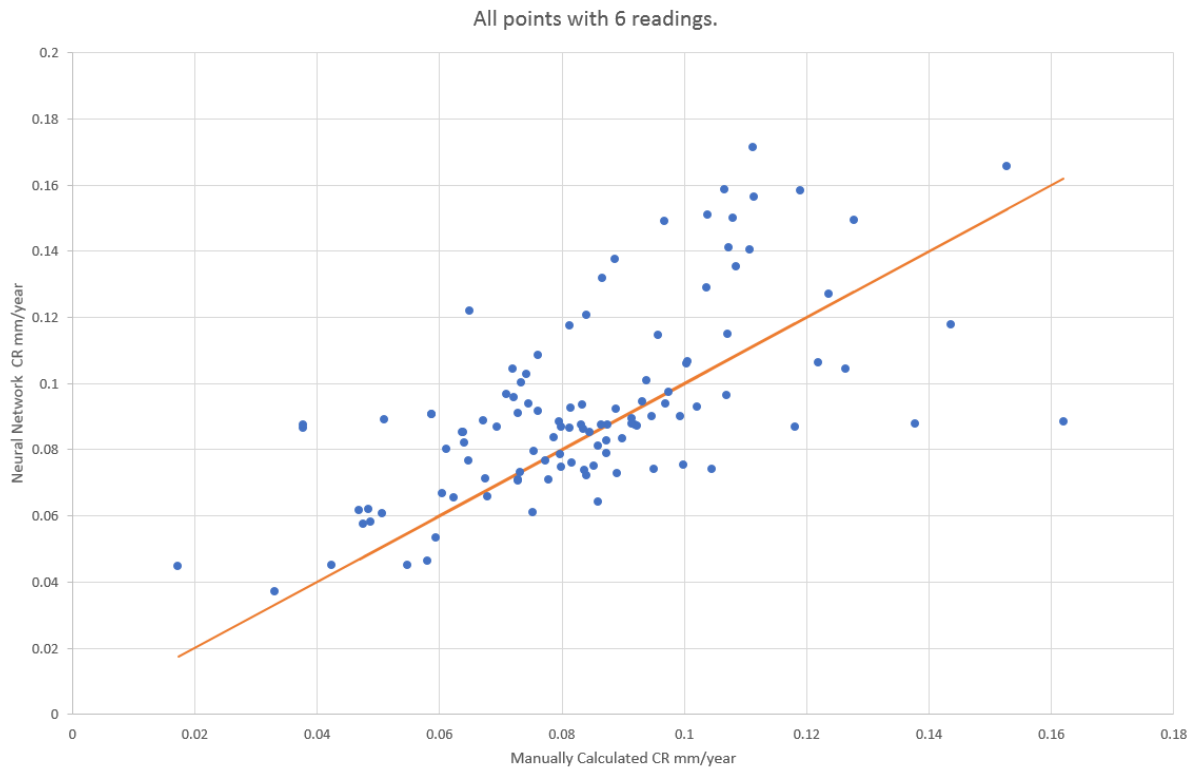


Figure B-0-14 All Corrosion rates vs Predicted Corrosion Rates where 6 readings have been taken

