

Article

Predictive Maintenance Framework for Cathodic Protection Systems Using Data Analytics

Estelle Rossouw ^{1,*} and Wesley Doorsamy ² 

¹ Postgraduate School of Engineering Management, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa

² Institute for Intelligent Systems, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa; wdoorsamy@uj.ac.za

* Correspondence: 200575264@student.uj.ac.za

Abstract: In the quest to achieve sustainable pipeline operations and improve pipeline safety, effective corrosion control and improved maintenance paradigms are required. For underground pipelines, external corrosion prevention mechanisms include either a pipeline coating or impressed current cathodic protection (ICCP). For extensive pipeline networks, time-based preventative maintenance of ICCP units can degrade the CP system's integrity between maintenance intervals since it can result in an undetected loss of CP (forced corrosion) or excessive supply of CP (pipeline wrapping disbondment). A conformance evaluation determines the CP system effectiveness to the CP pipe potentials criteria in the NACE SP0169-2013 CP standard for steel pipelines (as per intervals specified in the 49 CFR Part 192 statute). This paper presents a predictive maintenance framework based on the core function of the ICCP system (i.e., regulating the CP pipe potential according to the NACE SP0169-2013 operating window). The framework includes modeling and predicting the ICCP unit and the downstream test post (TP) state using historical CP data and machine learning techniques (regression and classification). The results are discussed for ICCP units operating either at steady state or with stray currents. This paper also presents a method to estimate the downstream TP's CP pipe potential based on the multiple linear regression coefficients for the supplying ICCP unit. A maintenance matrix is presented to remedy the defined ICCP unit states, and the maintenance time suggestion is evaluated using survival analysis, cycle times, and time-series trend analysis.

Keywords: cathodic protection; corrosion monitoring; data analysis; machine learning algorithms; pipelines; predictive maintenance; predictive models



Citation: Rossouw, E.; Doorsamy, W. Predictive Maintenance Framework for Cathodic Protection Systems Using Data Analytics. *Energies* **2021**, *14*, 5805. <https://doi.org/10.3390/en14185805>

Academic Editors: Mauro D'Arco and Francesco Bonaventura

Received: 6 July 2021

Accepted: 20 August 2021

Published: 14 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pipeline networks supply piped products to consumers that span over large geographic areas. Typical products include bulk and domestic water, oil, and gas. Underground pipelines enable the transfer of these products between the source and destination [1]. Pipeline networks often run through residential and industrial areas and remote areas such as agricultural properties [2]. The Charleston Advisor estimated in 2013 that South Africa had a total installed pipeline network of 3869 km and was expected at that stage to increase with economic activity [3]. However, the extension of pipeline networks presents significant safety and financial risk due to the corrosion process that starts when a pipeline is buried in the ground [4].

Corrosion is a natural phenomenon since a metal interacts with its environment and results in a metal loss. The corrosion process is related to the Gibbs energy theory, which states that metals will continuously try to achieve a lower oxide state by reducing their energy [5].

A study conducted by the National Association of Corrosion Engineers (NACE) estimated that the costs of corrosion amounted to USD 2.5 trillion in 2013, which was 3.4% of the global domestic product (GDP) [4]. Corrosion costs include product loss, plant

shutdowns, replacement of pipeline sections, and increased maintenance costs. NACE further suggests that an efficient corrosion management system (CMS) can reduce the cost of corrosion by 15–35% [6]. NACE suggests three methods to mitigate corrosion: change the environment, change the material, or place a barrier between the environment and material [7]. The last method refers to a protective coating or backfilling the pipe trench [8,9]. A secondary external corrosion prevention mechanism applicable to underground pipelines is impressed current CP (ICCP) rectifiers that supply direct current (DC) for corrosion control. The CP current is supplied to an anode ground bed, which forces all anodic areas on the pipeline to cathodic areas. This process is called polarization and results in corrosion of the ground bed instead of the pipeline [10]. Equilibrium is reached when all pipeline areas are cathodic, and at this stage, corrosion ceases [7].

