Article

Challenges in the application of DCVG-survey to predict coating defect size on pipelines

F. Anes-Arteche*, K. Yu, U. Bharadwaj, C. Lee, B. Wang

Francisco Anes-Arteche
TWI Ltd.
Granta Park, Cambridge, Cambridgeshire CB21 6AL
United Kingdom
francisco.arteche@twi.co.uk / franciscoantonio.anesarteche@brunel.ac.uk

Keming Yu
Brunel University London
Kingston Ln, Uxbridge, Middlesex UB8 3PH
United Kingdom
keming.yu@brunel.ac.uk

Ujjwal Bharadwaj
TWI Ltd.«QIQQA_BIBLIOGRAPHY»
Granta Park, Cambridge, Cambridgeshire CB21 6AL
United Kingdom
ujjwal.bharadwaj@twi.co.uk

Chi Lee
TWI Ltd.
Granta Park, Cambridge, Cambridgeshire CB21 6AL
United Kingdom
chi.lee@twi.co.uk

Bin Wang
Brunel University London
Kingston Ln, Uxbridge, Middlesex UB8 3PH
United Kingdom
bin.wang@brunel.ac.uk

Abstract

Corrosion is often the cause of pipeline failure potentially resulting in disasters causing damage and fatalities. To maintain the integrity of non-piggable lines, NACE's External Corrosion Direct Assessment (ECDA) methodology is commonly applied to assess external corrosion that can occur at coating defects on underground pipelines.

Work presented here is from a validation exercise carried out on the results of ECDA assessment using subsequent excavation data. The ECDA was carried out over 300 kilometres of crude oil pipelines with excavation carried out at 200 locations.

This paper models the relationships between pipeline coating defect area (area with coating breakdown) , corrosion depth, Direct Current Voltage Gradient (DCVG) measurements (in terms of %IR values), and factors capturing diverse environmental conditions through novel application of regression models.

This paper sheds light on the challenges in drawing conclusions in the assessment of corrosion from Direct Current Voltage Gradient (DCVG) inspection data and other types of data that form key inputs to ECDA. We expect that the analyses shown here using innovative regression models will support more reliable predictions of external corrosion in pipelines.

Keywords: ECDA, DCVG, Quantile Regression, Underground Pipeline, External Corrosion, Coating Damage

1 Introduction

All underground/underwater pipelines are subject to corrosion where the protective coating is damaged and there are inadequate levels of cathodic protection (CP). The issue is particularly significant in ageing pipelines. A commonly used approach for the assessment of external corrosion risk of buried, land pipelines is based on the NACE RP502 standard [1], often referred to External Corrosion Direct Assessment (ECDA).

Work reported in this paper builds on an integrity assessment carried out by TWI on pipelines. This paper presents the results from the application of this assessment that included ECDA. In this approach, which has been described in more detail in Section 2, initially, an assessment of the likelihood of external corrosion occurring on a pipeline was made from indirect measurements to prioritise further action. This formed the basis for a more comprehensive inspection that involved excavation at selected sites. Based on the correlations between actual observations (from excavations) regarding the condition of the pipeline and coatings at these locations with prior data (indirect measurements), this paper improves the understanding and interpretation of data used in ECDA in order to make more reliable predictions [1-3].

In existing approaches, the underlying assumption is that indirect measurements can provide data to reliably identify corrosion defects on the pipeline, so that excavation location can be prioritised. One established indirect method to determine the condition of the pipeline coating is to use an above-ground technique, such as DCVG, to locate and estimate the severity of the any coating defects that may be present on a pipeline [2]. Whilst the location aspect of this technique is very accurate and reliable, the severity, which is inferred from the %IR value, may not correlate very well with the actual size of the coating defect when examined after excavation [3]. Therefore, there is need to exercise caution in using %IR value to provide an indication of the severity (and/or size) of coating defects.

This paper is motivated from previous studies and the availability of both indirect and direct data from an industrial application at TWI, presented here as a case study. The analysis has confirmed existing beliefs and shown some insights that will benefit future assessment approaches. The sections that follow describe the case study, the type of data used, the analyses tools used, and the results obtained. This is followed by a discussion and conclusions where the authors attempt to elaborate on key results, the challenges involved and the way forward.

2 Case Study

TWI was commissioned to perform an integrity assessment for a number of underground crude oil pipelines operated by NISOC in Iran. These pipelines were not piggable and therefore the most appropriate solution was to apply the ECDA methodology. In order to identify coating damage, DCVG was performed along these pipelines.

2.1 Description of the data

For the purposes of this paper, data from Pre-Assessment, Indirect Inspection and Direct Examination (the first three steps of ECDA as specified by NACE) was gathered and analysed. The type of data gathered is described in the sub-sections below and the analyses carried out are shown in the sections that follow.

2.1.1 Pre-Assessment data

Preassessment data available included pipeline design specifications, operational data and time in service. This data was gathered from design and installation reports.

There were a total of 9 pipelines, covering 300 km, with diameters ranging from 26” to 42”. The material used was API 5L-X60 and X52. Operation pressure varied from 8 Bar to 17 Bar. The operation temperature ranged from 40 °C to 60 °C. The flow rate varied from 400 $m^3 h^{-1}$ to 1520 $m^3 h^{-1}$. The pipelines inspected in this project were commissioned between 1972 and 1992. The time in service was calculated from the time since commissioning of the line. The coating used to protect them was either cold wrap or coal tar.

2.1.2 Indirect Inspection data

During the Indirect Inspection phase of the integrity management project, DCVG was performed along the entire length of the pipelines using the pipeline CP system (impressed current) operating at its normal output. For each coating defect identified, the OL/RE (over-the-line to remote earth voltage) was measured. Then, as per NACE TM0109 [2], the P/RE (calculated pipe to remote earth potential at indication) was calculated using a linear interpolation between the pipe to remote earth voltage at the two closest Test Stations.

In this way, IR drop was calculated for each location at which coating damage had been detected with DCVG. Voltage drop (or %IR) is a relative value of the current waste through the coating defect and takes values from 0% to 100%.

2.1.3 Direct Examination data

In the third phase of the ECDA methodology, a series of measurements were taken and coating defects were examined after excavation, this was done not only at defect locations where DCVG data showed high severity, but also at some locations where the severity was not indicated as high. This was with a view to test the predictions using indirect against actual excavated (direct) data.

Two kinds of data were gathered: environmental and pipe-related data. Table 1 shows the nature of this data. It must be noted that for the regression analyses carried out, both quantitative data and categorical data (qualitative measurement or descriptive data) are used.

Soil resistivity has been taken using the four pins method detailed in NACE SP0502 [1] at 0.75, 1.5 and 3 m. The values used for the analyses are the closest measurements to the depth of the pipeline.