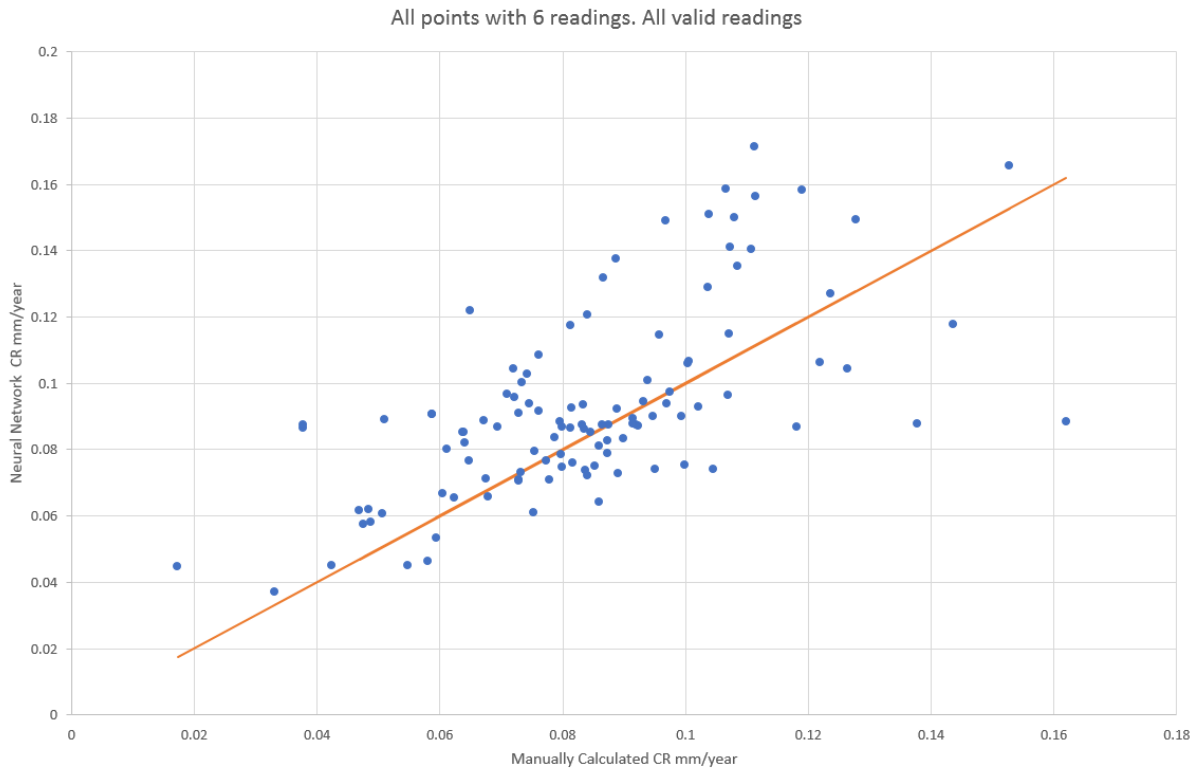


Figure B-0-15 All Corrosion rates vs Predicted Corrosion Rates where 6 readings all readings

