Journal: MACO | MSP No.: 201810085 | Dispatch: March 24, 2018 | CE: | No. of Pages: 20 | PE: Marita Beyer

Received: 4 February 2018 | Accepted: 14 March 2018
DOI: 10.1002/maco.201810085

Making use of external corrosion defect assessment (ECDA) data to predict DCVG %IR drop and coating defect area

Nik N.Q1 Bin Muhd Noor[1] | Keming Yu[1] | Ujjwal Bharadwaj[2] | Tat-Hean Gan[2]

[1] Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK

[2] TWI Ltd., Granta Park, Cambridge, Cambridgeshire CB21 6AL, UK

Correspondence
Nik N. Bin Muhd Noor, Brunel University London, Kingston Lane, Uxbridge UB8 3PH, UK.
Email: nik.noor1982@gmail.com, niknooruhafidzi.binmuhdnoor@brunel.ac.uk

Buried pipelines are vulnerable to the threat of corrosion. Hence these pipelines are coated with a protective layer (coating) to isolate the metal substrate from the surrounding environment. With time, the coating will deteriorate which could lead to corrosion. The condition of the coating can be investigated by the external corrosion direct assessment (ECDA) procedure to investigate and monitor corrosion activity on unpiggable pipelines and provides a guideline in maintaining its structural integrity. This paper highlights the results obtained from the ECDA process which was conducted on 250 km of buried pipelines. The results from the indirect and direct assessment part of the ECDA were modeled using the classical quantile regression (QR) and the Bayesian quantile regression (BQR) method to investigate the effect of factors toward the IR drop (%IR) and the coating defect size (TCDA). It was found that the classical method and the Bayesian approach produces similar predictions on the regression coefficients. However, the Bayesian method has the added advantage of the posterior distribution which considers parameter uncertainties and can be incorporated in future ECDAs.

KEY W O R DS

Bayesian quantile regression, buried pipelines, coating defect, direct current voltage gradient, external corrosion direct assessment

1 | INTRODUCTION

The safest form of transportation of oil and gas products is by the use of pipelines.[1] Failure of pipelines is rare but causes from third party interference such as excavation of pipeline locations, corrosion of the metal substrate, and operational issues can jeopardize a pipeline from operating normally.[2] A pipeline failure which results in the loss of containment has the potential to impact the society, environment and the company's economy.[3] Due to this, the structural integrity of pipelines is at the top of every operator's list in keeping the pipelines from failing and working in a safe and normal manner.

For buried pipelines, corrosion threats are a major concern.[4] The threat is minimized by the application of external coating on the outer surface of the pipeline.[5] In theory, the reaction of microscopic corrosion cells which is present on the metal surface is prevented by the application of a non-conductive material which separates the metal surface from the environment.[6]

The failure of pipeline coatings can occur in various ways. Normally, coatings are made from organic materials which makes them susceptible to deteriorate over time. The failure can also be due to the incorrect application of the coating, soil stresses experienced by the pipeline or the coating's adhesive properties has lost its functionality. Generally, failure of coatings can be summarized as the changes in any of the chemical, physical, or electrochemical properties of the coating.[7] The result of these failures is the discontinuity of the coating (defect) which leaves the metal substrate exposed to the environment. In the event of coating discontinuity, corrosion is likely to occur which could undermine the whole structural integrity of the pipeline.

Buried pipelines are normally protected with a cathodic protection (CP) system. This system acts as backup to the coating system and comes into play when defects are present on the pipeline's coating.[8] The monitoring of the CP system, the coating and the overall integrity of the pipeline is normally addressed by conducting an external corrosion direct assessment (ECDA).[9] As part of the ECDA process, an indirect assessment which is commonly used is the Direct Current Voltage Gradient (DCVG). This technique is used to identify the location of coating defects and to classify its severity. Based on a defect severity, a decision can be made on whether to proceed with further direct assessment which requires excavation of the defect location. The DCVG technique is considerably accurate in locating a defect location but lacks the accuracy in predicting its size (area).[10] The prediction of coating defect area has not yet been a popular research theme within the academic sphere and the pipeline industry. The authors found only a handful of literature relating to this topic. The most noticeable of which was done by Ref. [11]. In this paper, a quantile regression was used to model the relationship between the coating defect area and its possible contributors. The paper also sheds light on the challenges faced by pipeline operators when interpreting DCVG indications. McKinney,[12] has produced a model which estimates the coating defect area based on simulated data. The approach taken is deterministic where a finite element method (FEA) was used. Moghissi et al.[13] have identified that there is no simple solution toward prioritizing coating defects for further assessment. Data were collected from the Closed Interval Pipeline Survey, DCVG and current attenuation assessments and were used to derive basic formulations to model the relationship between coating defect area and its possible contributing factors. The approach taken in Ref. [13] uses similar methods as those found in the work by Ref. [12].

The motivation for the work reported in this paper is to supplement the body of knowledge highlighted above. Statistical models are proposed to better explain the inner workings of a DCVG indication for the prioritization of coating defects for subsequent direct examination of the affected pipeline.

Quantile regression is used to fully characterize the dependent variable without relying on assumptions of the response distribution e.g., normally distributed. As compared to the mean regression, quantile regression is much more robust to outliers since it employs absolute values of the error terms.[14] Judging by the distribution of the response variable from the MEOC data (which will be described in detail in the following section) which is represented by Figures 1 and 2, the total coating defect area (TCDA) and the %IR (IR drop) variable demonstrate a distribution which is not normal nor symmetric and have some degree of skewness. Distributions such as the ones above, are asymmetrical and hence need more complex solution in describing the entirety of the response variable's distribution.[15]

The Bayesian approach toward quantile regression was elaborated by Ref. [16]. Bayesian inferences is more advantageous than the classical approach in mainly two instances: 1) Bayesian statistics does not rely on asymptotic variances of the estimators and 2) the estimated parameter includes the parameter uncertainty in the form of a posterior distribution. Since the mechanism of cathodic protection are complex, uncertainty of parameter values becomes an inherent trait. The Bayesian approach helps us to quantify this uncertainty. The findings from this paper can then be used as prior information for the next iteration of the ECDA process.

This paper is divided into several sections. Section 2 describes the data that was obtained from a recent ECDA project conducted by TWI Ltd. Section 3 outlines the methodology used followed by Section 4 which presents the results from the models. Section 5 is discussion and Section 6 is the conclusion and future work.

2 | MEOC DATA

The Middle Eastern Oil Company (MEOC) has appointed TWI Ltd. to conduct an ECDA on its network of pipelines. There are a total of nine (9) pipelines, all of which are non-piggable. The ECDA conducted by TWI Ltd. complied with the ANSI/NACE SP0502-2010: Standard Practice Pipeline External Corrosion Direct Assessment Methodology. The ECDA is divided into four parts. Data from these parts were gathered and annotated by the authors to be used in the modeling process.

2.1 | Pre-assessment

The data in this section includes the design data of the pipe which included its design philosophy, material selection, and the pipe

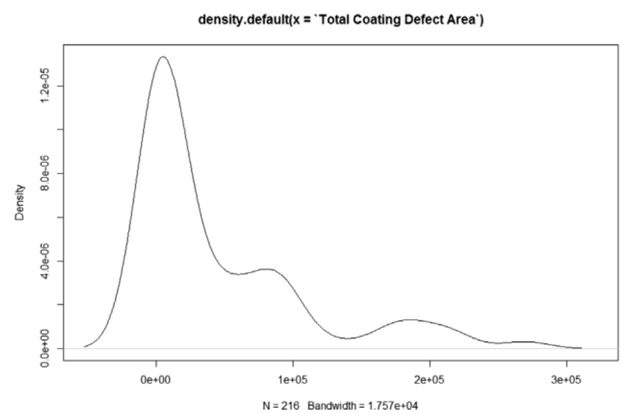


FIGURE 1 Probability density plot for TCDA. Reproduced with permission from TWI Ltd.

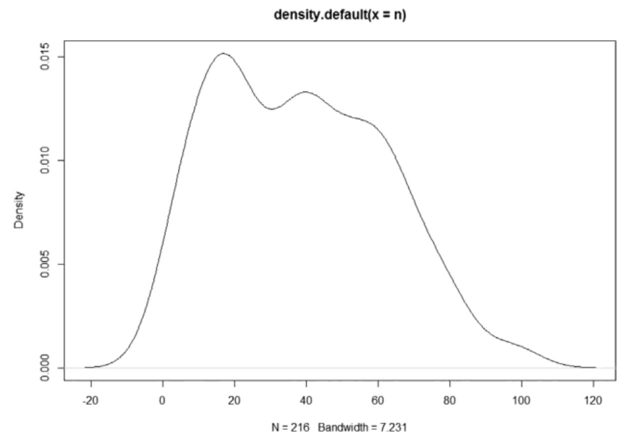


FIGURE 2 Probability density plot for %IR. Reproduced with permission from TWI Ltd.

physical characteristics. Historical operation data is also present in this section. It was found that the total length of the nine pipelines covers over 250 km. The age (time in service – TIS) for these pipes ranges from 19 to 39 years. The pipe sizes (PS) are from 26 to 42 inches. Operating pressure is from 8 to 17 Bar. The material grade for these pipes are API5L-X52 and X60. Working pressure of the pipes ranges from 40 to 60 °C with a 400 to 1520 $m^3 h^{-1}$ of fluid flow rate. Coatings types used for the nine pipelines ranges from cold wrap, coal tar, and polyethylene.