A pipeline section typically consists of one or more ICCP units, with downstream measurement stations, also referred to as test posts (TPs), typically less than 1.6 km apart [11]. The pipe-to-soil DC potential, or CP pipe potential, is a critical conformance measurement to determine if the CP is sufficient and is a statutory periodic requirement required for ICCP units and TPs [11,12]. The NACE SP0169-2013 standard provides three criteria to evaluate a CP system's effectiveness and is primarily concerned with the magnitude and polarity of the CP pipe potential [12]. For this study, the CP voltage threshold of the NACE instant-on CP criteria was selected as the maximum operating limit for the CP pipe potential, while the minimum operating limit was selected based on the operating conditions for CP systems in South Africa.

This paper considers two types of ICCP units: a transformer–rectifier unit (TRU) or a forced drainage unit (FDU). The latter is typically installed close to a DC transit system to drain stray current back to the rail that strayed onto the pipeline [13,14]. An FDU also presents an increased corrosion risk should it malfunction since a stray current can accelerate corrosion [12]. The 49 CFR Part 192 statute stipulates periodic inspection intervals for different CP equipment [11].

Apart from the absolute Code of Federal Regulations (CFR) inspection interval defined, pipeline operators can also consider additional periodic inspections of ICCP units and a reactive maintenance strategy to reinstate a failed ICCP unit. This strategy is economically sustainable; however, the maintenance cost can increase rapidly for extensive, cross-country pipeline networks, and a higher risk exists that the CP system is not functioning correctly. Historical data analysis can identify trends and aid in decision-making [12]. NACE suggests that historical CP performance can also be considered when analyzing the CP system state, and thus, the design of this paper includes both historical CP data analysis and the use of predictive modeling techniques. The aim is to use the CP pipe potential as the driving factor for the predictive modeling and maintenance suggestion.

The framework proposed in this paper utilizes the fundamental concepts of predictive maintenance principles for health and prognostic prediction (typically found in condition-monitoring systems) and reliability engineering principles (risk management and maintenance strategies). The presented framework does not predict equipment failures but rather the required maintenance based on the conformance of the CP pipe potential to the NACE SP0169-2013 CP criteria and the ICCP unit's state (relevant to the CP pipe potential).

The following section evaluates the key literature applicable to this paper. After that, the methodology, proposed framework, and study results are presented. A summary of the work is then given, with recommendations and future work.

2. Literature Review

2.1. Corrosion Basics

A basic corrosion cell consists of an electrically connected anode and cathode placed in a conductive electrolyte. If all mentioned elements are present, the corrosion process starts. At the anode, a loss of electrons occurs (oxidation), and consumption of electrons occurs at the cathode (reduction). This process results in a potential difference between the

anode and the cathode due to current flow [15]. The electrolyte can be water (freshwater or seawater) or soil, and its conductivity affects a metal's corrosion rate [16].

Anode and cathode potentials are not equal, and the practical galvanic series lists metals in descending order based on their native DC potential. The metal with the lower DC potential is the anode, while the metal with the higher DC potential is the cathode. The magnitude of the potential difference between the anode and cathode is the driving force for corrosion. Callister refers to metals placed in an ion solution as “electrodes”, each with its own DC potential [5].

2.2. Cathodic Protection

External corrosion prevention can include environment-specific inhibitors, a protective barrier such as a pipeline coating, or CP [5]. This paper only concerns the last, consisting of a sacrificial anode (SACP) or an impressed current CP system (ICCP). The theory suggests that the anode will corrode to protect the pipeline and the primary difference between the SACP and ICCP systems is the amount of current produced by each system. The ICCP system uses an external rectifier to supply the driving current to the anode ground bed [5]. The rectifier supplies a direct current in the opposite direction of the corrosion current, this yielding a zero-net current flow, and corrosion ceases [5].

The CFR pipeline safety statute stipulates that CP potentials must be recorded at periodic intervals at test stations or TPs to evaluate the CP system's effectiveness [11]. The NACE SP0169-2013 standard for steel pipelines presents three criteria to evaluate the effectiveness of a CP system, namely an instant-on CP pipe potential less than $-850 \text{ mV}_{\text{CSE}}$ with CP applied and considering any voltage drops, an instant-off CP pipe potential less than $-850 \text{ mV}_{\text{CSE}}$ without CP applied, and $100 \text{ mV}_{\text{CSE}}$ cathodic polarization [12]. The instant-on criterion was selected for this paper based on the historical CP data received. The CP data, however, did not include any known voltage drop (IR-drop).