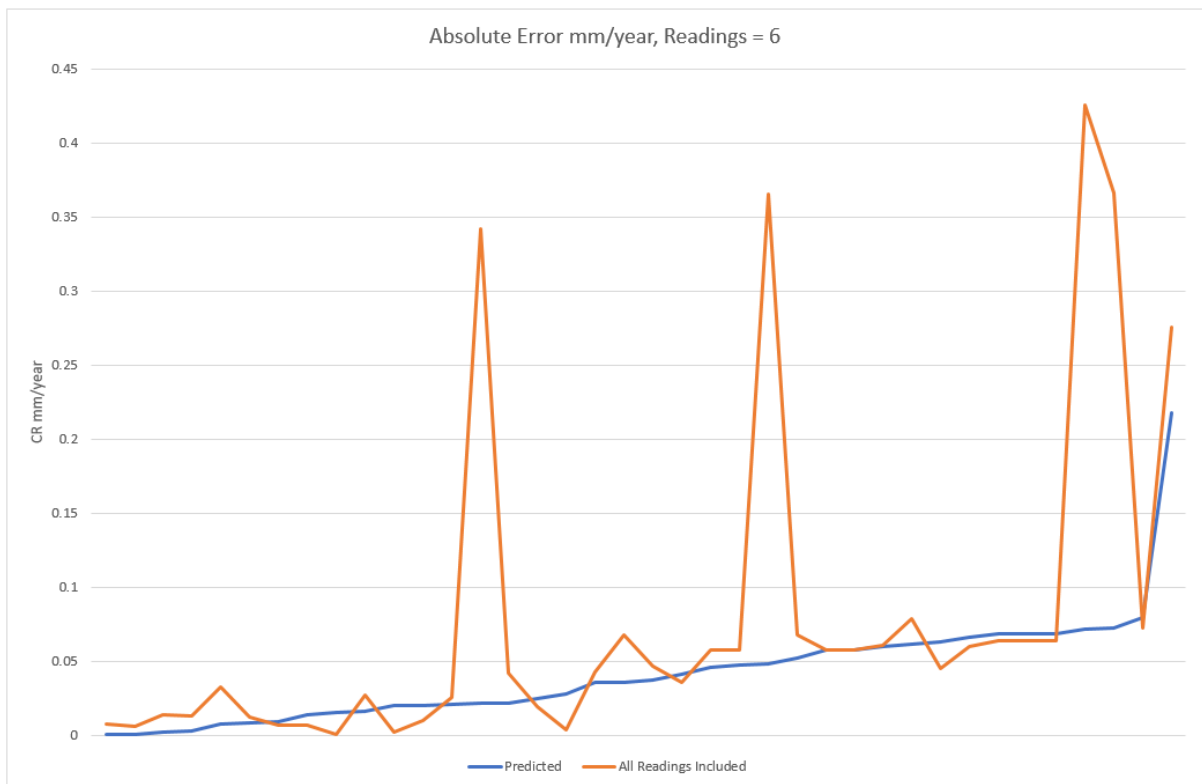


Figure B-0-16 Comparison of manual and automatically determined corrosion rates

APPENDIX C Corrosion Rate Distributions

Straight Sections

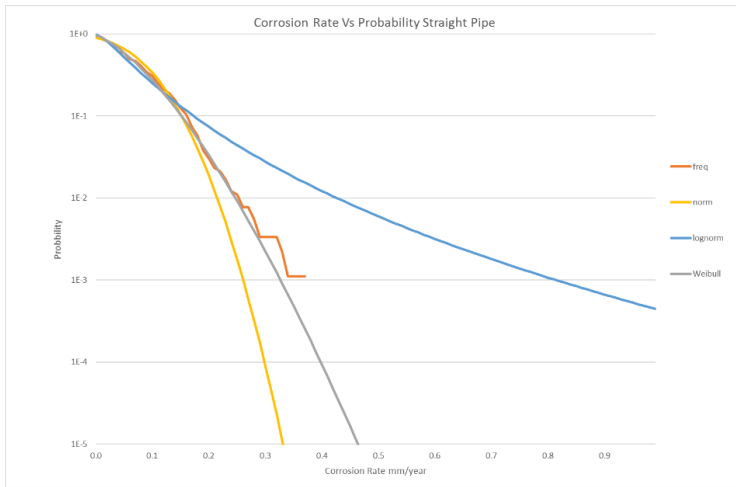


Figure C-1 Corrosion Rate Distribution of Straight Pipe

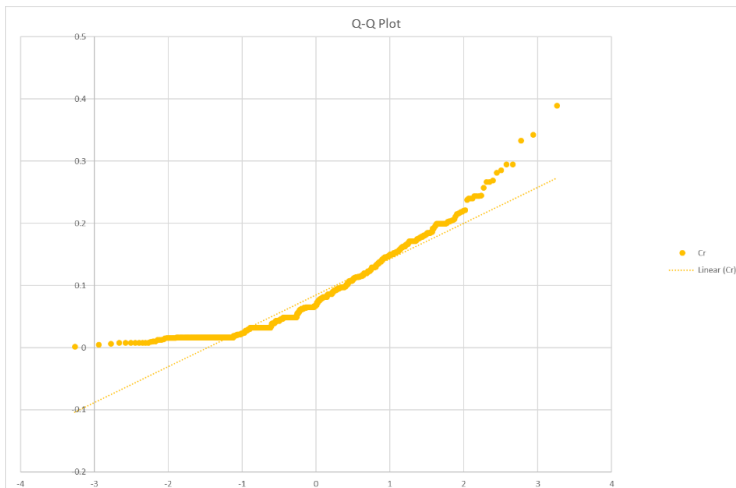


Figure C-2 Normal Distribution Q-Q Plot Straight Pipe

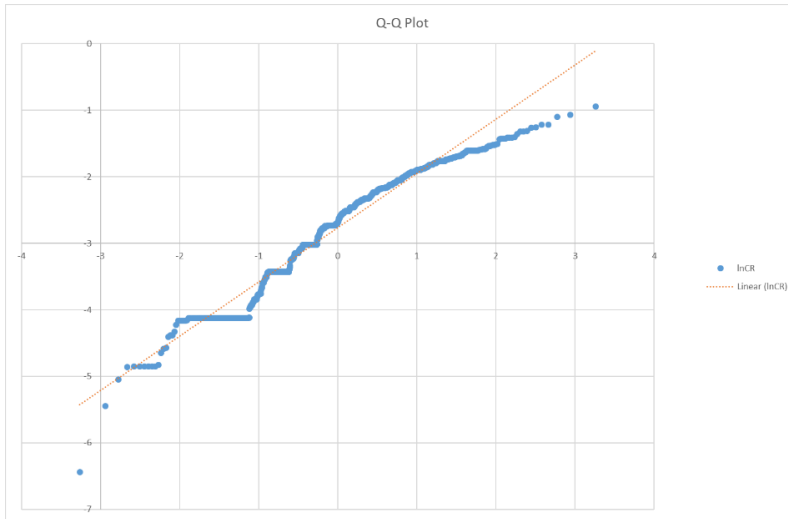


Figure C-3 Log Normal Distribution Q-Q Plot Straight Pipe

Fittings

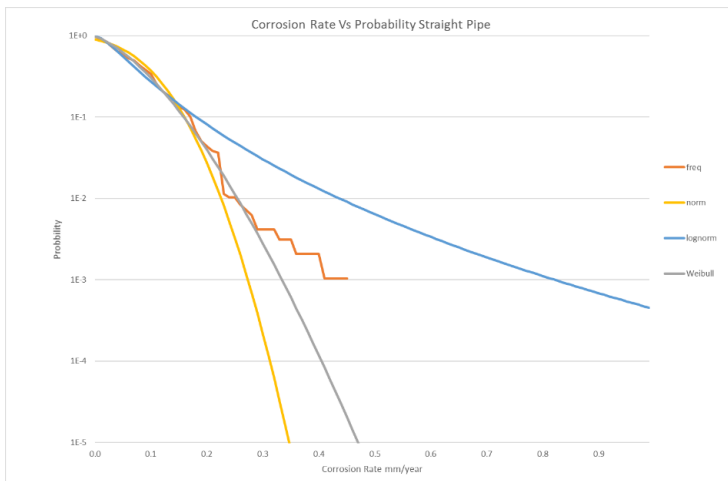


Figure C-4 Corrosion Rate Distribution of Fittings

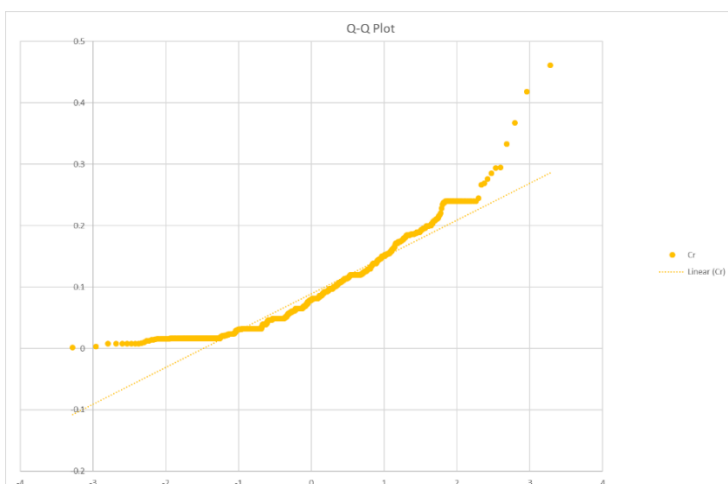


Figure C-5 Normal Distribution Q-Q Plot Pipe Fittings

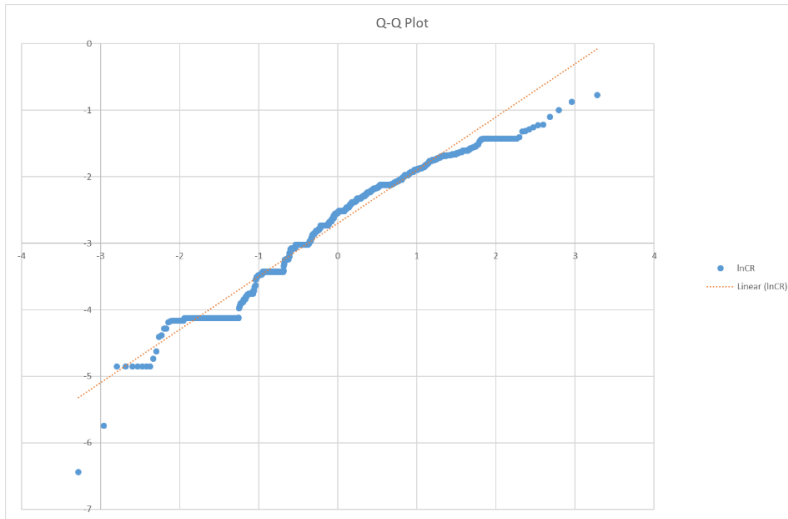


Figure C-6 Log Normal Distribution Q-Q Plot Pipe Fittings

Full Dataset

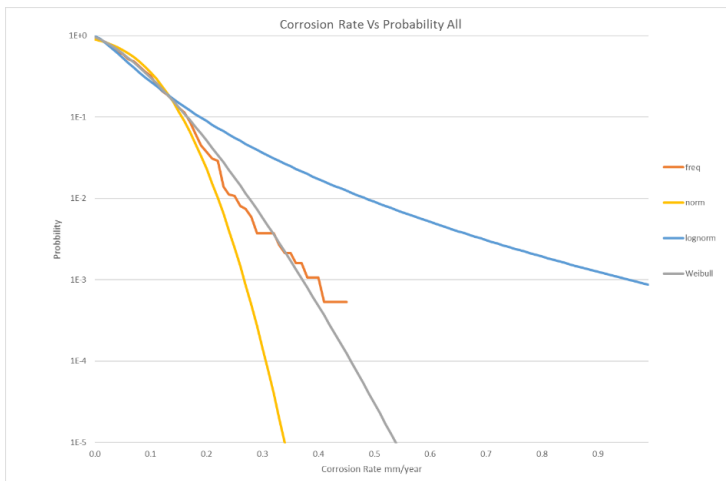


Figure C-7 Corrosion Rate Distribution of Entire System

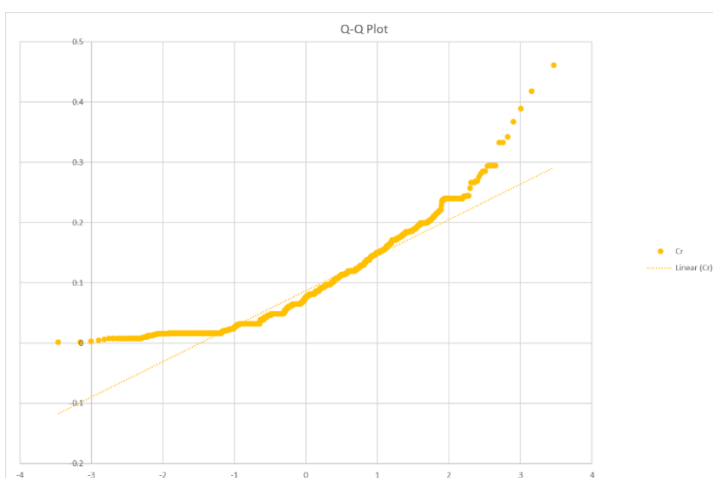


Figure C-8 Normal Distribution Q-Q Plot Full Dataset

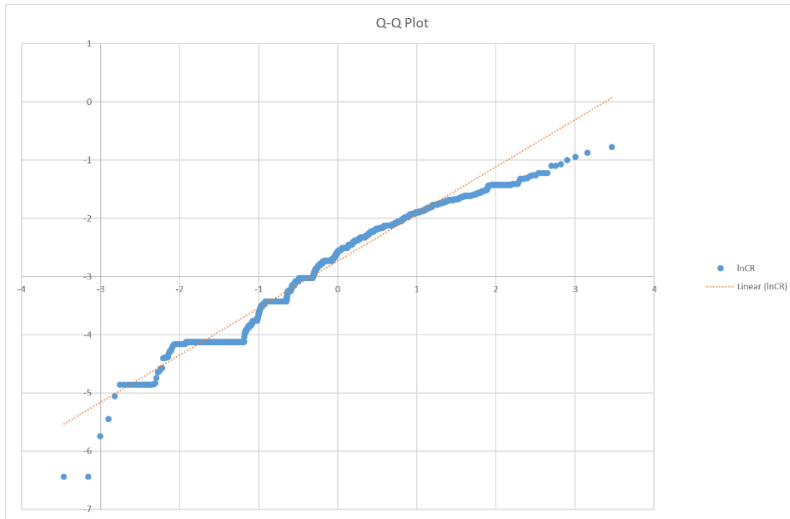


Figure C-9 Log Normal Distribution Q-Q Plot Full Dataset

APPENDIX D Code Used for Calculation