2.2 | Indirect assessment

This section of the ECDA process specifies the indirect tests that was used to investigate the CP condition and the coating condition of the pipeline. Techniques such as the close interval potential survey (CIPS), direct current voltage gradient (DCVG), alternate current voltage gradient (ACVG), and pipeline current mapper (PCM) were conducted to obtain information on the state of the pipeline.

The DCVG technique was identified as the most suitable technique to specify the coating defect area as it provided an established method of calculating the size of coating discontinuities. Once a defect is located, the total voltage (total mV) is calculated and divided by the pipeline's potential at that defect location. The pipe's potential is an interpolation of the relative distance of the defect to two bracketing test posts. This value is later multiplied by a hundred to get a percentage value which is called the percentage IR or IR drop (%IR). The values of the %IR are taken as the data needed for the construction of the model.

2.3 | Direct assessment

Direct assessment of defects provided us with a lot of useful data. After the identification of coating defects and calculation of its severity (based on %IR), decisions can be made on where to excavate to further analyze the defect. The decisions were based on the magnitude of the %IR and the pre-assessment data. As can be seen here, the decision relies on subjective judgment of the engineers. At excavation sites, data collected are the soil resistivity (SR) which is based on the four pin method,[17] the depth of buried pipe (DOC), the material of cover, the pH of the soil, the pH of water underneath the coatings and where there is corrosion activity the depth of the corrosion pits (POPD) using ultrasonic measurements and pit gauges. The size of the coating defects was also measured and were summed up (at one excavation site) to become the total coating defect area (TCDA). Deposits underneath coatings (DUC) were also annotated where present. The amount of deposit underneath the coating in terms of area is divided with the TCDA to gain a percentage value. All the data collected in this phase were considered as factors toward the prediction of the coating defect area. A complete list of the variables used for modeling based on the data gathered from the indirect and direct phase of the ECDA is listed in Table 1.

3 | METHODOLOGY

The objective for this paper can be divided into two. The first one is the construction of a model which summarizes all the contributing factors toward the %IR. A further refinement (lesser contributing variables) of the model is also constructed based on the industry's understanding on the system. This was done by consulting experts from the field. The second objective is to present a model which predicts the TCDA based on environmental data. Additionally, there are two versions of the dataset. The first version is the data set that included every measurement from the ECDA process. We shall name this the "Oriset" data. This is the original dataset received by the authors. The second version of the data set is called the "Filtset" and was scrutinized by the authors on what to expect from a DCVG indication relating to its size of coating defect. A total of four data points which were considered as outliers were taken out the data set. The data points removed was in the form of the outliers present in the distribution of TCDA where larger TCDA gives us lower values of %IR. For ease of referencing the models are named as follows in Table 2.

Two techniques were applied to the two datasets. The first approach is by the usage of the Bayesian quantile regression (BQR) to obtain model estimates. The second is the classical approach which employs quantile regression (QR).

3.1 | BQR

In classical statistics, assumptions were made on the estimated parameters where the value is considered fixed, but the quantity is unknown. Unlike the classical approach, Bayesian inference is a new way of thinking about statistics. The parameter of interest is not fixed but a random variable.

Based on the paper by Yu and Moyeed,[16] the *pth* regression quantile ($0 < p < 1$) can take on any solution, $\hat{\beta}(p)$;

TABLE 1 Types[Q2] of variables considered for the regression assessment

Symbol	Variables considered	Type of variable/summary statistics	
α	IR drop (%IR)	Quantitative	
		Min. value	0
		1st quantile	17.87%
		Median	37.8%
		Mean	38.48%
		3rd quantile	56.7%
		Max. value	98.9%
β	Soil resistivity (SR)	Quantitative	
		Min. value	75.36 Ω-cm
		1st quantile	560.25 Ω-cm
		Median	1282 Ω-cm
		Mean	2722.11 Ω-cm
		3rd quantile	2508.14 Ω-cm
		Max. value	43 332 Ω-cm
γ	Percentage of pit depth to wall thickness (POPD)	Quantitative	
		Min. value	0%
		1st quantile	0%
		Median	2.537%
		Mean	10.451%
		3rd quantile	17.471%
		Max. value	100%
δ	Deposits under coatings (DUC)	Quantitative	
		Min. value	0%
		1st quantile	3%
		Median	30%
		Mean	35.4%
		3rd quantile	60%
		Max. value	100%
ε	Depth of cover (DOC)	Quantitative	
		Min. value	0 cm
		1st quantile	100 cm
		Median	110 cm
		Mean	109.5 cm
		3rd quantile	130 cm
		Max. value	210 cm
ζ	Time in service (TIS)	Quantitative	
		Min. value	19 years
		1st quantile	20 years
		Median	36 years
		Mean	32.5 years
		3rd quantile	39 years
		Max. value	39 years
η	Pipe size (PS)	Quantitative	
		Min. value	26 inches

(Continues)

TABLE 1 (Continued)

			1st quantile	36 inches
			Mean	35.3 inches
			Max. value	42 inches
	Total coating defect area	Quantitative		
			Min. value	0 cm^2
			1st quantile	1200 cm^2
			Median	9985 cm^2
			3rd quantile	77 865 cm^2
			Max. value	269 894 cm^2
Backfill type				
θ	Rock	Qualitative		
κ	Sand + clay	Qualitative		
λ	Stones + clay	Qualitative		
Coating type				
μ	Coal tar	Qualitative		
ξ	Polyethylene	Qualitative		
CW	Cold wrap	Qualitative		
Backfill geometry				
ρ	Angular	Qualitative		
σ	Round + angular	Qualitative		
R	Rounded	Qualitative		
pH of water in soil				
φ	Acidic	Qualitative		
χ	Alkaline	Qualitative		
ψ	Neutral	Qualitative		
pH of water underneath coating				
ω	Acidic	Qualitative		
$\ddot{\upsilon}$	Alkaline	Qualitative		
$\ddot{\imath}$	Neutral	Qualitative		

Reproduced with permission from TWI Ltd.

and is associated to the aforementioned quantile regression minimization problem (minimization β)Q3

$$\min \sum_t \rho_p \left(y_t - x_t' \beta \right); \quad (1)$$

the *loss function* being

$$\rho_p(u) = u(p - I(u < 0)) \quad (2)$$

Yu and Moyeed[16] also showed that the minimization of the loss function above is exactly the same as maxi-mizing the likelihood function which is formed by joining independently distributed Asymmetric Laplace Densities (ALD).

The probability density function of the ALD is given as follows[18]

$$f(y; \mu, \sigma, p) = \frac{p(1-p)}{\sigma} \exp\left\{ -\rho \left(\frac{y-u}{\sigma} \right) \right\} \quad (3)$$

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

TABLE 2 Names of the various models corresponding to each dataset

Description of model	Dataset	Model name
Contribution to %IR model – full variables	Oriset	Model 1
Contribution to %IR model – refined variables	Oriset	Model 1a
Contribution to %IR model – full variables	Filtset	Model 2
Contribution to %IR model – refined variables	Filtset	Model 2a
TCDA model – full variables	Oriset	Model 3
TCDA model – refined variables	Filtset	Model 4

Reproduced with permission from TWI Ltd.

$$\pi(\beta|y) = \text{Likelihood}(y|\beta) \times g(\beta) \tag{4}$$

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

$$\text{Likelihood}(y|\beta) = p^n(1-p)^n \exp\left\{-\sum_i \rho_p\left(y_i - x_i'\beta\right)\right\} \tag{5}$$

As for the specification of priors, one can use any prior. But in the absence of a prior (as with the research presented in this paper – due to the lack of expert opinion and the limited amount of data), Yu and Moyeed[16] have proven that a non-informative improper prior yields a proper posterior distribution. In this method, there are no known conjugate priors but with the relative ease of using Markov chain Monte Carlo (MCMC) with the Metropolis Hastings algorithm, one is able to easily produce the posterior distribution of the parameters.

4 | MODEL ESTIMATION AND RESULT ANALYSES

Both the Oriset and the Filtset data were applied to the two regression techniques. Variables considered were the variables highlighted in Table 1. All the analyses were done in the statistical software R.