Backfill properties were not easy to quantify. They were classified as three different groups: sand, clay and clay with a mix of gravel/rock/stones due to the different properties. The geometry of the backfill refers to the shape that the soil particles have. They were divided into round and angular geometries: round where soil particles have rounded edges and angular where the soil particles have sharp edges.

For each excavation location the presence or absence of water was annotated; however, in some excavation reports, this data was missing. When water was present, the pH value was measured.

Coating defect areas were calculated from the excavation reports in which only length and width were annotated. However, photographs of each individual defect were taken. Using these, it was possible to estimate the true defect area. Only sections of the pipeline without coating were considered; disbonded defects were not considered in the analysis. The corrosion depth measurements were carried out using Ultrasonic testing equipment.

The amount of deposits under coating was also measured. It takes values from 0 to 100% and it was calculated by dividing the coating defect area with deposits by the total coating defect area.

2.2 Correlation between DCVG (%IR) and corrosion depth

During the Indirect Inspection phase of the ECDA methodology, Direct Current Voltage Gradient (DCVG) survey was carried out and during the Direct Assessment metal corrosion depth was measured. Linear regression has been applied to the data hence obtained in order to observe the existence or absence of correlation between the data. The techniques used for regression and the analyses are described in the sections that follow.

2.3 Correlation between DCVG (%IR) and coating defect area

Relationship between %IR and coating defect area is important. BS ISO 15589 states “DCVG surveys can be used to locate and establish the relative size of defects in protective coatings on buried pipelines” [5] so there is an acceptance by the pipeline industry that %IR provides an estimate of the size of coating defects. By “relative” it is understood that, voltage drop is proportional to the coating defect area, however this has not been quantified.

With a good understanding of how the voltage drop measured during DCVG is related with the coating defect area, it is possible to improve the accuracy in the prediction of coating defect area and consequently identify the high risk areas. This relationship has been investigated using the data described in section 2.1; the analyses are described in section 3.

2.3.1 Linear regression model

Linear regression is first used to model the relationship between %IR and coating defect area, with the aim of finding the relation between %IR and coating defect area. The advantages of applying linear regression are its simplicity and interpretability. Also, it provides a good initial understanding of the behaviour between these two parameters.

A linear regression model employs the least squares estimator to fit a single explanatory variable x to the dependent variabley. The target is to find the equation (Eq. 1) of the straight line that would give the best fit for the data points.

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (1)$$

In our case, y is coating defect area in cm^2, x is %IR, β_1 is the slope or regression coefficient, β_0 is the intercept and ε is the model error.

The simple linear model is unsuitable for multiple independent variables requiring a more complex regression method such as multiple linear regression.

2.3.2 Multiple linear regression model

Multiple linear regression (MLR) can be used to fit a predictive model to an observed data set in order to quantify the strength of the relationship between the dependent variable and the independent variables. The assumptions considered in order to apply MLR are:

- Variables have weak exogeneity, meaning they are free of error.
- Linearity.
- There is no correlation between the predictor variables.

In order to model the coating damage area, multiple linear regression (MLR) has been implemented using in R software taking in consideration variables described in Table 1 by using Eq. 2.

$$y = \propto + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (2)$$

Some of these factors being categorical (qualitative), dummy variables are used in order for them to be used in the regression model. Table 2 illustrates the factors included in the regression model with dummy variables.

When using dummy variables, at least one category needs to be omitted, which becomes the reference category against which the effects of other categories are assessed.

Nine independent variables were introduced in the mathematical model including %IR. The eight new variables, representing other factors (shown in Table 2) not considered in the simple regression model, were introduced to obtain their individual influence on the coating defect area.

MLR provides better predictive capability than simple linear regression, and provides an estimate of the relative importance of each variable. However, MLR is very sensible to outliers. Thus, to understand better the relationship between variables a more robust mathematical model, described below, is considered.

2.3.3 Quantile regression model

The probability density function of the coating defect area is not symmetric; it has a positive skewness (Fig. 1), leading us to use more complex models such as quantile regression, because we aim to assess how a factor or factors could cause larger or smaller coating defects. In such situations, mean-based regression models such as described above are not effective in finding solutions. Instead, quantile regression is used to identify the effect of key factors (%IR) for large and small coating defects.

Quantiles describe the distribution of the dependent variable in terms of quantile [6]. While the median is a special quantile to measure the middle location, extreme quantiles describe the tails of the distribution. Multiple Linear Regression models the relationship between one or more independent variables and the conditional mean of a dependent variable, whilst, quantile regression models the relationship between the independent variables and the conditional quantiles of the dependent variable rather than the conditional mean of aforementioned variable [7-8].

From basics statistics it is known that any real valued random variable, Y, is characterized by its distribution function,

$$F(y) = Prob(Y \leq y) \tag{3}$$

The τ-quantile of $F(y)$ is usually defined as the inverse of $F(y)$, i.e. . $F^{-1}(\tau)$.

Correspondingly, the τ-empirical quantile of the sample $\{Y_1, \dots, Y_N\}$ can be computed by the following minimization problem [9]:

$$\hat{q}_\tau = argmin_z \sum_{i=1}^{n} [\tau \cdot I(Y_i > z) + (1 - \tau) \cdot I(Y_i < z)] \cdot |Y_i - z| \tag{4}$$

Where I is an indicator function. While a conditional distribution of Y given independent variables X_s is concerned and replaced, quantile regression is defined correspondingly. The parameter z is the value which minimizes the function.

Quantile regression has been computed using R software. The variables included have been the same as described in 2.1.3. This quantile regression gives a more comprehensive view of the effect of the independent variables on the dependent variable (coating defect area) and will help to determine the factor combination affecting to high or low coating defects.

With quantile regression, we can study the effects of the %IR drop on coating defect area for low and high %IR drop. Although it is possible to do this with a normal distribution, it will not differentiate between those locations with low %IR drop and those with high %IR drop.

The advantages of applying quantile regression are the flexibility for modelling data with heterogeneous conditional distributions and more robustness relative to the use of linear regression. Also, quantile regression has richer characterization and the description of the data can show different effects of the independent variables on the dependent variable across the spectrum of the independent variable.

2.4 Correlation between coating defect area and corrosion depth

When performing ECDA, the %IR value plays an important role in determining the severity classification of an indication. The pipeline operator defines and applies criteria for classifying the severity of each indication. Small indications (%IR) are classified as minor severity, while large indications are classified as severe (Table 3 of NACE SP0502 [4]) . The severity indication allows an estimation of the extent of the coating defect area. When a section of a pipeline is exposed after coating breakdown, corrosion activity might occur with certain areas having high corrosion depths that compromise the integrity of the pipeline. It is not a straight forward process to predict the corrosion depth using indirect inspections. For this case study, 200 locations were examined directly and corrosion depth was measured as detailed in Section 2.1.3. Results are discussed in Section 3.3.