SQL Code For Data Normalisation

```
USE MASTERS
```

```
IF OBJECT_ID('tempdb..#TempTable') IS NOT NULL DROP TABLE #TempTable;  
IF OBJECT_ID('tempdb..#TempTable2') IS NOT NULL DROP TABLE #TempTable2;
```

```
Create Table #TempTable (  
  [PointID]    nvarchar(255) not null  
  , [Date]     datetime        not null  
  , [N]        int              not null  
); Insert Into #TempTable
```

```
Select  
  pointID  
  , min([date])  
  , count(PointID)
```

```
from masters.DBO.Thickness  
Where masters.DBO.Thickness.MinReading is not null  
group by masters.DBO.Thickness.PointID;
```

```
Create Table #TempTable2 (  
  [PointID]    nvarchar(255) not null  
  , [deltaY]   float          not null  
  , [tc]       float          null  
  , [N]        int              not null  
  , [Ntot]     int              null  
);
```

```
Insert Into #TempTable2  
select  
  t.PointID  
  , datediff(day, tt.[Date],t.[date])/365.25 [deltaY]  
  , t.MinReading [tc]  
  , ROW_NUMBER() over(Partition by t.PointID order by t.pointID, datediff(day,  
  tt.[Date],t.[date])/365.25) [N]  
  , tt.N [Ntot]  
from masters.DBO.Thickness t, #TempTable tt  
where  
  t.PointID = tt.PointID