Various CP equipment exists in the industry, such as AC mitigation stations, cross-bonds, sacrificial anodes, isolation joints, and natural drainage units [9,13]. The selection of this equipment depends on the application and CP system design. This paper focuses only on the ICCP unit and the downstream TPs due to the complexity of CP systems.

Stray current refers to any electrical current that does not follow an intended path and poses a significant corrosion risk to pipelines. Stray current sources can include overhead AC powerlines, foreign pipelines, DC transit systems, or telluric currents. Stray currents can cause high-magnitude spikes (up or down) in the CP pipe potential [17,18]. Zebang et al. suggest that geomagnetic storms can also induce fluctuation of the CP pipe potential on buried pipelines [19].

A typical CP system is illustrated in Figure 1, indicating the ICCP units and TPs.

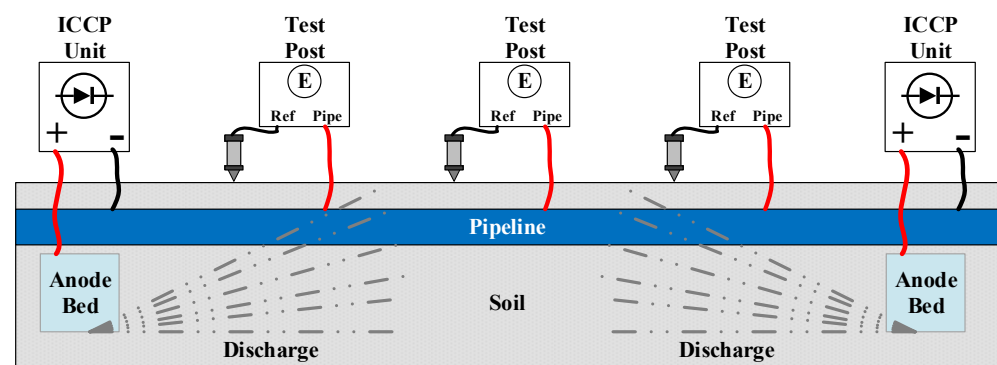


Figure 1. Typical pipeline CP system. Source: adapted from [20].

2.3. Corrosion Monitoring

A reference electrode (RE) is a device that provides a stable Galvani potential and consists of a porous conductive plug connected to a metal rod submerged in an ion

solution [21]. The native RE potential is determined by the metal-ion solution [22]. A saturated copper–copper sulfate (CSE) RE is frequently used for CP field measurements and is calibrated against a standard hydrogen electrode (SHE) [12].

The CP pipe potential is measured between the pipe and a RE with a DC voltmeter. The RE is typically a CSE for underground pipelines, and the DC voltage reading is denoted as “ V_{CSE} ” [19].

In industry, corrosion has been predominantly detected by recording the CP pipe potential, inline pipeline monitoring, and coating defects [12], and the use of more advanced technology is emerging with the advent of Industry 4.0. Tamhane et al. suggested using lead zirconate titanate (PZT) transducers for monitoring structural health, and corrosive changes are detected by a change in the electromechanical resonant frequency [23].

2.4. Remote Monitoring

Supervisory control and data acquisition (SCADA) systems collect data from various sensors and enable remote monitoring of CP systems [24]. Process data are stored in relational databases for historical data analysis and can highlight process inefficiencies and lead to process optimization [25]. Remote monitoring of CP systems enables a buildup of CP performance data [12] and is also the primary data source for this paper.

Abate et al. suggested the use of a networked control system for CP rectifiers using M-bus networks. Using connected CP equipment enables fuzzy logic control between CP rectifiers and recording and monitoring CP pipe potentials [26].