3 Analyses of data

Regression models have been performed and results have been reported in the following sub-chapters. These models include linear (simple and multiple) and quantile regressions. Relationships between DCVG measurements (%IR), corrosion depth and coating defect size have been addressed.

3.1 Correlation between DCVG (%IR) and corrosion depth

There is not a straightforward relation between corrosion depth (wall thickness lost) and the voltage drop (%IR) calculated during DCVG (Fig. 2). A total of 43% of the points correspond to regions where there was no corrosion activity and therefore the corrosion depth was zero.

3.2 Correlation between DCVG (%IR) and coating defect area

3.2.1 Linear regression model

The linear regression attempts to illustrate the correlation between voltage drop (%IR) and coating defect area (Fig. 3). Although a strong trend has not been observed, increase of the coating defect area (exposed pipe area) generally correlated to an increase in the %IR drop, especially at values above 30%.

Only defects where the exposed pipeline is in direct contact with the soil (bare sections of the pipe) have been t aken into consideration. Regions with disbonded coating have been omitted due the poor correlation with the voltage drop caused by CP shielding.

3.2.2 Findings and possible reasons for poor linear correlation between %IR and coating defect area.

Analysing the outliers of the graph (points far from the trend) by using individual inspection reports corresponding to such data, it has been found that the DCVG readings have been potentially influenced by features such as:

- Sections of the pipeline with scales correspond with lower voltage drops. Scale deposits in the pipeline surface effectively isolate the pipeline electrically reducing the measured coating defect area. This is a problem because DCVG will give a smaller measure for sections of the pipeline with a severe coating damage, thus invalidating the damage prediction.
- Reliability of the DCVG reading may be compromised in locations where old cable connections (cad welds) are present.
- Accurate measurements are not always possible at crossings with roads and watercourses due to local changes in the soil/ground conditions.
- The presence of nearby underground pipelines, in particular those with coating defects, reduce the accuracy of DCVG. The voltage signal is often interfered by the cathodic protection system of the nearby pipelines.
- The soil resistivity affects the %IR value for non-homogeneous soils along the pipeline. When performing DCVG, the pipe to remote earth potential at the indication (P/RE) is calculated by using a linear function of the voltage between the two nearest Test Stations

[2]. It is therefore assumed to be an homogeneous soil between the two Test Stations. However, soils are heterogeneous and the soil resistivity affects the %IR value. For the same coating defect area, %IR can be higher when soil resistivity has high values, whereas %IR can be lower when the local soil resistivity has low values. This is consistent with other work showing that a high resistivity could cause a small defect to yield a large %IR, for example as in [10]. The electrochemical process of cathodic protection causes the environment around the pipeline to become alkaline, in particular at the surface of the defect being protected [11]. The increase in the pH value can result in a decrease in soil resistivity near the defect and therefore increases the heterogeneity of the soil.

- The influence of high AC-high voltage lines nearby buried pipelines releasing stray currents to the ground: stray currents have important influence on long distance pipelines, in particular for those running in parallel or across high voltage AC lines. During DCVG measurements, currents from AC-high voltage lines affect the voltage gradient from defect indication epicenter to remote earth (OL/RE) [2]. This effect could lead to inaccurate DCVG measurements with the level of influence depending on the intensity and direction of the AC current released. The presence of stray currents makes a DCVG survey difficult to interpret as there may be AC current flowing on or off the pipeline.
- Orientation of the coating defect indication: when a coating anomaly is located in the bottom part of the pipeline the voltage signal is attenuated. This factor has not been included in the analysis due to limited availability of such data; it is an area that could benefit from future research.
- If there is physical contact between pipeline and metallic support of an aboveground pipeline, the voltage gradient measurements are affected.
- Depth of cover affects DCVG signal. DCVG indications decrease as depth of cover increases [10][12].

Some errors in the data could be attributed to excavation and Direct Examinations:

- During Direct Examination, coating defect area is measured. In many cases, the coating is just disbonded and during the excavation activity it breaks off. Therefore the measured defect area is higher than the area in contact with the soil before excavating. This is very common, in special for the bottom part of pipelines with coal tar coatings.

This study has identified some factors that can potentially cause poor linear correlation between DCVG data and defect area. There could be other factors at play, and, indeed, the factors could be different if the same pipeline system is in a different operating environment.

3.2.3 Multiple linear regression prediction model

As mentioned in Section 2.3.2, Multiple Linear Regression (MLR) has been performed in order to estimate the average of coating defect area. The following expression predicts the area of coating defect for the nine given parameters addressed in Section 2.

$$Average\ Coating\ Defect\ Area\ (cm^2) = 4.60 \cdot 10^4 - 1.18 \cdot 10^3\alpha + 4.68 \cdot 10^3 A1 - 2.58 \cdot 10^3 A2 + 1.91 \cdot 10^{-1}\beta + 3.35 \cdot 10^2\gamma + 4.42 \cdot 10^3 B1 + 9.16 \cdot 10^3 C1 - 1.35 \cdot 10^{-3} C2 - 7.58 \cdot 10^3 D1 - 1.52 \cdot 10^3 E1 + 1.56 \cdot 10^2\delta \quad (5)$$

Where:

α Time in service

β Soil resistivity at site location

γ %IR drop

δ Amount of deposits under coating (%)

$A1$ If 1, pH>7.5

$A2$ If 1, pH<7.5

$B1$ If 1, there is presence water at site location

$C1$ If 1, sand backfill at site location

$C2$ If 1, clay backfill at site location

$D1$ If 1, backfill have round geometry

$E1$ If 1, coating is cold wrap

Each of the coefficients in Eq. 5 has a standard deviation associated (Table 3). Dividing the estimate by the standard deviation (Std. Error) the “t value” is obtained. If this value is outside the range (-2,2), then, we have significantly different results from zero meaning that the statistic is reliable and therefore this factor has a strong correlation with the dependent variable. Other way of interpreting these results is by considering the “p value”. If it is less than 0.05, this also means that we have significantly different results from zero and we consider this variable as statistically significant.

Eq. 5 and Table 3 indicate the proportionality between voltage drop (%IR) and coating defect area. An increase of one unit in %IR, the predicted coating area increases 335 cm^2. Voltage drop is limited to 100%, therefore the maximum coating defect for the predictive model is 33,500 cm^2.