```

```

and t.MinReading is not null
and tt.N > 2
order by pointID, datediff(day, tt.[Date],t.[date])/365.25;

--select #TempTable2.Ntot, count(#TempTable2.PointID)/#TempTable2.Ntot from
#TempTable2 group by #TempTable2.Ntot order by Ntot

--3    34830
--4    16191
--5    4061
--6    1483
--7    275
--8    57
--9    4

select
r1.PointID [pointID], r1.Ntot [Ntot]
, r1.deltaY [Y1], r1.tc [t1]
, r2.deltaY [Y2], r2.tc [t2]
, r3.deltaY [Y3], r3.tc [t3]
, r4.deltaY [Y4], r4.tc [t4]
, r5.deltaY [Y5], r5.tc [t5]
, r6.deltaY [Y6], r6.tc [t6]
, r7.deltaY [Y7], r7.tc [t7]
, r8.deltaY [Y8], r8.tc [t8]
, r9.deltaY [Y9], r9.tc [t9]

from
#TempTable2 r1
,#TempTable2 r2
,#TempTable2 r3
,#TempTable2 r4
,#TempTable2 r5
,#TempTable2 r6
,#TempTable2 r7
,#TempTable2 r8
,#TempTable2 r9

where
r1.N = 1
And r2.N = 2

```

And r3.N = 3
And r4.N = 4
And r5.N = 5
And r6.N = 6
And r7.N = 7
And r8.N = 8
And r9.N = 9

And r1.Ntot = 9

And r1.PointID = r2.PointID
And r1.PointID = r3.PointID
And r1.PointID = r4.PointID
And r1.PointID = r5.PointID
And r1.PointID = r6.PointID
And r1.PointID = r7.PointID
And r1.PointID = r8.PointID
And r1.PointID = r9.PointID