2.5. Maintenance Strategies

Abundant literature exists that evaluates reliability engineering principles potentially applicable to this paper. Maintenance strategies can include preventative maintenance, reactive maintenance, time-based maintenance, risk-based maintenance, reliability-centered maintenance, or condition-based maintenance (CBM) or predictive maintenance (PdM) [17]. The last applies to this paper, and the Open Standards for Physical Asset Management (OSA-CBM) framework, guided by various ISO standards, describes various functional models for developing a condition-monitoring (CM) system [27].

CM systems are primarily concerned with diagnostics and prognostics, the former referring to a machine state and the latter to the progression of existing and future faults [18]. Process data received from SCADA systems can be used in CM systems to enable alarms and events, determine the system health, and perform a prognostic assessment on the equipment’s remaining useful life (RUL). A CM system can also recommend actions to assist with the decision-making process [28].

Prognostics consists of three common approaches: a data-driven approach, a model-driven approach, or a hybrid approach. The data-driven approach is concerned solely with sensor data, while the model-driven approach depends on a mathematical model of the equipment. A hybrid approach is a combination of the data and model approaches [29]. This paper adopts a data-driven approach since no CP design data were used for predictive modeling.

2.6. Data Analytics

A typical data analytics process consists of data preparation (collection, cleaning, and feature selection), preprocessing (cleaning, transformation, and standardization), analysis (visualization, regression, correlation, and forecasting), and postprocessing (documentation, forecasting, and evaluation) [25].

Data analytics evolved from mere statistics to complex machine learning (ML) and artificial intelligence (AI) systems. ML teaches a computer how to process data and enables outcome prediction. Furthermore, ML asks critical questions about the learning process (how and why) and continuously improves prediction accuracy. AI is primarily concerned with robotics and has expanded to industries such as process automation, sports analytics, and manufacturing [30].

Data analytics depend on probabilistic methods and enables decision-making where uncertainty exists [31]. Big data analytics enable knowledge discovery in databases (KDD) by extracting and analyzing data from large databases [25]. Feature engineering of datasets can include linear analysis of data (LAD) [32].

2.7. Machine Learning

ML consists of an outcome to predict based on the features included in the model design. Training and test datasets are usually created (at a specific ratio based on the application). The model will be trained using the known relationships in the training set and then evaluated against the test set [33]. For a basic model, the model's accuracy can be the sum of the correct predictions (as a percentage of the total number of predictions). Centered at the core of ML techniques are probabilistic methods, which describe an event's probability of occurring under a series of circumstances [31].

One key aspect of ML is the ability to learn automatically and improve without any user intervention. The data made available for ML play a significant role in which technique to use. Data can be structured, unstructured, semistructured, or metadata. Structured and unstructured data are usually saved in a relational database, whereas semi-structured or metadata can be access via JSON, XML, or HTML [34].

ML techniques include clustering, regression, classification, and association rule learning [35]. The data structure and required outcome will dictate which technique to use. Applicable to PdM systems are supervised and unsupervised learning. The latter predicts new cases from existing labeled cases [36] and can consist of techniques such as linear regression (LR), Naïve Bayes, support vector machines (SVMs), logistic regression, neural networks (NNs), and random forest (RF) [37]. Unsupervised learning aims to identify new patterns within data using bagging and boosting algorithms [36].

2.8. Predictive Modeling

Predictive modeling is concerned with defining a mathematical equation to make accurate predictions using ML techniques. The model building process includes data transformation, data exploration, feature engineering, data cleaning, identifying predictors, estimating performance with known quantitative statistics, evaluating different models, and selecting the most appropriate model [37].

Performance evaluation is a critical step in the ML model evaluation and ultimately informs selecting the best performing model. The mean absolute error (MAE) and root-mean-square error (RMSE) are two suggested metrics for evaluating the performance of a linear regression model, while the percentage error between the predicted and actual values is helpful to evaluate classification models [37]. The RMSE equation is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^n e_i^2} \quad (1)$$

2.9. Previous Work

Predictive modeling of CP systems is not a broad research topic, and hence related academic literature is limited. Various literature exists on PdM frameworks for different applications, and this paper employs some of the techniques suggested in the existing literature.

3. Methodology

3.1. Context

This paper's context involves modeling two pipeline sections based on the supplying ICCP unit type (FDU or TRU). The two pipeline sections enable analysis based on the unit type and associated risk. The equipment typically found as part of a CP system is shown in Figure 2.