Based on the assumption that $p = 0.05$, some of the factors have been addressed as having “near significant differences”. The independent variable $\propto$ and the intercept are not significant for $= 0.05$, but become significant if the p value is increased to 0.1. This means that $\propto$ and the intercept have some sort of correlation with the dependent variable, however not as strong as variables with $p < 0.05$.

Also, from this regression model we can determine that only 21.53% of the variation is explained by the regression and the rest is due to error.

3.2.4 Correlation of %IR and coating defect area: application of Quantile regression.

The calculated quantile regression estimates multiple rates of change (slopes) from the minimum to the maximum response, providing a more complete picture of the relationships between variables, which is an improvement over the regression model shown earlier. Equations for the most significant quantiles are presented in Eq. 6 to 8.

$$Coating\ defect\ area(0.5\ quantile) = 4.77 \cdot 10^4 - 1.19 \cdot 10^3\alpha + 3.00 \cdot 10^{-2}\beta + 1.08 \cdot 10^2\gamma + 7.89 \cdot 10^1\delta + 1.49 \cdot 10^3 A1 - 1.39 \cdot 10^3 A2 + 4.42 \cdot 10^3 B1 + 1.26 \cdot 10^4 C1 - 5.43 \cdot 10^3 C2 - 4.31 \cdot 10^3 D1 - 1.70 \cdot 10^2 E1 \quad (6)$$

$$Coating\ defect\ area(0.05\ quantile) = 2.09 \cdot 10^3 - 5.26 \cdot 10^1\alpha + 7.41 \cdot 10^{-3}\beta + 4.10 \cdot 10^0\gamma + 1.11 \cdot 10^{-1}\delta - 3.73 \cdot 10^1 A1 + 2.71 \cdot 10^3 A2 + 1.52 \cdot 10^2 B1 + 2.72 \cdot 10^3 C1 - 1.57 \cdot 10^2 C2 - 2.10 \cdot 10^2 D1 - 9.36 \cdot 10^2 E1 \quad (7)$$

$$Coating\ defect\ area(0.95\ quantile) = 4.96 \cdot 10^4 - 1.23 \cdot 10^3\alpha + 6.69 \cdot 10^{-1}\beta + 8.61 \cdot 10^2\gamma + 3.35 \cdot 10^2\delta - 2.07 \cdot 10^4 A1 - 2.71 \cdot 10^4 A2 + 1.94 \cdot 10^4 B1 + 2.45 \cdot 10^3 C1 - 1.85 \cdot 10^4 C2 - 3.29 \cdot 10^3 D1 - 1.32 \cdot 10^2 E1 \quad (8)$$

In Fig. 4, the quantiles of the dependent variable are on the horizontal axis and the coefficient magnitudes on the vertical axis. The MLR coefficients are plotted as a horizontal line with the confidence intervals as two horizontal lines around the coefficient line. The MLR coefficients do not vary by quantiles.

The quantile regression coefficients are plotted as lines varying across the quantiles with confidence intervals above and below them. If the quantile coefficient is outside the MLR confidence interval, then we have significant differences between the quantile and MLR coefficients.

The quantile coefficients for the %IR drop (independent variable) on coating defect size (dependent variable) are significantly different from the MLR coefficients. Moreover, the effect of the %IR drop increases for locations with higher coating defect size (higher quantiles).

For the 5th quantile, which represents the small coating defects, an increase of one unit in the %IR value, increases 4.10 cm^2 the coating defect area. Whereas for the 95th quantile, which represents the large coating defects, an increase of one unit in the %IR value, increases

861 cm^2 the coating defect area (around 200 times more than for the 5^{th} quantile). The MLR coefficients cross with the quantile coefficients at the 75^{th} quantile.

The analyses show that DCVG readings are more sensitive to large coating defect areas than small to medium coating defect areas.

3.3 Correlation between coating defect area and corrosion depth.

There is no straightforward relation between corrosion depth and voltage drop (%IR). This is illustrated in Fig.2, which shows the correlation between voltage drop (%IR) and corrosion depth for the case study considered. A significant portion of the points correspond to regions where there was no corrosion activity and therefore the corrosion depth is zero.

Fig. 2 shows that %IR is not a good parameter in order to quantify corrosion depth. External corrosion is strongly dependent on environmental factors and therefore they should be taken in consideration. That is the reason why it is not possible to rely on %IR as a factor to determine the extent of corrosion damage.

Fig. 5 models the relationship between coating defect area and corrosion depth (peak depth). The cumulative corrosion feature red line shows the corrosion feature count starting from the biggest and progressing to the smallest in terms of calculated area. As studied before by Argent [13] and demonstrated here, the changing slope of this line shows that most of the coating defects are relatively small in area, and this number is decreasing with the increment of exposed area.

Regression models have been applied in order to determine the strength of correlation for the cited variables including zero inflated models (binomial and Poisson) in order to draw conclusions. However, the regressors obtained with zero inflated models didn’t show any kind of direct relation between the coating defect area and the corrosion depth.

4 Discussion and the way forward

4.1 The use of DCVG data in ECDA to predict corrosion depth

DCVG data is used in ECDA to identify the pipeline locations to be excavated for Direct Examination; this is consistent with studies that have shown DCVG data to be reliable in knowing the location of coating breakdown. However, the correlation between %IR and corrosion depth is not strong and this case study confirms this aspect (Section 3.1). Corrosion depth is dependent on environmental factors and cathodic protection performance. This can be confirmed in further research in which a more comprehensive dataset that includes CP levels is analysed in a multiple regression model, thus improving corrosion depth prediction.

4.2 The use of DCVG data in ECDA to predict coating defect area

A substantial improvement in the reliability of prediction can be made by considering not just DCVG data, but also other such as those relating to environment some of which are shown in Table 1. Factors such as prior corrosion and repair history that have been included in other studies may help make better predictions[14-15]. In the case study shown here multiple factors (to the extent possible, given the data available) have been taken into account in the regression analyses resulting in more reliable prediction (Section 3.2.4).This it is supported by Masilela and Pereira [16] whose study states that DCVG enables comparison of located defects with other defects found in the same area. The %IR is used to reflect size/importance of a defect.

When ECDA is performed, at first instance, pipeline operators usually rely on DCVG values in order to provide an initial assessment of the line. This is a good practice to detect coating anomalies, typically used for new pipelines where the coatings more likely to be damaged during pipe construction [16].

4.3 Correlation between coating defect area and corrosion depth

For corrosion to be present, two conditions must be active, a damaged coating and inadequate levels of cathodic protection [17]. However, pipeline corrosion depth cannot be predicted by the only use of coating defect area data.

Corrosion might appear in small coating areas (Fig. 5). Deep pits materialize in small coating defect areas, meaning in locations that, following the severity classification given by NACE SP0502 [4], should be considered as minor severity. The relative size of the anode and cathode areas could be a critical factor in determining the amount of corrosion damage at these locations.