R Code for corrosion rate calculation

```
#CHECK IF ALREADY INSTALLED
```

```
#Steps for talking to the sql DB
```

```
#install.packages("RODBC")
```

```
setwd("C:/@develop/Phd/R")
```

```
getwd()
```

```
library(RODBC)
```

```
con <- RODBC::odbcConnect("R")
```

```
#RODBC::sqlQuery(con,"select count(*) from masters.DBO.THICKNESS")
```

```
#load point thickness and thickness summary tables
```

```
sqlThickness = "SELECT T.POINTID, T.DATE, T.MINREADING, T.Valid FROM MASTERS.DBO.tTHICKNESS  
T WHERE POINTID in (select pointID from masters.DBO.points) ORDER BY T.MINREADING DESC"
```

```
sqlPoints = "select p.PointID ,p.LineDesignation ,p.DesignPressure ,p.[System] ,f.[Type], f.[Straight]  
,m.code [Material] ,s.NPS ,p.Tnom ,sy.[service] from masters.DBO.points p ,masters.DBO.FittingDesc  
f ,masters.DBO.Material m ,masters.DBO.Size s ,masters.DBO.[System] sy where POINTID in (select  
Distinct pointID from masters.DBO.Thickness) and p.FittingDesc = f.FittingDesc and p.Material =  
m.material and p.Size = s.Size and p.[System] = sy.[system] ORDER BY P.PointID DESC"
```

```
sqlTsum = "select t.PointID PointID, count(*) N, Sum(datediff(DAY,'1970-01-01',t.[date])/365.25)  
sumx, Sum(t.MinReading) sumy , Sum(power(datediff(DAY,'1970-01-01',t.[date])/365.25,2)) sumx2 ,  
Sum(power(t.MinReading,2)) sumy2 , Sum(datediff(DAY,'1970-01-  
01',t.[date])*t.MinReading/365.25) sumxy from Masters.dbo.Thickness t where Valid = 1 Group by  
t.PointID"
```

```
#get thickness from DB
```

```
Thickness<-data.frame(RODBC::sqlQuery(con,sqlThickness))
```

```
Points<-data.frame(RODBC::sqlQuery(con,sqlPoints))
```

```
TSum<-data.frame(RODBC::sqlQuery(con,sqlTsum))
```

```
Tnorm<-read.csv("Tnorm.csv")
```

```
TSum = TSum[TSum$N > 1,]
```

```
#Read in Values generated by normalise.sql
```

```
Thick3<-read.csv("Thick3.csv")
```

```
Thick4<-read.csv("Thick4.csv")
```

```
Thick5<-read.csv("Thick5.csv")
```

```
Thick6<-read.csv("Thick6.csv")
```

```
Thick7<-read.csv("Thick7.csv")
```

```
Thick8<-read.csv("Thick8.csv")
```

```
Thick9<-read.csv("Thick9.csv")
```

```
#Calculate corrosion rate from sql query results from TSUM
```

```
TSum$CR<- -1 * ((TSum$N*TSum$sumxy) -(TSum$sumx*TSum$sumy))/((TSum$N*TSum$sumx2)-  
TSum$sumx^2)#((TSum$N*TSum$sumxy) -(TSum$sumx*TSum$sumy))/((TSum$N*TSum$sumx2)-1)  
#(TSum$sumx^2)
```

```
TSum$RSQ <- ((TSum$N*TSum$sumxy-TSum$sumx*TSum$sumy)^2)/((TSum$N*TSum$sumx2-  
TSum$sumx^2)*(TSum$N*TSum$sumy2-TSum$sumy^2))
```

```
TSum = TSum[TSum$CR > 0,]
```

```
TSum = TSum[TSum$RSQ > 0,]
```

```
TSum = TSum[TSum$RSQ <= 1,]
```

```
#Add calculated CR to Thick Table for each N
```

```
#tidy up data frames
```

```
names(Thick3)[names(Thick3) == "i..pointID"] <- "PointID"
```

```
names(Thick4)[names(Thick4) == "i..pointID"] <- "PointID"
```

```
names(Thick5)[names(Thick5) == "i..pointID"] <- "PointID"
```

```
names(Thick6)[names(Thick6) == "i..pointID"] <- "PointID"
```

```
names(Thick7)[names(Thick7) == "i..pointID"] <- "PointID"
```

```
names(Thick8)[names(Thick8) == "i..pointID"] <- "PointID"
```

```
names(Thick9)[names(Thick9) == "i..pointID"] <- "PointID"
```

```
Thick3<-merge(Thick3, TSum , by.x = "PointID", by.y = "PointID")
```

```
Thick3$sumx<-NULL
```

```
Thick3$sumy<-NULL
```

```
Thick3$sumxy<-NULL
```

```
Thick3$sumx2<-NULL
```

```
Thick3$sumy2<-NULL
```

```
Thick4<-merge(Thick4, TSum , by.x = "PointID", by.y = "PointID")
```

Thick4\$sumx<-NULL

Thick4\$sumy<-NULL

Thick4\$sumxy<-NULL

Thick4\$sumx2<-NULL

Thick4\$sumy2<-NULL

Thick5<-merge(Thick5, TSum , by.x = "PointID", by.y = "PointID")

Thick5\$sumx<-NULL

Thick5\$sumy<-NULL

Thick5\$sumxy<-NULL

Thick5\$sumx2<-NULL

Thick5\$sumy2<-NULL

Thick6<-merge(Thick6, TSum , by.x = "PointID", by.y = "PointID")

Thick6\$sumx<-NULL

Thick6\$sumy<-NULL

Thick6\$sumxy<-NULL

Thick6\$sumx2<-NULL

Thick6\$sumy2<-NULL

Thick7<-merge(Thick7, TSum , by.x = "PointID", by.y = "PointID")