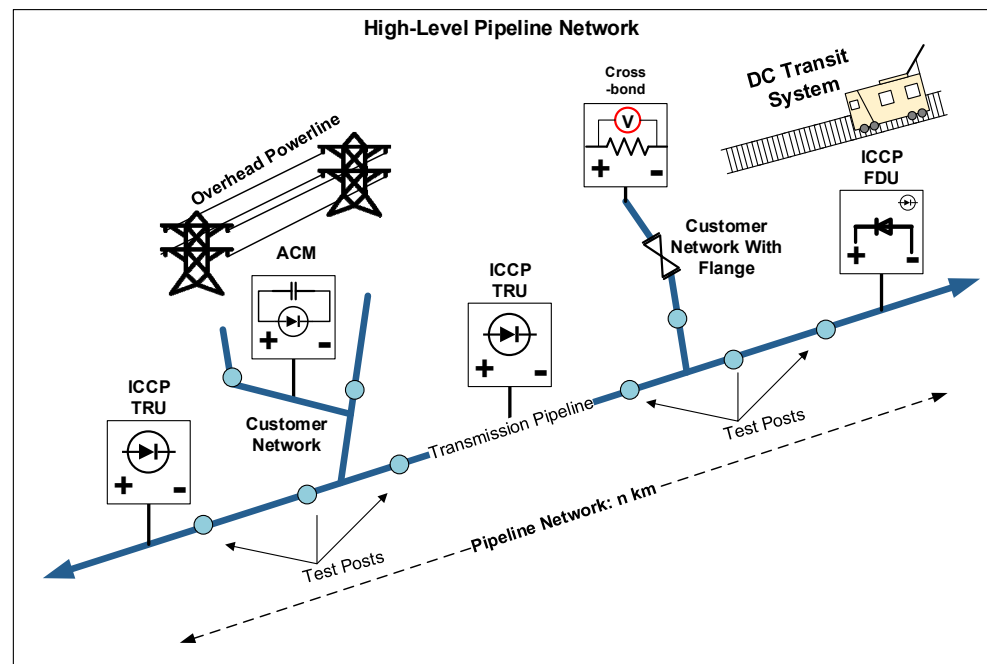


Figure 2. Typical CP system elements. Source: adapted from [12,38–40].

3.2. Historical Data

3.2.1. Data Collection Method

Instant-on CP pipe potentials were collected for this study. The structured data used for the ML modeling consists of CP operating data collected from the following:

1. A CP-SCADA system that receives data recorded from ICCP rectifiers using a combination of communication technologies, such as GPRS, LoRaWAN, or industrial radio networks. The SCADA server is either hosted in the Microsoft Azure cloud or installed on-premise. The following data are historized in the SCADA system from each ICCP rectifier:
 - Rectifier output voltage (V_{OUT});
 - Rectifier output current (I_{OUT});
 - Rectifier drainage current (I_{DRAIN});
 - CP pipe potential (V_{CSE}).
2. A remote data logger database that contains continuous CP pipe potentials at TPs (transmitted via General Packet Radio Service (GPRS) to a cloud-hosted database server).
3. A database of manual CP recordings taken with specialized CP data loggers at TPs.
4. Any related operational and maintenance documentation to provide additional information for the ML model setup (such as pipeline asset information and operating standards).

Figure 3 depicts a high-level communication overview. The private access-point-name (APN) allows for an additional layer of data security because a dedicated username and password are required. Some routers allow for an additional layer of security using layer two tunneling protocol (L2TP) [41].

Since this paper focuses on using data for decision-making, no CP design data were used for the framework presented.