For a given potential difference, if the anode area is large compared with the cathode area, the anode current density will be low and the corrosion is widely distributed, resulting in a more general corrosion loss in the absence of any interference effects. Whereas, if the anode area is small (high anode current density) with respect to the cathode area, the corrosive action is localized and severe local damage may result [19].

Anodic interference from stray currents such as grounded electric power sources, equipments or electric railways, causes corrosion. This type of corrosion is a combined effect of a relatively large potential difference or current plus the fact that the anode area, where the current leaves the pipe, is small.

AC corrosion, due to its own characteristics, usually happens at small/very small coating faults [19]. When a defect is small, the AC required to induce pitting corrosion is low. However, current density decreases because of the blocking effect of corrosion product accumulating at the defect [20].

4.4 Potential causes of anomalies in DCVG readings

DCVG is a good estimator for locating coating defects; however DCVG readings are potentially affected by factors discussed in Section 3.2.2.

By analysing the outliers of the linear regression model and supported by literature review, it is felt that the presence of the following events will affect the performance of this technique:

- Surface scales.
- Presence of connections to old sacrificial anode protection systems (cad welds).

- Presence of nearby underground pipelines.
- Presence of high voltage AC lines.
- Physical contact between pipeline and metallic support of an aboveground pipeline.

Corrosion activity is hard to model using indirect inspection techniques. It is subjected to uncertainty given the underlying factors which are especially difficult to model with the available data.

Some studies [21] assume homogeneous soil resistivity for the DCVG survey interpretation, nonetheless in this project it is used the on-site soil resistivity for each of the defect locations, resulting in better results as proposed by McKinney et al. [22] whose research dictates “soil resistivity plays a larger role in determining DCVG signals than coating flaw size”.

4.5 Challenges and ongoing research

As discussed above, to rely only on DCVG data to assess damage (from both, coating breakdown and a reduction in the thickness of the pipeline as measured in depth of corrosion) is potentially misleading. A multiple regression such as shown in this paper that takes account of environmental and other factors is more accurate in predicting coating defect area. However, it requires specific data to be available. To be able to make more accurate predictions, updating techniques are being used so that new information can be used in the analyses as and when it is available. Also, there are techniques that enable the combination of data from different sources using Bayesian methods [23].

The regression techniques for prediction of corrosion damage must be viewed as complementary to other techniques such as Bayesian Belief Networks [24-26]. A pipeline integrity management approach may have inputs from elements of an ECDA approach, physics based corrosion models, structural reliability models such as in [27], and risk based decision support models that include the impact of consequential failure such as shown in [28].

The choice of approach and the techniques used often depend on the sort of data that is available. There is a strong case for sharing of corrosion data among stakeholders and the use of data mining techniques to analyse such data for common benefit [29]. Getting data from a wider sample may be particularly useful when situation/ location specific data is not easily available; such data could then be calibrated with specific inspection data when it becomes available.

5 Concluding remarks

This work confirms that DCVG is an inspection technique used to locate coating anomalies with accuracy. It gives a reasonable idea of the coating defect size but it is not possible to predict it with high confidence. The voltage drop (%IR) is an indicator of coating defect size, but it is not correlated with corrosion depth and therefore it may be misleading to use it in an integrity assessment to identify and screen out potentially less susceptible to failure sections in a pipeline. Likewise, the approach does not necessarily help in identifying the more susceptible-to-failure sections.

Corrosion depth is not necessarily proportional to coating defect size, and therefore, small indications detected during pipeline survey need to be treated with caution. It is required to consider both small and large coating anomalies for cathodically unprotected pipelines, otherwise high corrosion rates might occur in small defects in the presence of adverse environmental conditions. On the other hand, large coating defects are easy to detect as demonstrated with the application of the quantile regression model. High values of %IR will be generally linked to large coating defects.

We expect that the analyses shown here using novel regression models will support more reliable predictions of external corrosion in pipelines. The application presented here is in the petroleum sector but the results will be of interest in other sectors such as water distribution pipelines.

Acknowledgements

The authors are grateful to NISOC for allowing us to use the data in the case study. We also acknowledge the funding received through NSIRC (National Structural Integrity Research Centre) enabling this research to be carried out by TWI Ltd and Brunel University London.

6 References

[1] N. S. SP0502, *NACE International*, Houston, Texas, USA, **2010**.

[2] N. S. R. TM0109, *NACE International*, Houston, Texas, USA, **2009**.

[3] F. Varela, M. Y. Tan, M. Forsyth, *Corros. Eng., Sci. Technol*. **2015**, *50*, pp. 226-235.

[4] N. S. SP0502, *NACE International*, Houston, Texas, USA, **2002**.

[5] BS ISO 15589-1, *BSI Standards Publication*, **2015**.

[6] K. Yu, Z. Lu, J. Stander, *Journal of the Royal Statistical Society: Series D (The Statistician*, **2003**, *52, Issue 3*, pp. 331-350.

[7] R. Koenker, K. F. Hallock, *Journal of Economic Perspectives*, **2001**, *15, Issue 4*, pp. 143-156.

[8] C. Chen, presented at *Proceedings of the Thirtieth Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc, USA, **2005**.

[9] R. Koenker, Quantile Regression in R: A Vignette, **2012**.

[10] J. P. McKinney, *Master Thesis*, University of Florida, USA, **2006**.

[11] R. Norsworthy and others, presented at *17th International Corrosion Congress*, Las Vegas, Nevada, **2008**, Paper No. 4017.

[12] C. Ukiwe, S. McDonnell, presented at *NACE Corrosion 2011*, Houston, Texas, USA, **2011**, Paper No. 11132.

[13] C. J. Argent, C. Frolish, J. Healy, *BHR Group Pipeline Integrity*, MACAW Engineering Ltd., UK, **2008**.

[14] L. Welsh, presented at *NACE Corrosion 2011*, Houston, Texas, USA, **2011**, Paper No. 11129.

[15] A. Francis, M. McCallum, T. Illson, M. McQueen, presented at *NACE Corrosion 2004*, Houston, Texas, USA, **2004**, Paper No. 04196.

[16] Z. Masilela, J. Pereira, *Eng. Failure Anal*. **1998**, *5, Issue 2,* pp. 99-104.

[17] J. L. Mora-Mendoza, L. C. Saucedo-Robles, H. Rodriguez-Clemente, M. A. Gonzalez-Nunez, G. Zavala-Olivares, M. J. Hernandez-Gayosso, *Mater. Corros*. **2011**, *62, Issue 8*, pp. 796-801.

[18] M. Romanoff, *Underground Corrosion*, US Government Printing Office, **1957**.