4.1 | Contributing factors to %IR (Model 1)

The estimates from Table 3 showed interesting results particularly on the TCDA variable. Iterations of up to 1 million of the MCMC was conducted to achieve convergence. This can be seen in Figure 3 where the trace plot and posterior histogram of various quantiles is presented. A quantile plot of the variable TCDA is shown in Figure 4. The maximum estimated coefficient value occurs at the 0.5 quantile where a 1 cm^2 increase in coating defect size reflects in an increase of 0.0000687% of %IR. If we increase the percentage values to 100% (maximum reading of the DCVG indication), the max coating defect size the DCVG technique can detect is 1 455 604 cm^2. The lowest estimated value for the TCDA occurs at the 0.05 quantile. The estimated coefficient reveals a 1 cm^2 increase in TCDA will increase the %IR value by 0.0000022%. This shows that medium-sized coating defects give the largest signal on the DCVG indication where small defects contribute the least. Also, at the 0.5 quantile, the credible interval is much narrower compared to the ones at the two opposite ends indicating lower uncertainty. Equations of various quantiles are presented in the following:

$$\begin{aligned}\%IR_{0.05} = {} & 44.2 + 0.0000022\text{TCDA} - 0.0000235\beta \\ & + 0.00611\gamma + 0.0079\delta + 0.0549\varepsilon \\ & - 0.336\zeta - 0.0818\eta + 5.2\theta \\ & - 1.03\kappa + 1.72\lambda - 3.26\mu \\ & - 6.28\xi + 0.754\rho - 2.64\sigma \\ & + 1.17\varphi + 8.41\chi + 7.24\psi \\ & - 0.943\omega - 2\ddot{\upsilon} + 2.56\ddot{\imath}\end{aligned} \tag{6}$$

$$\begin{aligned}\%IR_{0.5} = {} & 86.1 + 0.0000687\text{TCDA} - 0.000567\beta \\ & + 0.0439\gamma - 0.0372\delta + 0.0933\varepsilon - 0.374\zeta \\ & - 1.31\eta + 50.8\theta + 16.3\kappa + 0.562\lambda \\ & - 0.215\mu + 0.368\xi - 19.9\rho - 0.835\sigma \\ & - 8.1\varphi + 0.753\chi + 7.03\psi \\ & - 3.24\omega - 7.78\ddot{\upsilon} - 0.125\ddot{\imath}\end{aligned} \tag{7}$$

$$\begin{aligned}\%IR_{0.95} = {} & 23.6 + 0.0000532\text{TCDA} - 0.000346\beta \\ & + 0.108\gamma - 0.0704\delta + 0.0364\varepsilon \\ & + 1.19\zeta + 0.285\eta + 10.7\theta - 11.6\kappa \\ & + 0.411\lambda + 11\mu + 2.44\xi - 8.69\rho \\ & - 0.446\sigma - 4.8\varphi - 0.804\chi - 14\psi \\ & - 11\omega + 0.991\ddot{\upsilon} - 1.99\ddot{\imath}\end{aligned} \tag{8}$$

Soil resistivity also plays a role in the contribution to the %IR. The maximum (lowest) estimated value for soil resistivity occur at the 0.5 quantile with a value of −0.000567. This can be interpreted as a 1 unit increase of soil resistivity will lead to a decrease of 0.000567% with respect to %IR. However, the variable backfill type – rock which is related to the resistant nature of the soil, showed an inverse effect. Across the quantiles, the estimated coefficients point to meaningful contribution toward the %IR readings especially within the range of 0.25 to the 0.75 quantile.

TABLE 3 Bayesian quantile regression (BQR) estimates with 95% credible intervals for quantiles 0.05, 0.5, and 0.95 for Model 1

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

Reproduced with permission from TWI Ltd.

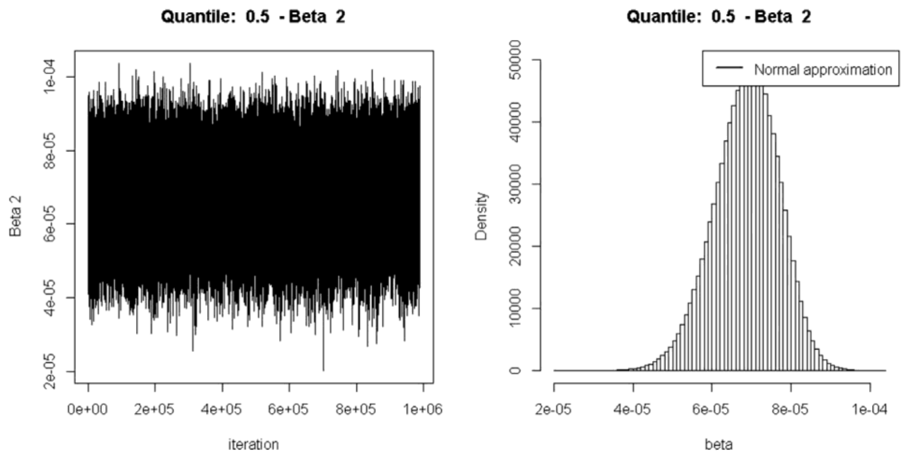


FIGURE 3 Example of a trace plot and posterior histogram of the 0.5 quantile for the estimated coefficient, TCDA for Model 1. Reproduced with permission from TWI Ltd.

4.2 | Refined %IR (Model 1a)

The results of the estimated coefficients by BQR for Model 1a is presented in Table 4 below. In achieving convergence for all the variables, iterations of up to 300 000 were determined with the initial 5000 steps regarded as burn-in. For the variable of interest, the TCDA, the maximum estimated value occurs at the 0.5 quantile. This prediction is similar to the one predicted by Model 1. As for the overall estimated trend, it follows the same pattern as Model 1 with Model 1a being more pronounced. The value of the coefficient at the maximum is 0.0000828. This means that a 1 cm^2 of TCDA will have an effect on the %IR by 0.0000828%. At the 0.05 quantile, the coefficient value is at its lowest with a value of −0.0000353. The negative value signifies that with a 1 cm^2 increase in TCDA will yield a 0.0000353% decrease in %IR.

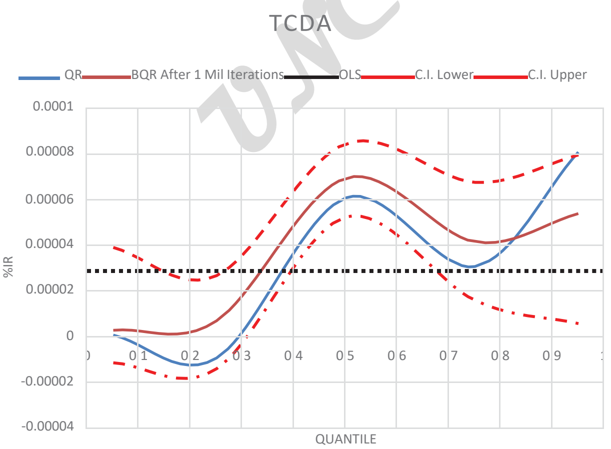


FIGURE 4 Example of a quantile plot of the TCDA variable for Model 1. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

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

$$\%IR_{0.05} = 4.74 - 0.0000353TCDA + 0.000000565\beta + 0.0508\varepsilon - 0.158\eta + 5.23\theta - 0.939\kappa + 1.56\lambda + {}^{2}\mu - 0.113\xi + 0.434\rho - 3.65\sigma + 1.2\varphi + 8.11\chi + 5.06\psi \quad (9)$$

$$\%IR_{0.5} = 87.5 + 0.0000828TCDA - 0.000668\beta + 0.0722\varepsilon - 1.77\eta + 53.4\theta + 25.4\kappa + 0.619\lambda + 5.54\mu + 6.77\xi - 18.2\rho + 0.251\sigma - 6.07\varphi + 1.76\chi + 1.14\psi \quad (10)$$

$$\%IR_{0.95} = 64.9 + 0.000073TCDA - 0.000296\beta + 0.0228\varepsilon + 0.432\eta + 6.45\theta - 8.21\kappa - 1.46\lambda - 6.78\mu - 15.3\xi - 6.73\rho - 1.24\sigma - 4.67\varphi - 0.575\chi - 12.6\psi \quad (11)$$

4.3 | Contributing factors to %IR (Model 2)

The estimated coefficients for Model 2 are given in Table 5. However, the reference variable is substituted to be backfill type – rock, coating type – polyethylene and backfill geometry – angular. This is due to investigate on the factors regarding soft soils which included clay with rounded grain

TABLE 4 Bayesian quantile regression (BQR) estimates with 95% credible intervals for quantiles 0.05, 0.5, and 0.95 for Model 1a

(Intercept)	4.74	−1.53	14.4	87.5	78.2	97	64.9	56	84.8
Soil resistivity (SR)	0.000000565	−0.000344	0.000364	−0.000668	−0.000863	−0.000416	−0.000296	−0.000641	0.0000428
Pipe size (PS)	−0.158	−0.388	0.0759	−1.77	−2.01	−1.53	0.432	−0.199	0.64
Backfill type (sand + clay)	−0.939	−7.93	3.08	25.4	16.1	36.8	−8.21	−22.3	2.47
Coating type (coal tar)	2	−0.739	6.36	5.54	3	8.17	−6.78	−11.9	0.445
Backfill geometry (angular)	0.434	−3.6	4.73	−18.2	−22.7	−14.4	−6.73	−30.5	2.56
pH of water in soil (acidic)	1.2	−10.2	16.7	−6.07	−12.7	0.598	−4.67	−42.2	9.47
pH of water in soil (neutral)	5.06	−1.05	12.4	1.14	−1.34	4.89	−12.6	−19.3	0.294

Reproduced with permission from TWI Ltd.

structure. A total of 400 000 iterations were made to get to the point of convergence with the initial 5000 readings as burn-in. Table 5 shows the TCDA variable coefficients has an upward trend with a slight dip at the 0.25 quantile. The highest value is reached at the 0.95 quantile with a value of 0.000229%. For a 1 cm^2 increase in the size of coating defect area, a 0.000229% increase in percentage IR is expected. This is higher than the maximum obtained by Model 1. Additionally, this happens at the 0.95 quantile which goes well with established understanding of the technique as compared to Model 1 where the maximum occurred at the 0.5 quantile. This is mainly due to the contribution of the careful judgement of the authors which obliterated four points from the original set.