Thick7\$sumx<-NULL

Thick7\$sumy<-NULL

Thick7\$sumxy<-NULL

Thick7\$sumx2<-NULL

```
Thick7$sumy2<-NULL
```

```
Thick8<-merge(Thick8, TSum , by.x = "PointID", by.y = "PointID")
```

```
Thick8$sumx<-NULL
```

```
Thick8$sumy<-NULL
```

```
Thick8$sumxy<-NULL
```

```
Thick8$sumx2<-NULL
```

```
Thick8$sumy2<-NULL
```

```
Thick9<-merge(Thick9, TSum , by.x = "PointID", by.y = "PointID")
```

```
Thick9$sumx<-NULL
```

```
Thick9$sumy<-NULL
```

```
Thick9$sumxy<-NULL
```

```
Thick9$sumx2<-NULL
```

```
Thick9$sumy2<-NULL
```

```
#get neural net
```

```
install.packages("neuralnet")
```

```
require(neuralnet)
```

```
#add in other paramters to Thick model
```

```
Thick3<-merge(Thick3, Points , by.x = "PointID", by.y = "PointID")
```

```
Thick4<-merge(Thick4, Points , by.x = "PointID", by.y = "PointID")
```



```
Thick5<-merge(Thick5, Points , by.x = "PointID", by.y = "PointID")
Thick6<-merge(Thick6, Points , by.x = "PointID", by.y = "PointID")
Thick7<-merge(Thick7, Points , by.x = "PointID", by.y = "PointID")
Thick8<-merge(Thick8, Points , by.x = "PointID", by.y = "PointID")
Thick9<-merge(Thick9, Points , by.x = "PointID", by.y = "PointID")
```

```
#generate random sample of thick
```

```
set.seed(222)
```

```
ind3 <- sample(2,nrow(Thick3),replace = TRUE, prob = c(0.05,0.95))
```

```
ind4 <- sample(2,nrow(Thick4),replace = TRUE, prob = c(0.05,0.95))
```

```
ind5 <- sample(2,nrow(Thick5),replace = TRUE, prob = c(0.05,0.95))
```

```
ind6 <- sample(2,nrow(Thick6),replace = TRUE, prob = c(0.05,0.95))
```

```
ind7 <- sample(2,nrow(Thick7),replace = TRUE, prob = c(0.05,0.95))
```

```
ind8 <- sample(2,nrow(Thick8),replace = TRUE, prob = c(0.05,0.95))
```

```
ind9 <- sample(2,nrow(Thick9),replace = TRUE, prob = c(0.05,0.95))
```

```
Thick3$AllValid<-(Thick3$Ntot == Thick3$N)
```

```
Thick4$AllValid<-(Thick4$Ntot == Thick4$N)
```

```
Thick5$AllValid<-(Thick5$Ntot == Thick5$N)
```

```
Thick6$AllValid<-(Thick6$Ntot == Thick6$N)
```

```
Thick7$AllValid<-(Thick7$Ntot == Thick7$N)
```

```
Thick8$AllValid<-(Thick8$Ntot == Thick8$N)
```

```
Thick9$AllValid<-(Thick9$Ntot == Thick9$N)
```

```

#First attempt, zoomed in on a specific type

#Training <- Thick3[which(Thick3$Type == "PIP" & Thick3$Material=="API 5L Gr.B"&Thick3$service
=="Gas"),]

Training3 <- Thick3[ind3==1,]

Training4 <- Thick4[ind4==1,]

Training5 <- Thick5[ind5==1,]

Training6 <- Thick6[ind6==1,]

Training7 <- Thick7[ind7==1,]

Training8 <- Thick8[ind8==1,]

Training9 <- Thick9[ind9==1,]

Testing3 <- Thick3[ind3==2,]

Testing4 <- Thick4[ind4==2,]

Testing5 <- Thick5[ind5==2,]

Testing6 <- Thick6[ind6==2,]

Testing7 <- Thick7[ind7==2,]

Testing8 <- Thick8[ind8==2,]

Testing9 <- Thick9[ind9==2,]

#sqr(n) layers and nodes

nn3 = neuralnet(CR~t1+Y2+t2+Y3+t3+Straight+NPS,data = Training3, hidden = c(7,7,7), linear.output
= FALSE)

nn4 = neuralnet(CR~t1+Y2+t2+Y3+t3+Y4+t4+Straight+NPS,data = Training4, hidden = c(9,9,9),
linear.output = FALSE)

```

```

nn5 = neuralnet(CR~t1+Y2+t2+Y3+t3+Y4+t4+Y5+t5+Straight+NPS,data = Training5, hidden =
c(11,11,11,11), linear.output = FALSE)

nn6 = neuralnet(CR~t1+Y2+t2+Y3+t3+Y4+t4+Y5+t5+Y6+t6+Straight+NPS,data = Training6, hidden =
c(13,13,13,13), linear.output = FALSE)

nn7 = neuralnet(CR~t1+Y2+t2+Y3+t3+Y4+t4+Y5+t5+Y6+t6+Y7+t7+Straight+NPS,data = Training7,
hidden = c(15,15,15,15), linear.output = FALSE)

nn8 = neuralnet(CR~t1+Y2+t2+Y3+t3+Y4+t4+Y5+t5+Y6+t6+Y7+t7+Y8+t8+Straight+NPS,data =
Training8, hidden = c(17,17,17,17), linear.output = FALSE)