[19] L. Di Biase, R. Cigna, O. Fumei, presented at *NACE Corrosion 2010*, Houston, Texas, USA, **2010**, Paper No. 10108.

[20] D. Kuang and Y. Cheng, *Corros*. **2014**, *71*, *Issue 3*, pp. 267-276.

[21] H. Castaneda, M. Urquidi-Macdonald, *Corros.* **2004**, *60, Issue 6*, pp. 538-547.

[22] J. P. McKinney, M. E. Orazem, O. C. Moghissi, presented at *NACE Corrosion 2006*, Houston, Texas, USA, **2006**, Paper No. 06188.

[23] ASME International, *Risk Based Methods for Equipment Life Management*, CRTD, Obataining and combining data, New York, USA, **2003**, 41, pp.115-46

[24] N. Sridhar, F. Ayello, F. Candreva, *Managing Aging Plants*, **2015**, pp. 47-51.

[25] S. Jain, F. Ayello, J. A. Beavers, N. Sridhar, presented at *NACE Corrosion 2014*, Houston, Texas, USA, **2014**, Paper No. 3851.

[26] F. Ayello, N. Sridhar, G. Koch, V. Khare, presented at *NACE Corrosion 2014*, San Antonio, Texas, USA, **2013**, Paper No. 2616.

[27] M. Lamb, T. Zimmerman, presented at *The Onshore Pipeline Cost Reduction Conference*, Amsterdam, Netherlands, 3-5 April, **2000**.

[28] U. R. Bharadwaj, V. V. Silberschmidt, J. B. Wintle, *Journal of Quality in Maintenance Engineering*, **2012**, *18, Issue 4*, pp. 417-431.

[29] X. Li, D. Zhang, Z. Liu, C. Du, C. Dong, *Nat. Mater*. **2015**, *527*, pp. 441-442

Received: ((will be filled in by the editorial staff))

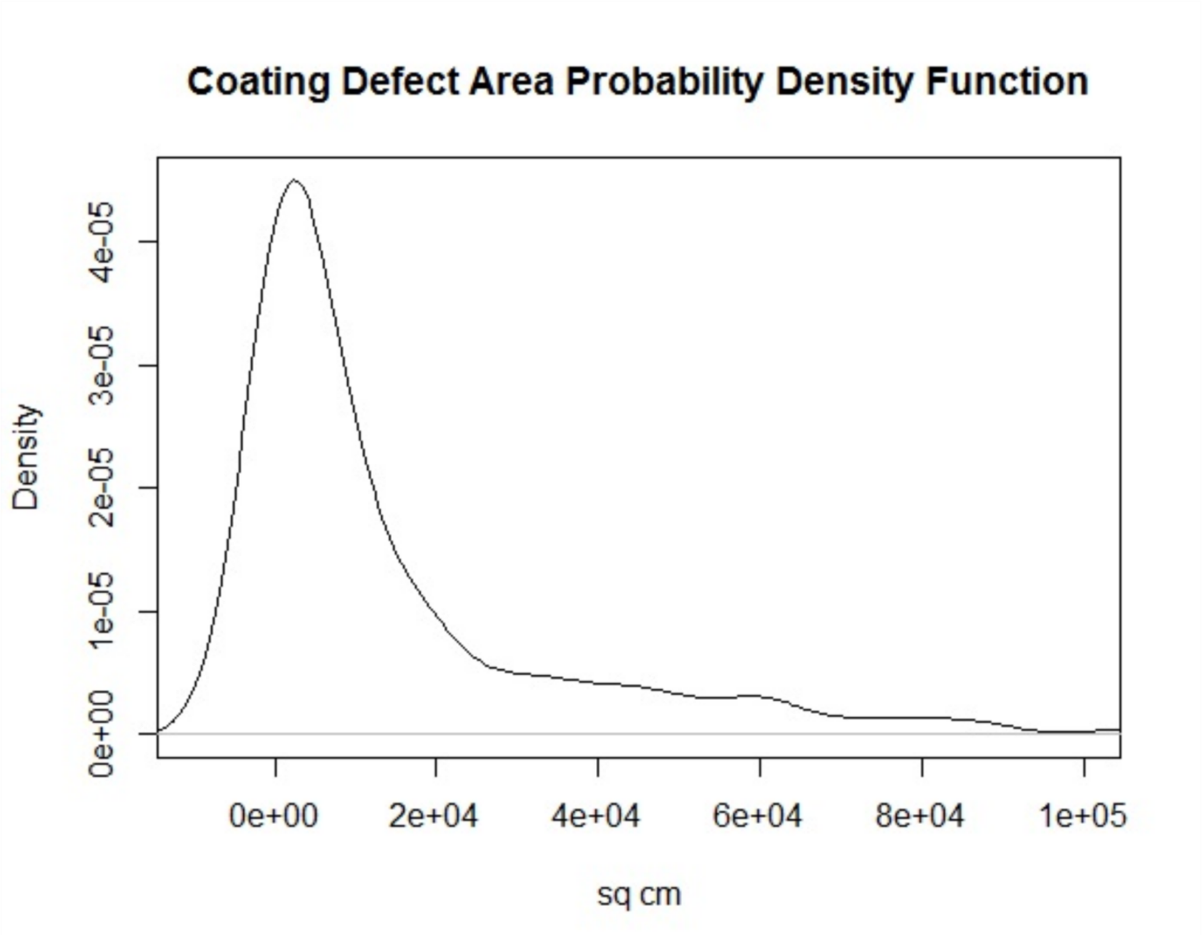

Figure 1. Probability density function for coating defect area variable. 'Reproduced Courtesy of TWI Ltd'.