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

The variable clay showed increasing trends across the % IR distribution with a dip at the 0.95 quantile. The maximum estimated coefficient was noticed to be at the 0.75 quantile with a value of 60.8. This can be translated as the effect of the presence of clay to the %IR will be the most at the 0.75 quantile of the %IR distribution.

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

$$\begin{aligned}\%IR_{0.05} = {} & 14.7 + 0.0000741\,\mathrm{TCDA} + 0.0000293\beta + 0.0334\gamma - 0.0209\delta + 0.0668\varepsilon - 0.116\zeta \\ & - 0.126\eta - 2.26C - 1.69\kappa + 11.4\lambda \\ & - 9.26\,\mathrm{CW} + 7.57\xi + 1.73R - 0.246\sigma \\ & + 3.98\varphi - 1.84\chi - 11.4\psi + 2.8\omega \\ & - 1.62\ddot{\upsilon} - 4.96\ddot{\iota} \end{aligned} \tag{12}$$

$$\begin{aligned}\%IR_{0.05} = {} & 79.4 + 0.0000618\,\mathrm{TCDA} + 0.000206\beta + 0.161\gamma - 0.0373\delta + 0.00696\varepsilon - 0.234\zeta \\ & - 0.3\eta - 3.35C - 16.7\kappa + 3.29\lambda \\ & - 31.4\,\mathrm{CW} + 1.45\xi + 1.02R \\ & - 0.156\sigma + 7.4\varphi - 21.2\chi \\ & - 11.2\psi + 1.02\omega - 8.36\ddot{\upsilon} - 6.67\ddot{\iota} \end{aligned} \tag{13}$$

$$\begin{aligned}\%IR_{0.95} = {} & 22 + 0.000229\,\mathrm{TCDA} + 0.000373\beta + 0.0558\gamma - 0.05\delta + 0.0982\varepsilon + 0.179\zeta \\ & + 0.186\eta + 16.3C + 4.9\kappa + 25.4\lambda \\ & - 5.69\,\mathrm{CW} + 29.7\xi - 1.58R + 1.58\sigma \\ & + 9.2\varphi - 2.62\chi - 16.8\psi - 4.98\omega \\ & - 3.43\ddot{\upsilon} - 13.1\ddot{\iota} \end{aligned} \tag{14}$$

4.4 | Refined %IR (Model 2a)

Table 6 shows the estimated coefficients predicted by the BQR method with the Filtset data for Model 2a. 400 000 iterations were made to achieve convergence with the initial 5000 recordings regarded as burn-ins. At the 0.05 quantile, the predicted TCDA coefficient showed similar results to the one obtained for Model 2. The coefficient value drops at the 0.25 quantile and rising steadily after this all the way up to the 0.95 quantile where it reaches its maximum. Maximum predicted value stands at 0.000221 which means a 1 cm^2 increase in TCDA will give an increase of 0.000221% in %IR. Previously for Model 2, similar characteristics were observed with only slight differences in the predicted values.

Soil resistivity plays a role in Model 2a where an increasing trend is observed starting from the 0.25 quantile all the way up to the 0.95 quantile. The highest predicted value is at the 0.95 quantile with a Bayes estimate of 0.000482. At the 0.95 quantile, a 1 unit increase in the value of soil resistivity will mean a 0.000482% increase in %IR.

The presence of clay as the backfill material will affect the %IR differently across the quantile of the %IR distribution when compared to the soil resistivity variable. Clay affect the 0.75 quantile the most with the 0.05 the least affected. The value of the maximum estimate coefficient is 57. This is not far off than the estimated value at the same quantile for Model 2. The upward trend up to the 0.75 quantile reflects that clay has a positive effect in the contribution of the %IR reading.

Models of various quantiles are presented in the following equations.

$$\begin{aligned}\%IR_{0.05} = {} & 30.3 + 0.0000956\,\mathrm{TCDA} - 0.000132\beta \\ & + 0.0561\varepsilon - 0.301\eta - 18.9C \\ & - 19.1\kappa - 6.25\lambda - 6.48\,\mathrm{CW} \\ & + 8.51\xi + 1.22R - 0.0602\sigma \\ & + 4.13\varphi - 1.26\chi - 11.3\psi \end{aligned} \tag{15}$$

$$\begin{aligned}\%IR_{0.5} = {} & 86.2 + 0.0000768\,\mathrm{TCDA} - 0.000178\beta \\ & - 0.0665\varepsilon - 0.452\eta + 0.785C \\ & - 20.3\kappa - 3.21\lambda - 32.1\,\mathrm{CW} \\ & - 0.72\xi - 0.279R - 0.243\sigma \\ & + 6.63\varphi - 18.9\chi - 11.6\psi \end{aligned} \tag{16}$$

$$\begin{aligned}\%IR_{0.95} = {} & 31.3 + 0.000221\,\mathrm{TCDA} + 0.000482\beta \\ & + 0.0639\varepsilon - 0.00579\eta + 18.6C + 10.9\kappa \\ & + 25.9\lambda - 8.59\,\mathrm{CW} + 29.7\xi \\ & - 0.829R + 1.22\sigma + 11.9\varphi + 4\chi - 16.8\psi \end{aligned} \tag{17}$$

4.5 | Total coating defect area (TCDA) models

With the establishment of the %IR models utilizing both the Oriset and the Filtset data (Models 1, 1a, 2, and 2a), the construction of the TCDA model will further increase the capability of operators and decision makers in prioritizing coating defects based on their severity. To add to this enhancement, we propose TCDA models (Models 3 and 4) which predict the coating defect area based on variables from

TABLE 5 Bayesian quantile regression (BQR) estimates with 95% credible intervals for quantiles 0.05, 0.5 and 0.95 for Model 2

(Intercept)	14.7	−1.76	40.681215	79.4	43.7	92.1	22	−415	69.1
Soil resistivity (SR)	0.0000293	−0.000316	0.000383	0.000206	0.0000175	0.000329	0.000373	−0.0000612	0.000848
Deposits under coatings (DUC)	−0.0209	−0.0789	0.029226	−0.0373	−0.0655	−0.0111	−0.05	−0.115	0.00565
Time in service (TIS)	−0.116	−0.328	0.085368	−0.234	−0.338	−0.128	0.179	−0.22	0.788
Backfill type (clay)	−2.23	−13.6	3.720633	3.35	−2.41	23.9	16.3	−7.2	260
Backfill type (stones + clay)	11.4	−0.0934	18.838816	−3.29	−9.69	18.8	25.4	0.026	262
Coating type (Coal tar)	7.57	−11	20.359542	−1.45	−8.49	7.7	29.7	16.4	38.7
Backfill geometry (round + angular)	−0.246	−4.13	3.671415	−0.156	−2.4	2.01	1.58	−2.14	10.4
pH of water in soil (alkaline)	−1.84	−7.75	0.907865	−21.2	−24.3	−17.7	2.62	−1.7	12.2
pH of water underneath coating (acidic)	2.8	−6.1	18.678275	1.02	−2.44	9.92	−4.98	−27.4	5.78
pH of water underneath coating (neutral)	−4.96	−16.5	0.619875	−6.67	−14.2	0.543	−13.1	−33.1	2.77

TABLE 6 Bayesian quantile regression (BQR) estimates with 95% credible intervals for quantiles 0.05, 0.5, and 0.95 for Model 2a

	Quantiles								
	0.05			0.5			0.95		
		Credible intervals			Credible intervals			Credible intervals	
Variables	Posterior mean	0.025	0.975	Posterior mean	0.025	0.975	Posterior mean	0.025	0.975
(Intercept)	30.3	−2.29	501	86.2	77.7	94.1	31.3	−308	79.5
Total coating defect area (TCDA)	0.0000956	0.0000936	0.000138	0.0000768	0.0000547	0.0000989	0.000221	0.000186	0.000222
Soil resistivity (SR)	−0.000132	−0.000408	0.000202	−0.000178	−0.000395	0.0000902	0.000482	0.0000361	0.00108
Depth of cover (DOC)	0.0561	0.0195	0.0981	−0.0665	−0.0907	−0.0337	0.0639	0.0178	0.0949
Pipe size (PS)	−0.301	−0.333	0.00686	−0.452	−0.622	−0.309	−0.00579	−0.571	0.214
Backfill type (clay)	−18.9	−393	4	0.785	−2.71	5.56	18.6	−6.5	323
Backfill type (sand + clay)	−19.1	−392	3.51	−20.3	−26.2	−13.6	10.9	−24.2	300
Backfill type (stones + clay)	−6.25	−381	17.7	−3.21	−7.47	0.767	25.9	−0.151	324
Coating type (PVC cold wrap)	−6.48	−51	2.72	−32.1	−36.8	−28.5	−8.59	−29.3	0.1
Coating type (Coal tar)	8.51	−37.8	18.5	−0.72	−5.65	2.67	29.7	9.06	38.9
Backfill geometry (round)	1.22	−5.87	5.43	−0.279	−2.77	1.51	−0.829	−6.18	2.5
Backfill geometry (round + angular)	−0.0602	−5.43	3.04	−0.243	−3.27	2.1	1.22	−1.56	7.78
pH of water in soil (acidic)	4.13	−9.69	39.6	6.63	−0.543	14	11.9	0.0936	31.7
pH of water in soil (alkaline)	−1.26	−6.84	1.42	−18.9	−21.5	−15.8	4	−1.39	12.3
pH of water in soil (neutral)	−11.3	−18.2	−0.912	−11.6	−15.9	−6.64	−16.8	−24.1	−0.0591