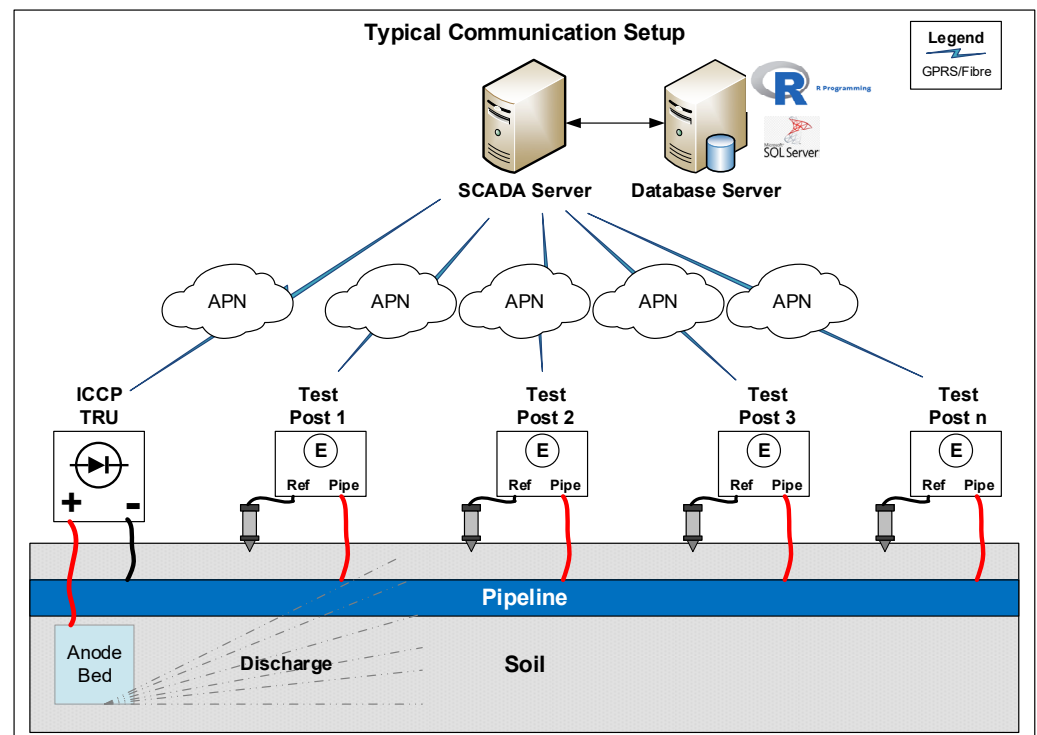


Figure 3. Typical CP remote monitoring architecture. Source: adapted from [42].

3.2.2. Metrology

The NACE TM0497-2018 standard provides the guidelines for the instrumentation specifications, maintenance procedures, and the proposed polarity for performing CP pipe potential (V_{CSE}) measurements. All the measurements collected for this paper consisted of instruments with a minimum input impedance of 10 M Ω (except for shunt readings). Table 1 is a summary of the measurement strategy used for this paper:

Table 1. Instrument Overview.

Instrument Overview					
Variable	Description	Range	Measurement Type	Input Impedance	Measurement Interval/Strategy
V_{CSE}	Instant-on pipe-to-soil potential	± 25 VDC	Direct	10 M Ω	5 min or deadband
I_{OUT}	Rectifier output current	0–100 A	Current shunt/CT	Highest possible	5 min or deadband
V_{OUT}	Rectifier output voltage	0–150 VDC	Direct	10 M Ω	5 min or deadband
I_{DRAIN}	Rectifier drainage current	0–100 A	Current shunt/CT	Highest possible	5 min or deadband

The deadband measurement strategy is based on statistical process control (SPC) concepts [43], whereby a value is only reported to the SCADA if the variable being measured exceeds the predefined control limits. This strategy reduces the number of data transmitted, analyzed, and stored (which can significantly impact the system's operational expenditure (OPEX) cost).

The data loggers utilized in this paper are powered by either an external battery pack or an embedded lithium-ion battery. Data loggers with external battery packs transmit logged data to a central server using GPRS, while manual loggers send data to a database using a USB interface.

3.3. Data Analysis Framework

Figure 4 illustrates the data analysis framework for this paper.