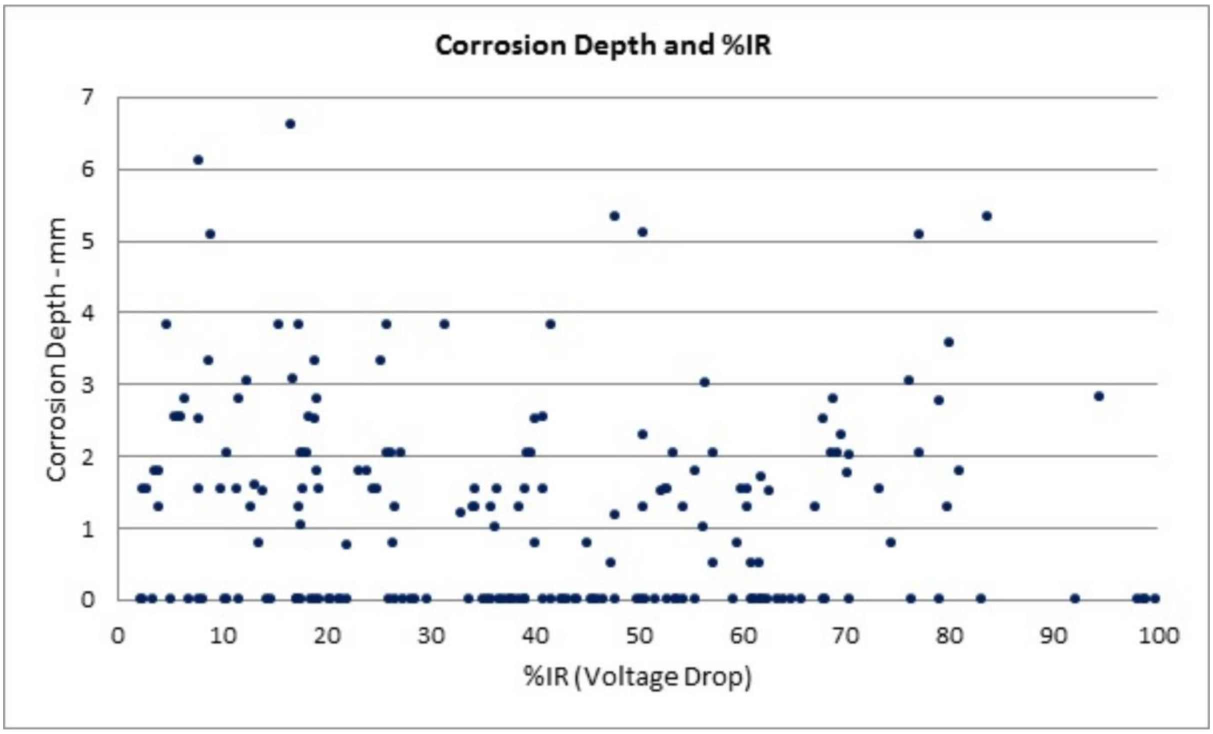

Figure 2. Corrosion depth and %IR. 'Reproduced Courtesy of TWI Ltd'.

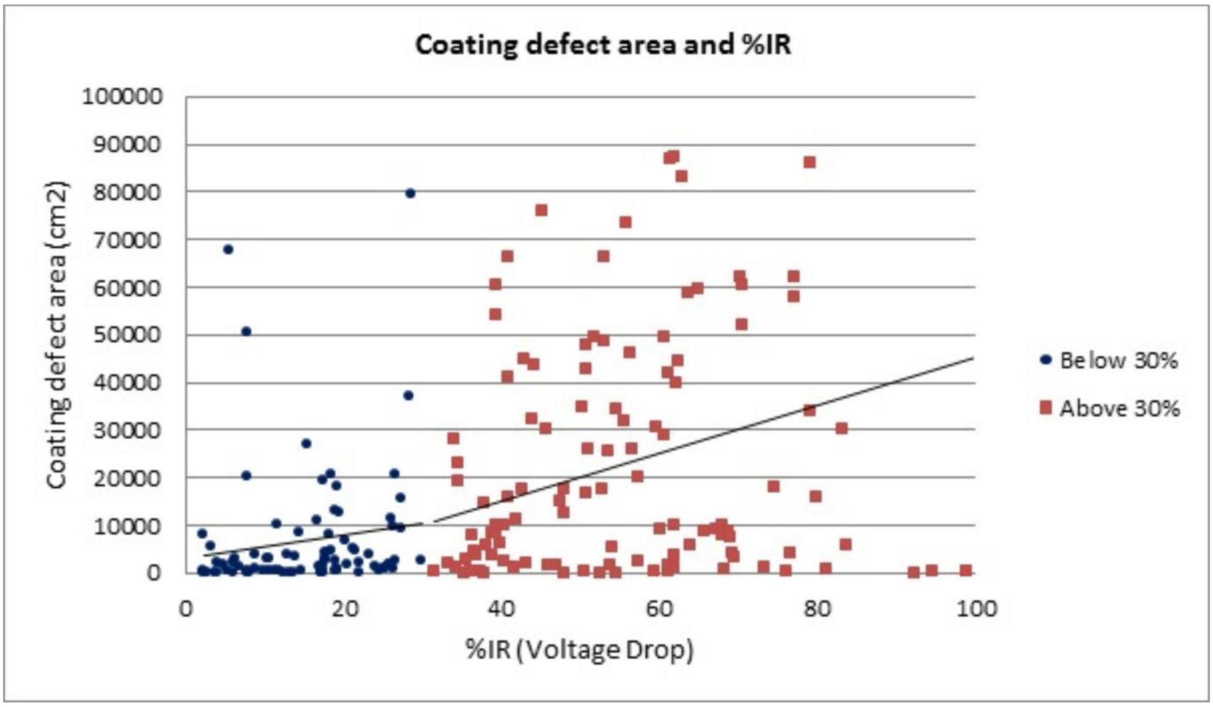

Figure 3. Coating defect area and %IR. 'Reproduced Courtesy of TWI Ltd'.

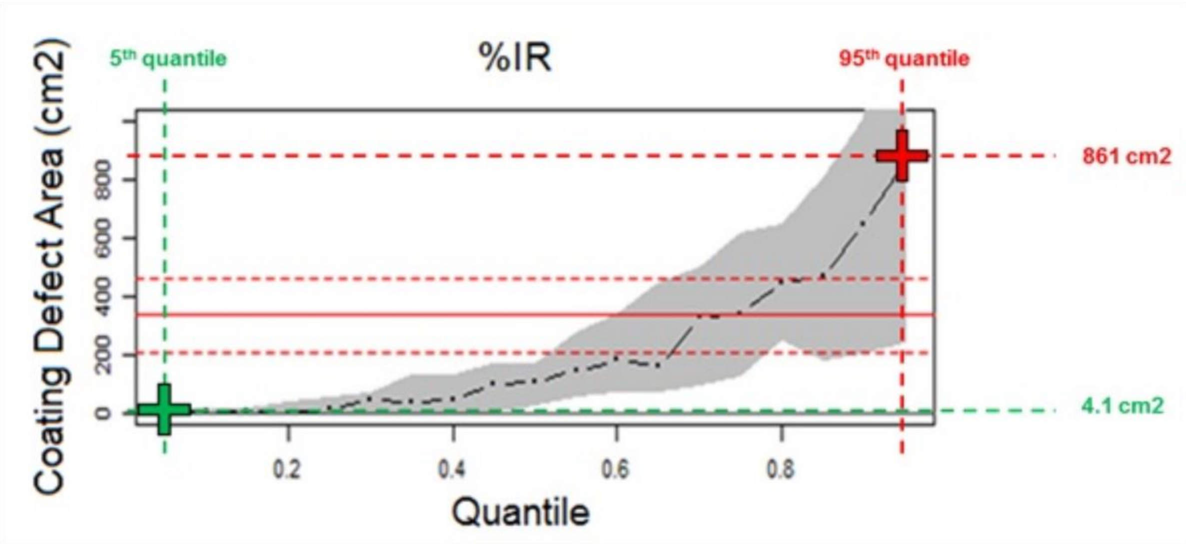

Figure 4. Representation of the quantile coefficients for the %IR drop on coating defect size. 'Reproduced Courtesy of TWI Ltd'.