TABLE 7 Bayesian quantile regression (BQR) estimates with 95% credible intervals for quantiles 0.05, 0.5, and 0.95 for Model 3

	Quantiles								
	0.05			0.5			0.95		
		Credible intervals			Credible intervals			Credible intervals	
Variables	**Posterior mean**	**0.25**	**0.975**	**Posterior mean**	**0.25**	**0.975**	**Posterior mean**	**0.25**	**0.975**
(Intercept)	−465	−496.4146	−445	78 687.177	74 052.516	79 414.664	189 000	128 000	207 000
%IR	−0.0178	−0.1091	0.0559	84.428	82.292	100.219	−93.1	−162	28.6
Soil resistivity (SR)	−0.0034	−0.0037	−0.00304	−0.524	−0.529	−0.476	0.151	−0.0154	0.482
Percentage of pit depth to wall thickness (POPD)	6.04	5.4327	6.79	232.204	228.826	264.055	2740	2640	2790
Deposits under coatings (DUC)	−0.901	−0.9835	−0.821	19.543	10.358	20.721	−257	−387	−212
Depth of cover (DOC)	−0.00321	−0.0523	0.0392	−69.776	−71.118	−53.681	−111	−122	−70.9
Time in service (TIS)	4.92	4.6761	5.16	−2351.485	−2362.104	−2319.907	−8030	−8200	−7690
Pipe size (PS)	10.4	9.831	11.3	707.098	702.75	728.063	6040	5670	7410

Reproduced with permission from TWI Ltd.

its environment. The variables chosen in this model are limited to only quantitative values due to the large amounts of subjective interpretations on the qualitative variables. Another reason for this is to avoid higher computational cost as Bayesian inference with the Metropolis-Hastings algorithm (MCMC – MH) is known to take large amounts of computational memory when dealing with large quantities of data. As from the previous section, the variable %IR is of prime interest as this is one of the first measurements obtained when conducting a DCVG assessment. By correctly interpreting what the signal mean, one can make sound judgment on the state of the coating under inspection.

4.6 | TCDA Model 3

The coefficients estimated by the BQR for Model 3 data are presented in Table 7. Convergence took 11 million iterations. The initial 5000 iterations were treated as burn-ins and are disregarded. Primary interest for the model is the %IR variable, which shows a close to zero estimates for the 0.05 and 0.25 quantiles. Beginning at the 0.5 quantile, we can see the trend increasing up to the 0.75 quantile and back down again at the 0.95 quantile. The maximum estimated coefficient is at the 0.75 quantile with a value of 849. This represents a 1 unit increase in %IR and represents an increase of 849 cm^2 in terms of TCDA. Therefore a 100% reading of the %IR will translate into 84 900 cm^2 of TCDA. Although this estimation is promising in determining the size of coating defects, the 0.95 quantile illustrates a different picture. The estimated coefficient for this quantile is −93.1. The negative values signify that a 1 unit increase in %IR equals to a decrease of 93.1 cm^2 in TCDA. Equations below are selected models for the 0.05, 0.5, and 0.95 quantile for Model 3.

$$\mathrm{TCDA}\left(\mathrm{cm}^2\right)_{0.05} = -465 - 0.0178\alpha - 0.0034\beta + 6.04\gamma - 0.901\delta - 0.00321\varepsilon + 4.92\zeta + 10.4\eta \tag{18}$$

$$\mathrm{TCDA}\left(\mathrm{cm}^2\right)_{0.5} = 78687.177 + 84.428\alpha - 0.524\beta + 232.204\gamma + 19.543\delta - 69.776\varepsilon - 2351.485\zeta + 707.098\eta \tag{19}$$

$$\mathrm{TCDA}\left(\mathrm{cm}^2\right)_{0.95} = 189000 - 93.1\alpha + 0.151\beta + 2740\gamma - 257\delta - 111\varepsilon - 8030\zeta + 6040\eta \tag{20}$$

The variable POPD (pit depth) showed useful insights into the correlation between TCDA and corrosion. Based on the trend shown, as the quantile increases, so does the estimated coefficient values. However, there appears to be a sudden dip at quantile 0.75 and picks up again at quantile 0.95. The maximum value estimated by the BQR is 2740 which equates to a 1 unit increase in depth of the corrosion pit corresponding to a 2740 cm^2 in TCDA. This occurs at the 0.95 quantile.

TABLE 8 Bayesian quantile regression (BQR) estimates with 95% credible intervals for quantiles 0.05, 0.5, and 0.95 for Model 4

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

Reproduced with permission from TWI Ltd.

Also, the width of the credible intervals across the quantiles are narrow as compared to previous predicted models.

4.7 | TCDA Model 4

As was previously mentioned, the data considered for this assessment is the Filtset. Results of the analyses is highlighted in Table 8. As was expected, the %IR variable showed a positive consistent increasing manner across the quantile. Starting at the 0.05 and 0.25 quantile, the increase of the estimated coefficients is subtle but for the 0.5 quantile the changes are much more abrupt with the values tapering back at the 0.75 and 0.95 quantile. The maximum value occurs at the 0.95 quantile with an estimated coefficient of 1481.9. In other words, an increase in 1 unit of %IR will reflect an increase in the TCDA of 1481.9 cm^2. Therefore, for larger defects (0.95 quantile) a reading of 100% in the %IR value corresponds to a 148 190 cm^2 in TCDA which is the maximum size the model can predict. For the lowest quantile, the maximum predicted size is 2.21 cm^2. The maximum predicted defect sizes for all the quantiles are shown in Figure 5.

The POPD variable represents the amount of corrosion activity present on pipelines under consideration. Referring to Table 8, the estimated coefficients showed increasing trend. From quantile 0.25 up to 0.75 the predicted values do not show significant differences. Abrupt changes can only be seen at the tails of the TCDA distribution i.e., the 0.05 and the 0.95 quantile.

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

$$TCDA\left(cm^2\right)_{0.05} = -464 + 0.0221\alpha - 0.00336\beta + 6.06\gamma - 0.916\delta - 0.00871\varepsilon + 4.95\zeta + 10.3\eta \quad (21)$$

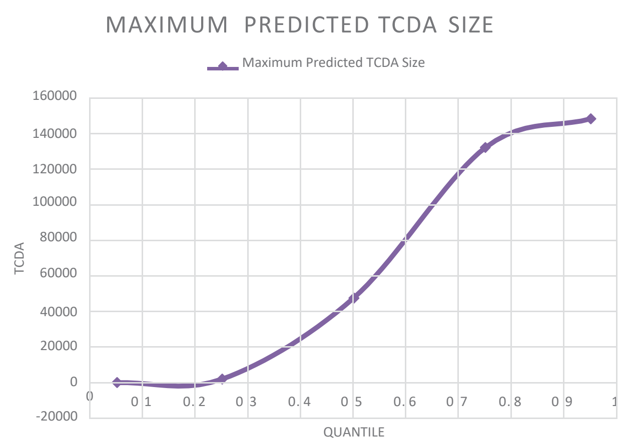


FIGURE 5 The maximum predicted TCDA size based on BQR for different quantiles of the TCDA Model. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

$$\mathrm{TCDA}\ (\mathrm{cm}^2) = 28978.772 + 475.876\alpha - 0.453\beta_5 + 219.149\gamma + 34.219\delta - 20.033\varepsilon - 1681.945\zeta + 963.069\eta \quad (22)$$

$$\mathrm{TCDA}\ (\mathrm{cm}^2) = 77655.6 + 1481.9\alpha + 12.2\beta_{95} + 406\gamma - 218.8\delta - 97.8\varepsilon - 5066.9\zeta + 4610.6\eta \quad (23)$$

5 | DISCUSSION

5.1 | Contributing factors to %IR – (Models 1, 1a, 2, and 2a)

5.1.1 | TCDA variable

The low coefficient values estimated for the TCDA variable (Models 1, 1a, 2, and 2a) was unexpected since the concept of a DCVG technique relied primarily upon coating defects to generate voltage drops. The results show coating defects in general have a mild effect on the %IR reading. Other known and unknown factors might also be a contributor toward %IR. One of these factors could be SR and the nature of the backfill geometry. Other factors could include the presence of interference in the form of stray currents especially if the pipeline is situated adjacent to other pipelines or is located near overhanging power cables. Although an interruption technique was used to eliminate foreign currents contributing to %IR indication, large structures such as buried pipelines need longer periods for it to depolarize and considered IR free.[19] To picture this more clearly, the following figures show the relationship between TCDA and %IR while keeping other variables constant. As was previously mentioned, other factors which gave rise to the %IR readings such as the POPD, DUC, DOC, TIS, PS, and SR were used to generate the models. These variables take on 10.5%, 35.2%, 109.5 cm, 32.5 years, 35.3 inch, and 2722.1 Ω-cm, respectively which represents the mean value of each variable.