```

```
#check network converged
```

```
plot(nn3)
```

```
plot(nn4)
```

```
plot(nn5)
```

```
plot(nn6)
```

```
plot(nn7)
```

```
plot(nn8)
```

```
#compare to real
```

```
Testing3$pred <- predict(nn3,Testing3[,c("t1","Y2","t2","Y3","t3","Straight","NPS")])
```

```
Testing4$pred <- predict(nn4,Testing4[,c("t1","Y2","t2","Y3","t3","Y4","t4","Straight","NPS")])
```

```
Testing5$pred <-
```

```
predict(nn5,Testing5[,c("t1","Y2","t2","Y3","t3","Y4","t4","Y5","t5","Straight","NPS")])
```

```
Testing6$pred <-
```

```
predict(nn6,Testing6[,c("t1","Y2","t2","Y3","t3","Y4","t4","Y5","t5","Y6","t6","Straight","NPS")])
```

```
Testing7$pred <-  
predict(nn7,Testing7[,c("t1","Y2","t2","Y3","t3","Y4","t4","Y5","t5","Y6","t6","Y7","t7","Straight","NPS")])
```

```
Testing8$pred <-  
predict(nn8,Testing8[,c("t1","Y2","t2","Y3","t3","Y4","t4","Y5","t5","Y6","t6","Y7","t7","Y8","t8","Straight","NPS")])
```

```
#output to excel
```

```
write.csv(Testing3,"predNN3OUT.csv")
```

```
write.csv(Testing4,"predNN4OUT.csv")
```

```
write.csv(Testing5,"predNN5OUT.csv")
```

```
write.csv(Testing6,"predNN6OUT.csv")
```

```
write.csv(Testing7,"predNN7OUT.csv")
```

```
write.csv(Testing8,"predNN8OUT.csv")
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
ggplot(Training,aes(x=CR,y=pred))
```

```
write.csv(Training,"NN3OUT.csv")
```

```
write.csv(Thick3,"TestingNN3OUT.csv")
```

```
sumTest3<-
```

```
Testing3[,c("PointID","Ntot","N","CR","RSQ","LineDesignation","DesignPressure","System","Type","Straight","Material","NPS","Tnom","service","AllValid","pred")]
```

```
sumTest4<-
```

```
Testing4[,c("PointID","Ntot","N","CR","RSQ","LineDesignation","DesignPressure","System","Type","Straight","Material","NPS","Tnom","service","AllValid","pred")]
```

```
sumTest5<-
```

```
Testing5[,c("PointID","Ntot","N","CR","RSQ","LineDesignation","DesignPressure","System","Type","Straight","Material","NPS","Tnom","service","AllValid","pred")]
```

```
sumTest6<-
```

```
Testing6[,c("PointID","Ntot","N","CR","RSQ","LineDesignation","DesignPressure","System","Type","Straight","Material","NPS","Tnom","service","AllValid","pred")]
```

```
sumTest7<-
```

```
Testing7[,c("PointID","Ntot","N","CR","RSQ","LineDesignation","DesignPressure","System","Type","Straight","Material","NPS","Tnom","service","AllValid","pred")]
```

```
sumTest8<-
```

```
Testing8[,c("PointID","Ntot","N","CR","RSQ","LineDesignation","DesignPressure","System","Type","Straight","Material","NPS","Tnom","service","AllValid","pred")]
```

```
sumTestAll<-rbind(sumTest3,sumTest4,sumTest5,sumTest6,sumTest7,sumTest8)
```

```
write.csv(sumTestAll,"SumTestAll.csv")
```

R Code for PoF calculation

R Code for corrosion classification

```
#install.packages("neuralnet")
```

```
require(neuralnet)
```

```
##neuralnet
```

```
setwd("C:/@develop/phd/R")
```

```
train <- read.csv("tcc.csv")
```

```
#dim(train)
```

```
##neuralnet
```

```
nn = neuralnet(y1+y2~i1+i2+i3+i4+i5+i6,data=train,hidden = c(6,6,6,6,6)) ##i2+i3+i4+i5+i6 +y2
```

```
nn = neuralnet(A+B+C+D+E+F+G+H+I+J+M+O+R+S+V+Y+AA~t1+t2+t3+t4+t5+t6+t7,data=train,hidden  
= c(6,6,6,6,6,6)) ##i2+i3+i4+i5+i6 +y2
```

```
plot(nn)
```

```
write.csv(nn$net.result, "result.csv")
```

```
write.csv(nn$weights[[1]][[1]][,], "weightsW1.csv") # , row.names = FALSE, check.rows = FALSE,  
check.names = FALSE)
```

```
write.csv(nn$weights[[1]][[2]][,], "weightsW2.csv")
```

```
write.csv(nn$weights[[1]][[3]][,], "weightsW3.csv")
```

```
write.csv(nn$weights[[1]][[4]][,], "weightsW4.csv")
```

```
write.csv(nn$weights[[1]][[5]][,], "weightsW5.csv")
```

```
write.csv(nn$weights[[1]][[6]][,], "weightsW6.csv")
```

```
write.csv(nn$weights[[1]][[7]][,], "weightsW7.csv")
```

```
plot(nn)
```

```

nn$weights

nn$net.result

nn$response

write.csv(nn$net.result, "result.csv")

write.csv(nn$data, "data.csv")

predNN <- function(i)
{
setwd("C:/@develop/phd/R")

#test is the data to be analysed
test<-data.frame(read.csv((paste("test" , i , ".csv", sep = ""))))

#unique list of pointids with i readings
ids <- data.frame(test[,1])

#NN is the neural network for i readings
NN<-readRDS(paste("NN", i , ".RDS", sep = ""))

#pred is the predicted cr type per point id
pred <- predict(NN,test [,-1])

#classify
pr <- apply(pred, 1, which.max)

pc<-c("a","b","c","d","e")

```

```
crtype <- data.frame(ids, pc[pr])

#output

write.csv(pred, paste("pred" , i , ".csv", sep = ""))

write.csv(test, "ids.csv")

}

#source("PredNN.r")
```