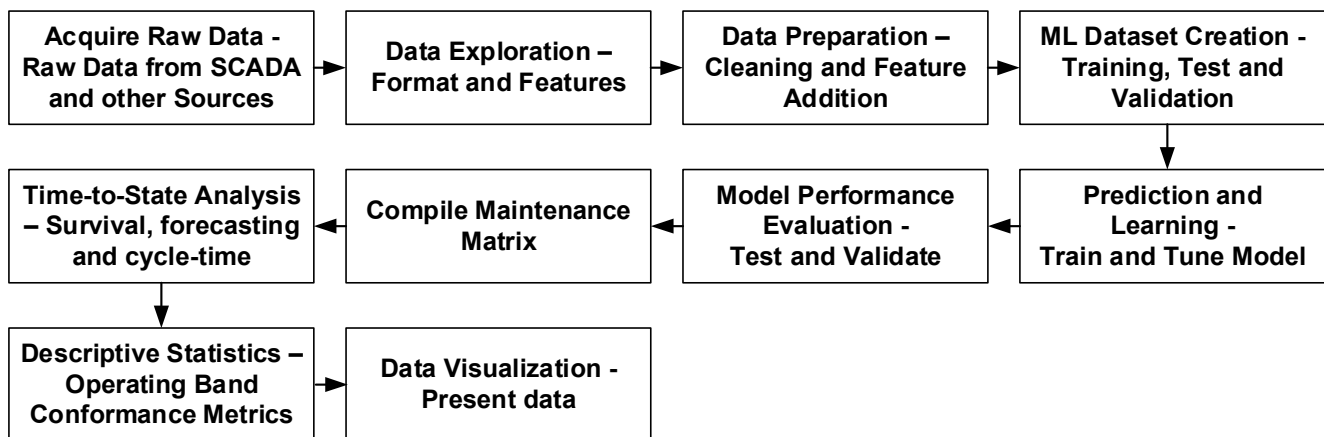


Figure 4. Data analysis framework.

3.3.1. Data Acquisition

The data points collected include the rectifier output voltage, current, and drainage current and the CP pipe potential (V_{CSE}) at both the TPs and ICCP units. All CP pipe potentials (V_{CSE}) are instant-on and do not consider the IR-drop.

3.3.2. Data Exploration

Data exploration considers an in-depth evaluation of the data received from the SCADA system and logger databases to determine relationships between variables, evaluate distributions, and visually inspect the CP pipe potential based on the defined operating window (OW). The authors used the R programming language extensively for modeling and prediction.

3.3.3. Data Preparation

Incomplete columns and rows are removed in this step, and column data types are altered based on the analysis requirements. Additional columns are created using logic combinations. The primary columns include the status column, which assigns a state label based on the CP pipe potential's conformance to the NACE criteria. The OW defined for this paper is an instant-on CP pipe potential between $-5.00 V_{CSE}$ and $-0.85 V_{CSE}$.

The state labels include the following: P, protected (V_{CSE} within OW); UP, underprotected (V_{CSE} more electropositive than OW); and OP, overprotected (V_{CSE} more electronegative than OW). A numeric risk value between 1 and 4 is assigned to a risk column based on the CP pipe potential (V_{CSE}) magnitude within each of the three states.

Two conditions drive a column reflecting the rectifier operating status: whether the status is "P" and the rectifier is supplying current for a specific time limit. A numeric unit type column identifies the CP equipment, 1 for TP, 2 for TRU, and 3 for an FDU. The ability to detect a stray current or a significant voltage shift requires a column that evaluates the absolute numerical difference between the CP pipe potential of two sequential rows. The OP, UP, and P event times and cumulative event time are required and computed per line for event time analysis.

3.3.4. Dataset Creation

Two datasets are required for learning and predicting, namely a training set and a test set. These two datasets are created from the original CP dataset, and different ratios were used for the two datasets based on the computational overhead and the size of the original

CP dataset. The optimal dataset split was 75:25% for training and test sets. The number of data points varies between 82,000 and 95,000 based on the different simulations.

3.3.5. Learning and Prediction

Learning and prediction are iterative processes to evaluate different ML models based on the selected technique and predictor combinations [44]. The caret package in R was selected for learning and prediction in this paper.

3.3.6. Model Tuning

The kNN cross-validation was implemented to evaluate the model fit (i.e., either overfitting or underfitting). The train control parameter was selected as 10, and the optimal grid-tuning parameter was determined by iterating through a sequence from 1 to 71 in increments of 2.