Figure 6 shows the predictions made by Model 1 of the %IR with increasing TCDA. Generally, the models highlight an upward trend which is in parallel with the current understanding of the system. However, the slope of the models indicates a small effect of TCDA toward %IR. This can clearly be seen at the lower quantiles (0.05 and 0.25) where the line is almost flat. Also, the median quantile has the highest prediction value and the steepest slope which corresponds to the estimated coefficient values in Figure 4. A refined version of Model 1 is given by Model 1a which is presented in Figure 7. Similarly, the models take on the mean values of each contributing variable.

The prediction of the resulting %IR in Figure 7 shows an improvement in terms of the effect of TCDA on %IR with steeper slopes being observed. Similar to Model 1, the median of the %IR received the largest effect from the TCDA. The estimated %IR values based on the median is also higher with Model 1a as compared to Model 1. The removal of certain variables which do not contribute to the %IR has improved the %IR estimation for the top three quantiles. For the 0.25 quantile, small effects of the TCDA toward %IR are seen which is similar to the previous Model 1. However, the estimated values here are higher. The 0.05 quantile show decreasing trend where increasing TCDA relates to a decreasing of %IR.

The inconsistency (higher TCDA does not reflect a higher %IR values) for Models 1 and 1a with respect to the 0.05 quantile could possibly be attributed to the outliers present at higher and lower quantiles of the TCDA distribution – large defect areas are paired to low reading of the %IR and vice versa. Additionally, credible intervals at higher and lower quantiles for Models 1 and 1a are much wider indicating

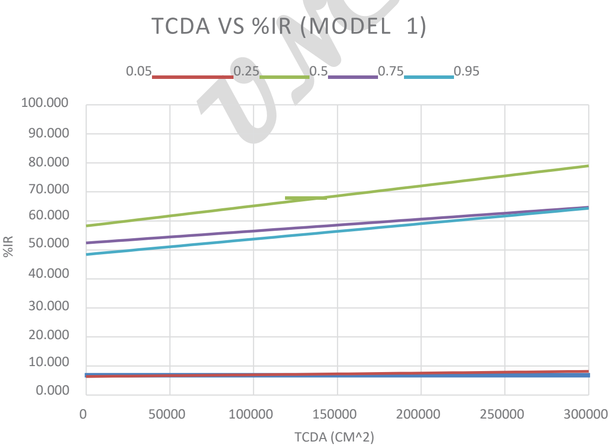


FIGURE 6 TCDA versus %IR for Model 1. Each color represent a different quantile. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

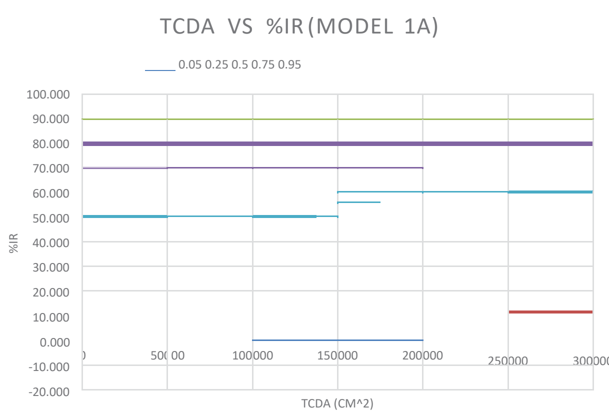


FIGURE 7 TCDA versus %IR for Model 1a. Each color represent a different quantile. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

higher uncertainty as compared to the median quantile where the maximum estimated values have occurred. Inconsistent results can also be summarized in the following bullet points.

- Interference in the form of stray or telluric currents which will interfere with the %IR signal. Currents from adjacent ICCP system, electrified railway tracks (DC traction system), overhead power cables, etc. have the potential of compromising the %IR signal. This can be seen at certain locations of the pipeline.
- Adjacent transmitting power cables could compromise the DCVG signal in the form of AC currents. AC currents can also lead to accelerated corrosion to the pipelines running below.[20] In the case of the MEOC pipelines, power cables can be seen running closely along and perpendicular to the direction of the buried lines.
- The heterogeneous nature of soils compromises or alters the measured voltage signal. The calculation of the %IR value requires input in the form of the pipeline-to-electrolyte interface resistance. The resistant value is related to the SR value which is measured at test posts. However, DCVG readings are conducted away from test posts where the magnitude of SR changes. The changes will contribute to the inconsistencies of the %IR measurements where the heterogeneity of soil is not considered in the %IR formula. Although SR measurements were taken for every excavated area, this was not included into the %IR calculation.
- Defects occurring at the 6 o'clock position will tend to attenuate the voltage signal and will not correspond to the true size of a defect.[11]

- Based on the report provided by TWI Ltd., there is a possibility that some of the coating defects were caused by the excavator during excavation of bell holes for the direct examination process. These defects were not present during the indirect assessment (DCVG measurements).
- Deposits of scales due to the cathodic protection current on the metal substrate will mask the true size of a coating defect. Measurements are perceived to be small based on the %IR reading. This is an erroneous representation of the true size of the defect.

The assessment on Models 2 and 2a utilizes the Filtset data. The estimated %IR readings based on Model 2 (Figure 8) and 2a (Figure 9) are given as follows. Similar to the previous Models 1 and 1a assessments, the mean of POPD, DUC, DOC, TIS, PS, and SR was used to generate these models.

Immediately, it can be seen that in Figure 8 the estimated values of the %IR are much improved than Models 1 and 1a. The effect of TCDA on %IR is also greater which reflects the underlying intention of a DCVG assessment. The highest predicted value of the %IR is at the 0.95 quantile which indicates TCDA has the highest effect on higher readings of the %IR. Additionally, narrower credible intervals were obtained highlighting in lesser uncertainty of the estimated coefficients. Therefore, the removal of four excavation points improves the overall estimation of the role of TCDA on %IR. However, looking at quantiles 0.05 and 0.25 shows an apparent effect of TCDA toward %IR. However, these estimates are below the zero line. For the 0.25 quantile, all the predicted readings of %IR are negative and it sits lower than the 0.05 quantile. Although the apparent outliers were removed for this assessment, there are other factors that might give an overall effect on the %IR predictions. Model 2a tries to find this answer by further refining the model through the omission of variables which in theory does not contribute to the generation of %IR. Model 2a's prediction of %IR is given as follows in Figure 9.

COLOR

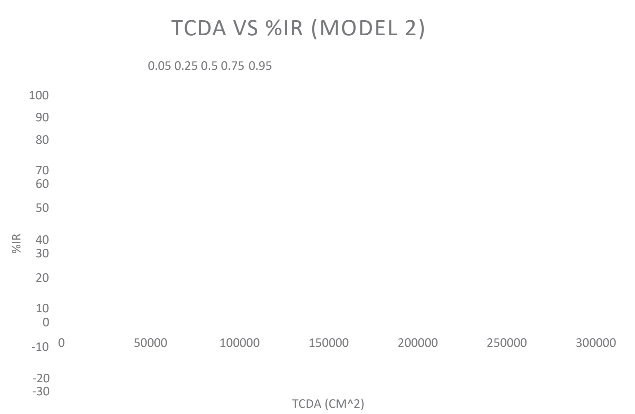


FIGURE 8 TCDA versus %IR for Model 2. Each color represent a different quantile. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

COLOR

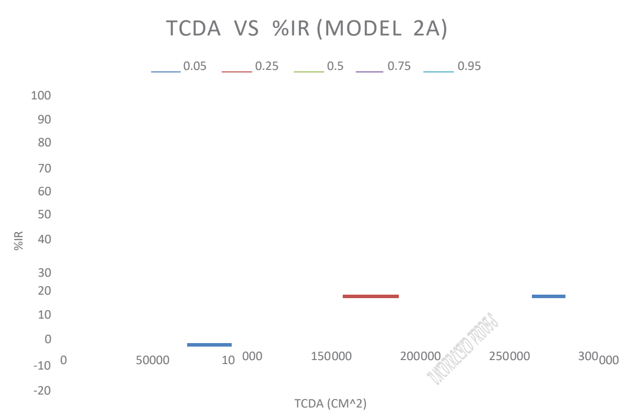


FIGURE 9 TCDA versus %IR for Model 2a. Each color represent a different quantile. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

From Figure 9 the omission of certain variables has improved the overall prediction of the %IR based on TCDA. Significant effects of the TCDA toward the %IR is seen across all the quantiles. The effect of higher TCDA on higher readings of %IR is seen with the highest predicted values of %IR occurring at the 0.95 quantile. This can also be said with other quantiles where lower values of TCDA effects the lowest part of the %IR readings. However, for the predicted values with regard to the 0.05 quantile, shows prediction values of %IR of less than zero. This small inconvenience can be stipulated as the cause of the linear approach taken by the authors when modeling the relationship. Overall, Model 2a is an acceptable model in the prediction of %IR (based on established literature on the DCVG technique) with the added bonus of simplicity and brevity due to its utilization of fewer variables.