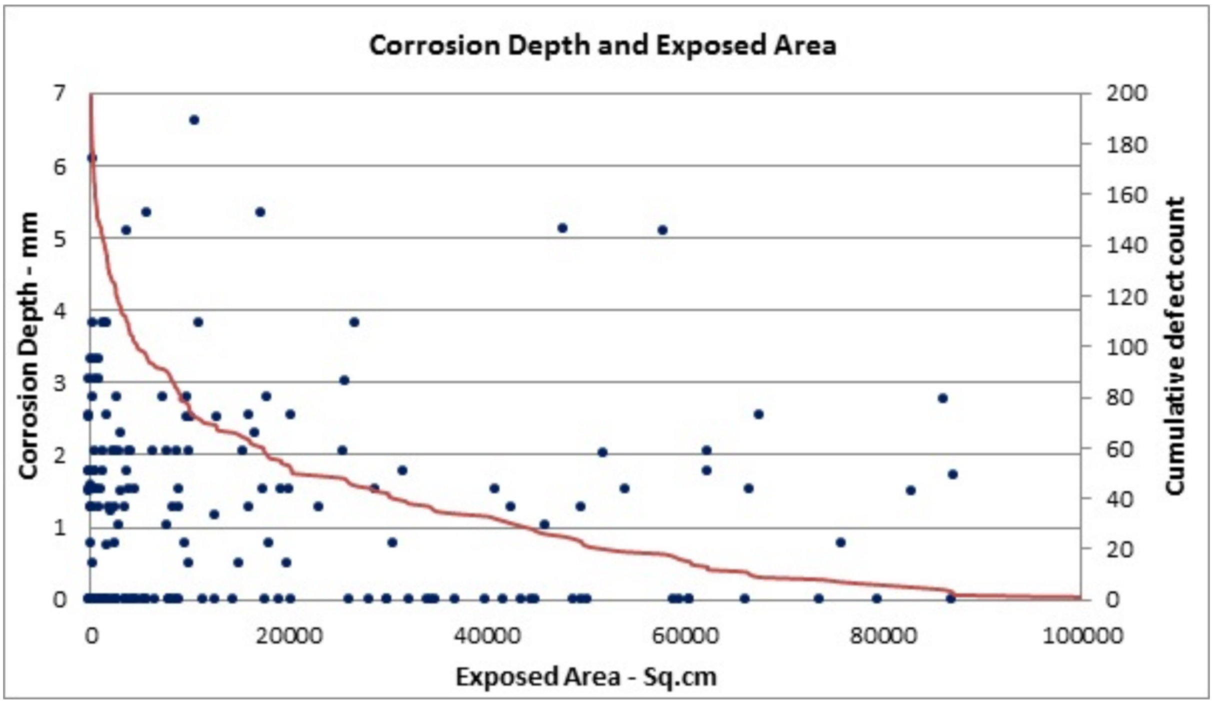

Figure 5. Relationship between coating defect area and corrosion depth. 'Reproduced Courtesy of TWI Ltd'.

Factor		Type of data	Range of values
Soil resistivity	*(Environmental)*	Quantitative	Values from 75 to 43,332 Ωcm
Backfill type	*(Environmental)*	Categorical	See Table 2
Backfill geometry	*(Environmental)*	Categorical	See Table 2
Presence of water	*(Environmental)*	Categorical	See Table 2
pH of water	*(Environmental)*	Quantitative	Values from 6 to 14 when applicable
Coating damage area	*(Pipe related)*	Quantitative	Values from 0 to 21,550 cm^2
Corrosion depth	*(Pipe related)*	Quantitative	Values from 0 to 6.6mm
Deposits under coating	*(Pipe related)*	Quantitative	Values from 0 to 100%

Table 1. Types of data used in regression analyses. 'Reproduced Courtesy of TWI Ltd'.

Factors with Dummy variables	States
pH of water from excavation	>7.5
	<7.5
	Unknown
Presence of water	Yes
	No
Type of backfill	Sand
	Clay
	Clay with mix of gravel/rock/stone
Geometry of backfill	Round
	Angular
Type of coating	Cold wrap
	Coal Tar

Table 2. Factors with dummy variables and their states. 'Reproduced Courtesy of TWI Ltd'.

Variable	Estimate	Std. Error	t value	p value	Interpretation
Intercept	$4.60 \cdot 10^{4}$	2.78 E+04	1.651	0.10071	Near significant differences
α	$-1.18 \cdot 10^{3}$	6.67 E+02	-1.769	0.07876	Near significant differences
β	$1.91 \cdot 10^{-1}$	3.83 E-01	0.500	0.61767	No significant differences
γ	$3.35 \cdot 10^{2}$	7.86 E+01	4.263	$3.4 \cdot 10^{-5}$	Significant Differences
δ	$1.56 \cdot 10^{2}$	6.85 E+01	2.280	0.02392	Significant Differences
A1	$4.68 \cdot 10^{3}$	1.03 E+04	0.455	0.64993	No significant differences
A2	$-2.58 \cdot 10^{3}$	9.56 E+03	-0.270	0.78735	No significant differences
B1	$4.42 \cdot 10^{3}$	6.09 E+03	0.726	0.46873	No significant differences
C1	$9.16 \cdot 10^{3}$	1.13 E+04	0.811	0.41883	No significant differences
C2	$-1.35 \cdot 10^{-3}$	5.06 E+02	-2.671	0.00833	Significant Differences
D1	$-7.58 \cdot 10^{3}$	8.39 E+03	-0.904	0.36740	No significant differences
E1	$-1.52 \cdot 10^{3}$	1.14 E+04	-0.134	0.89386	No significant differences

Table 3. Multiple Linear Regression Coeficients. 'Reproduced Courtesy of TWI Ltd'.

Graphical Abstract

The overarching aim of this study is to improve the understanding and interpretation of the data typically used in ECDA, thereby complementing expert judgement or supporting experts to make reliable prediction regarding external corrosion. Particularly interesting is the use of the advanced Quantile Regression model to study the effects of the %IR values on the coating defect area for low and high %IR values.