5.1.2 | SR and backfill type variable

The SR estimated coefficients for Models 1 and 1a show a decreasing trend with its lowest value occurring at the 0.5 quantile region. However, the estimated effect of the rock variable on the contribution of %IR indicated an inverse trend with the maximum estimated coefficients occurring also within the region of 0.25–0.5 quantile. Since these two variables are somewhat related, the opposite predictions seem to complement each other and highlights the heterogenous nature of soil. Highly resistive electrolyte which contain materials such as rocks will produce large amounts of voltage drop as current passes through it. These voltage drops are likely to be detected by the DCVG instrument which indicates a defect more severe than it actually is. This is confirmed by the works of Mckinney[12] in his thesis which states that prioritization of DCVG indication will be more accurate if SR is taken into account. The higher quantiles highlight a relatively weak effect of the rock variable to %IR. However, this can be understood by also observing the value estimated for the general SR variable which highlights a stronger effect.

With respect to Models 2 and 2a, the reference variable for the models were changed and the variable backfill type – clay, shows increasing trend until it reduces at the 0.95 quantile. Clay is considered to have high degree of compactness thus possessing low resistance toward current flow. The low resistance would not produce large voltage drops and hence one would not expect the raising trend of the estimated coefficients. However, if we were to look at the backfill geometry – round variable, the estimates are much more streamlined with common understanding. The presence of rounded soil grains creates an environment which is less resistant to electrical currents (similar to clay). Across the quantile, the estimated coefficient values show a downward trend with a slight increase at the highest quantile. This is the inverse of the clay variable's trend. Similar to Models 1 and 1a, the two variables seem to complement each other and is only understood when both of them are looked at together. The decrease in the estimated value at the 0.95 quantile for the clay variable and the increase of the predicted value at the 0.95 quantile for the backfill geometry – round variable is the cause of a possible mixture of fine to coarse grain soils in the backfill. Moreover, there are also the possibility of foreign currents interfering with the measured signal as was mentioned above. Coupled this with the heterogeneous nature of soils, unexpected outcomes like this are not unusual to find.

5.2 | TCDA model _ (Models 3 and 4)

5.2.1 | %IR variable

The estimated coefficients for Model 3 has shown that the trend does not sit well with current industrial understanding on DCVG. A better way of visualizing this is by plotting the predicted TCDA based on increasing %IR using model. Other variables in the model were kept constant where the mean of the POPD, DUC, DOC, TIS, PS, and SR similar to previous assessments in this paper were used as the contributing factors.

Figure 10 shows the linear effect of %IR toward the resulting TCDA estimation. At the lowest quantiles (0.05 and 0.25) the effect is almost zero which is represented by the flat line. The trend in Figure 10 is not surprising if one is to look at the Oriset data where small indications of %IR has been paired to very large coating defects and vice versa. The same scenario is encountered during the construction of Models 1 and 1a. The irregularities we see here can also be explained by the bullet points given in Section 5.1.1.

The most probable cause for this trend is due to the disturbance coming from stray and telluric currents. Most of

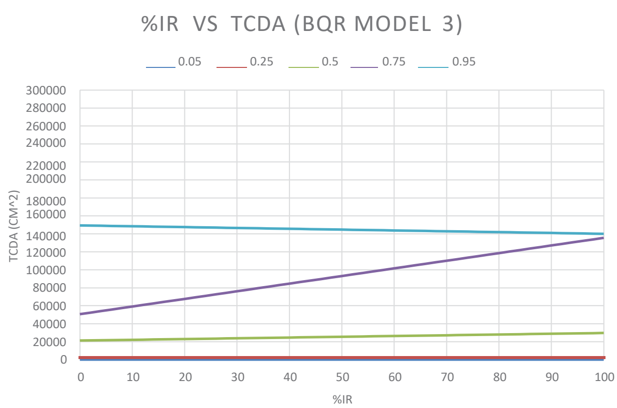


FIGURE 10 %IR versus TCDA for Model 3. Each color represent a different quantile. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

the pipes under assessment were situated within a network of pipelines which runs in parallel and perpendicular with the one that is under investigation. Currents from adjacent CP systems which is protecting other pipelines has the potential of leaving its intended path and is being picked up by the DCVG instrument. Kutz[19] has explained this problem in greater detail.

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

The relationship of %IR and TCDA based on Model 4 is illustrated in Figure 11. The trend in Figure 11 illustrates the general industrial understanding of the relationship between %IR and TCDA. As the quantiles increase, so does the effect of %IR on TCDA which leads to the conclusion of higher %IR affecting larger coating defect areas in a positive way. It can also be said that the sensitivity of the DCVG technique relies on the size of the coating defect. Medium to large defects give a reasonable approximation of the defect size. However, the interpretation based on the %IR on smaller defects should be treated with caution due to large amounts of zero readings present at lower quantiles. As was mentioned earlier, outliers were omitted based on careful judgment. Due to this, Model 4 does not suffer from the problems faced by Model 3, Model 1, and Model 1a where outliers play a role in the estimation of

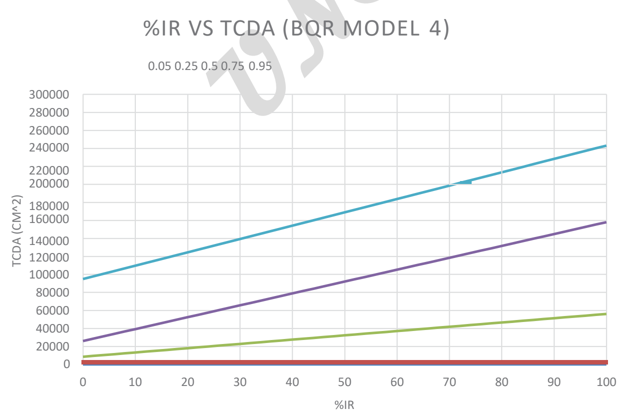


FIGURE 11 %IR versus TCDA for Model 4. Each color represent a different quantile. Reproduced with permission from TWI Ltd. [Color figure can be viewed at wileyonlinelibrary.com]

coefficients. As such, the models are more general and are sufficient for the case of subsequent inspection of the MEOC pipelines.

5.2.2 | POPD variable

Findings from Model 3 indicated that at large coating defect area the possibility of finding deeper corrosion pits are more likely. With larger TCDA, the amount of current provided by the cathodic protection system also should be large. When the level of protection current is inadequate or obstruction of the current's path in the form of a shielding electrolyte is present, one is to expect corrosion activity to be highly likely. However, a dip at quantile 0.75 tells us that at pipelines with medium to large TCDA corresponds to corrosion pits with shallower depths which goes against the normal assumption that a pit's depth is directly proportional to the size of TCDA. At first glance, Model 4 does not exhibit such issues. At the same quantile, the coefficient predicted shows a smooth increase from the median quantile to the largest quantile. Moreover, for Model 4, a consistent upward motion can be seen across the TCDA quantiles. Between the 0.25 and the 0.75 quantile, shows a plateau of estimates suggesting that for these defect sizes, the effect of an increasing POPD is minimal. The increase in values from the 0.05 quantile to the 0.25 quantile can be judged as an initial step toward the corrosion process. At this stage, corrosion is initiated and coating defects grow in tandem. The plateau is an indication that the pit growth rate is faster than the growth of TCDA. This will produce deeper pits at smaller TCDA which solidifies the notion that pit depth is not proportional to the size of coating defect – at least not linearly. This finding was also observed in Ref. [11]. Deeper pits at smaller coating defect should be treated with caution as defects of such characteristics will normally go unnoticed with the consequence of failure being very severe. The effect of direct proportionality between pit depths and coating defect size can be seen between the 0.75 and 0.95 quantile. However, between these quantiles the credible interval increases in wideness indicating a less certain prediction. The OLS prediction is also located in the negative region which means that all the above observation would be missed with the average approach.

5.2.3 | SR variable

Model 4's predicted coefficient quantile trend can be interpreted as highly resistive soil having a large effect on the size of coating defects. Coarse grained soil is known to be highly resistant to electrical current flow hence soils such as sand, silt or even rocks poses high units of SR. These types of soil with its angular particle geometry have the possibility of damaging the pipe coatings through the process of abrasion.

Pipe or soil movement have the possibility of creating abrasion between the coating interface and the electrolyte. Another factor to consider is the stresses created by the self-weight of the backfill.[21,22] The backfill weight applies stresses on to the pipe's coating creating a wrinkling affect normally found at 8 and 4 o'clock position of the pipe. The wrinkling of the coating combined with the abrasion effects of the angular particle size (high SR) will sometime result in the coating tearing apart.

5.3 | Why use Bayesian quantile regression?

The coefficient estimates illustrated by both the Bayesian and classical method in this paper are somewhat similar. Both approaches consider parameter uncertainty with the Bayesian approach being more reliable as it does not rely on asymptotic approximation of the variances. Classical approach such as bootstrapping in the construction of confidence intervals uses estimation of the asymptotic variances and depend on the model error density which is difficult to reliably estimate. Hence, the coverage probabilities of the true parameter of these methods is sufficient at best but not necessarily 100% reliable. This is supported by a paper from by Ref. [23] which shows the classical approach estimated a lower probability of containing the parameter value from the confidence interval as compared to the Bayesian approach. This seems to suggest that a Bayesian method is better in terms of coverage and thus includes all parameter uncertainty. Other advantages of the Bayesian method are that it provides a simple explanation based on the credible interval. For this paper, the credible intervals are set to be 95% and thus the true value of the coefficients can be explained as "having a probability of 0.95 of falling within the credible intervals." For the classical method, the interpretation is not as direct.

Additionally, the BQR method uses the ALD as the likelihood function. Since the likelihood function (ALD) disregard the original distribution of the data, specifying a specific distribution is not needed. The paper[16] goes on to say that the use of the ALD is a "very natural and effective way for modelling Bayesian quantile regression." After the Bayesian process, the resulting posterior statistics such as the mean estimates of the quantiles and the calculated credible intervals can be used as new information for future ECDA. This process is often referred to as Bayesian updating.

In the process of conducting this research, the authors found some drawbacks in employing the Bayesian method. One of them being the problem of convergence. As was seen in the results of Model 4, up to 11 million iterations were needed to achieve convergence. This is due to the nature of the sampling algorithm (Metropolis-Hastings) which uses the accept and reject approach in the goal of achieving convergence at the stationary distribution. Also, there are no known methods to check the convergence of MCMC at this moment.[24] The authors had to rely on the graphical representation of the trace plots which lacks mathematical justification.

6 | CONCLUSION AND FUTURE WORK

This paper has showed that Bayesian techniques on quantile regression is an essential tool for engineers in assessing uncertain data. ECDA pipeline data particularly for the DCVG technique incorporates large amounts of uncertainty due to the unknown factors such as the factors highlighted in Section 5.1.1, the heterogeneity of soils, the levels of CP current, and human factors. As was mentioned earlier, Bayesian techniques allow an assessor to quantify the full spectrum of uncertainty in the prediction of parameters.

In certain countries, the law dictates that an ECDA should be performed on a periodic schedule to ensure the safe continual operation of the pipeline.[25] The NACE SP0502-2010[9] highlights the importance of periodic assessments where "through successive applications of the ECDA method, an operator will be able to identify and address locations of corrosion activity which has occurred, is occurring and at locations where there is a potential to occur." This makes the ECDA a continuous updating process. The Bayesian principle fits this philosophy nicely since updating the findings from this paper is made possible with future ECDA. It is expected that future findings will produce better estimates with every iteration of the ECDA process.

The MEOC data was divided into two for the purpose of investigating the influence of outliers occurring at the upper quantiles of the TCDA distribution. These outliers are thought to be produced from one of the factors highlighted in Section 5.1.1. One of the dataset had a total of four points removed based and the results of the removal can be seen in three of the six models produced namely Models 2, 2a, and 4. Although it is widely known that median regression are robust to outliers, this was not the case for the other three models (Model 1s, 1a, and 3). A dip in the largest quantile for Model 3 with regard to the TCDA variable suggest that it was influenced by outliers. This was not seen in Model 4 (after removal of outliers). Clearly, the quantile regression applied here does not eliminate the problem of outliers entirely. An alternative way to solve this is to construct the model with a non-linear approach. However, with the already established Bayesian approach here, this is not necessary. All that is required is new data and the Bayesian method will update the findings here.

For the estimation of the effect of soil resistivity (SR) on %IR, it is concluded that one must look at both the soil resistivity measurements and the effect of soil grain geometry together. The two variables seem to have some relation and provide a much more holistic picture on the effect its having on the contribution toward %IR. Tests such as the variance inflation factors[26] for multicollinearity effects could be used for future work to see whether the variables are statistically correlated.

As for the case of pit depths (POPD), the rate of growth between the depth of pits and the size of coating defects is not proportional. At some point in time the rate of corrosion is faster which resulted in very deep pits occurring in smaller coating defect area. This is illustrated in Model 4. In this situation, the chances of locating small coating defects is low and hence elevating the risk of failure of the pipeline. It can be said that small coating defects should not be taken lightly especially if the environment for corrosion is highly likely.

Overall, each model represents a unique trait which or which does not agree with established theories. The differences are largely due to the influence of external factors which disrupts the obtained DCVG indications and thus influences the outcome of the analyses. Fortunately, these uncertainties were considered and by continually updating the results through successive iterations of the ECDA, one can only improve the understanding of the state of the pipeline translating into the reduction of operating costs, enhancement of safety and keeping failure risks at bay.

ACKNOWLEDGMENT

The authors would like to thank TWI Ltd. for providing the necessary data for the work presented in this paper.

ORCID

Nik N. Bin Muhd Noor http://orcid.org/0000-0001-9527-2874

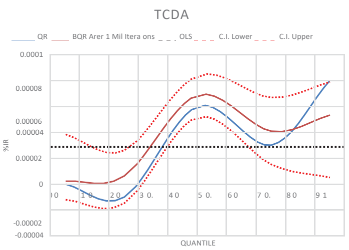
REFERENCES

[1] D. Furchtgott-Roth, *Pipelines Are Safest For Transportation of Oil and Gas*. Manhattan Institute for Policy Research, New York **2013**.
[2] C. Lam, *Master Thesis*, UWO Ontario, Canada, **2015**.
[3] M. Nasir, K. wee Chong, S. Osman, W. S. Khur, *Mater. Sci. Eng.* **2015**, *78*, 012026.
[4] N. Noor, S. R. Othman, L. K. Sing, N. Yahya, M. N. M. A. Napiah, Z. Abdullah, presented at *2nd International Conference on Plant Equipment and Reliability (ICPER 2010)*, Kuala Lumpur, Malaysia, **2010**.
[5] M. Romano, M. Dabiri, A. Kehr, *JPCL* **2005**, *22*, 7.
[6] K. R. Larsen, *MP* **2016**, *55*, 32.
[7] S. Papavinasam, M. Attard, R. W. Revie, *MP* **2006**, *45*, 28.
[8] J. M. Esteban, K. J. Kennelley, M. E. Orazem, R. M. Degerstedt, *Corrosion* **1997**, *53*, 264.
[9] *NACE, SP0502-2010*: Pipeline External Corrosion Direct Assessment Methodology, NACE International, Houston **2010**.
[10] F. Varela, M. Yongjun Tan, M. Forsyth, *CSET* **2015**, *50*, 226.
[11] F. Anes-Arteche, K. Yu, U. Bharadwaj, C. Lee, B. Wang, *Mater. Corros*. **2017**, *68*, 329.
[12] J. P. McKinney, *Master Thesis*, University of Florida, USA, **2006**.
[13] O. C. Moghissi, J. P. McKinney, M. E. Orazem, D. D'Zurko, presented at *Corrosion 2009*, 22–26 March, Atlanta, Georgia, **2009**.
[14] R. Koenker, K. Hallock, *JEP* **2001**, *15*, 43.
[15] L. Hao, D. Q. Naiman, *Quantile Regression*. Sage Publication, Thousand Oaks, CA, USA **2007**.
[16] K. Yu, R. A. Moyeed, *Statist. Probab. Lett*. **2001**, *54*, 437.
[17] *ASTM G187-12a* Standard Test Method for Measurement of Soil Resistivity Using Two Electrode Soil Box Method, American Society for Testing Materials, West Conshohocken **2013**.
[18] K. Yu, J. Zhang, *Commun. Stat. Theory Methods* **2005**, *34*, 1867.
[19] M. Kutz, *Handbook of Environmental Degradation of Materials*. William Andrew, Elsevier, Amsterdam **2005**.
[20] M. Shwehdi, U. Johar, *International conference (ICNIR2003) EMFI*, **2003**, pp. 20–22.
[21] R. Norsworthy, *Corrosion* 2009, March 22nd, Atlanta, Georgia, **2009**.
[22] B. C. Yen, G. D. Tofani, *OGJ* **1985**, *83*, 34.
[23] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, Boston **2011**.
[24] K. Yu, P. Van Kerm, J. Zhang, *Sankhya Ser. A* **2005**, *67*, 359.
[25] *U. S. DOT 49 CFR 192.925* – What are the requirements for using External Corrosion Direct Assessment (ECDA)? *U. S. Department of Transportation*, New Jersey **2011**.
[26] J. Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman, *Applied Linear Statistical Models*. McGraw-Hill/Irwin, Chicago, USA **2004**.

How to cite this article: Bin Muhd Noor NN, Yu K, Bharadwaj U, Gan T-H. Making use of external corrosion defect assessment (ECDA) data to predict DCVG %IR drop and coating defect area. *Materials and Corrosion*. 2018;1–20. https://doi.org/10.1002/maco.201810085

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only. It will not be published as part of main.



This paper highlights the results obtained from the external corrosion direct assessment (ECDA) process which was conducted on 250 km of buried pipelines. The results from the indirect and direct assessment part of the ECDA were modeled using the classical quantile regression (QR) and the Bayesian quantile regression (BQR) method to investigate the effect of factors toward the IR drop and the coating defect size.