3.3.7. Model Performance Evaluation

The RMSE (1) and percentage accuracy describe the prediction error or accuracy for linear regression and classification models.

3.3.8. Maintenance Matrix

A maintenance matrix is presented in this paper that suggests maintenance activities based on the state (OP, UP, P), risk, and maximum allowable time (also referred to as the cycle time or time limit).

3.3.9. Time-To-State Analysis

The time-to-state analysis is concerned with the time to a specific event and is the CP pipe potential state (OP, UP, P) in the context of this paper. Three methods are evaluated: survival analysis based on the Kaplan–Meier survival curve, cycle times, and time-series forecasting at different frequencies.

3.3.10. Descriptive Statistics

Descriptive statistics present the CP pipe potential conformance as a time or percentage statistic over a designated reporting interval.

4. Exploratory Data Analysis

4.1. CP Pipe Potential Evaluation

The CP pipe potential and OW are shown on each line graph.

4.1.1. Regulating within Operating Window

Figure 5 indicates that the CP pipe potential is regulated within the OW. The resultant assigned state is “P”.

4.1.2. Overprotection

Figure 6 indicates that the CP pipe potential operates below the minimum limit of the defined OW (selected as $-5.0 V_{CSE}$). The resultant assigned state is “OP” and can result in the disbondment of the pipeline wrapping if active over prolonged periods.

4.1.3. Underprotection

Figure 7 indicates that the CP pipe potential operates above the defined OW’s maximum limit (the NACE $-0.85 V_{CSE}$ guideline). The resultant assigned state is “UP” and can result in forced corrosion over time.

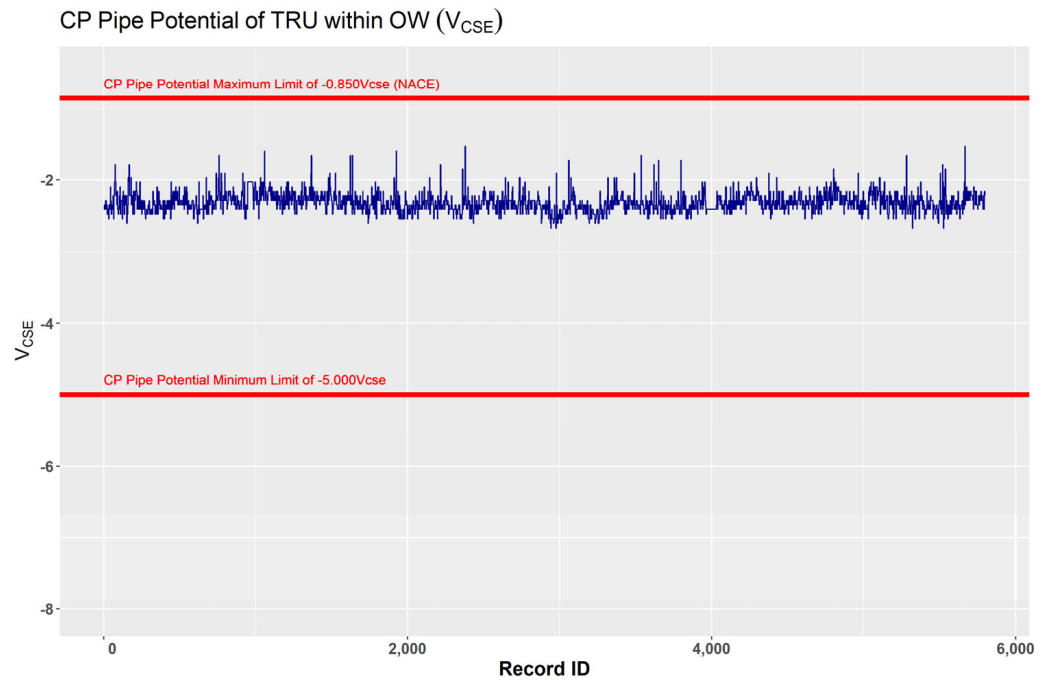


Figure 5. CP pipe potential within operating window.

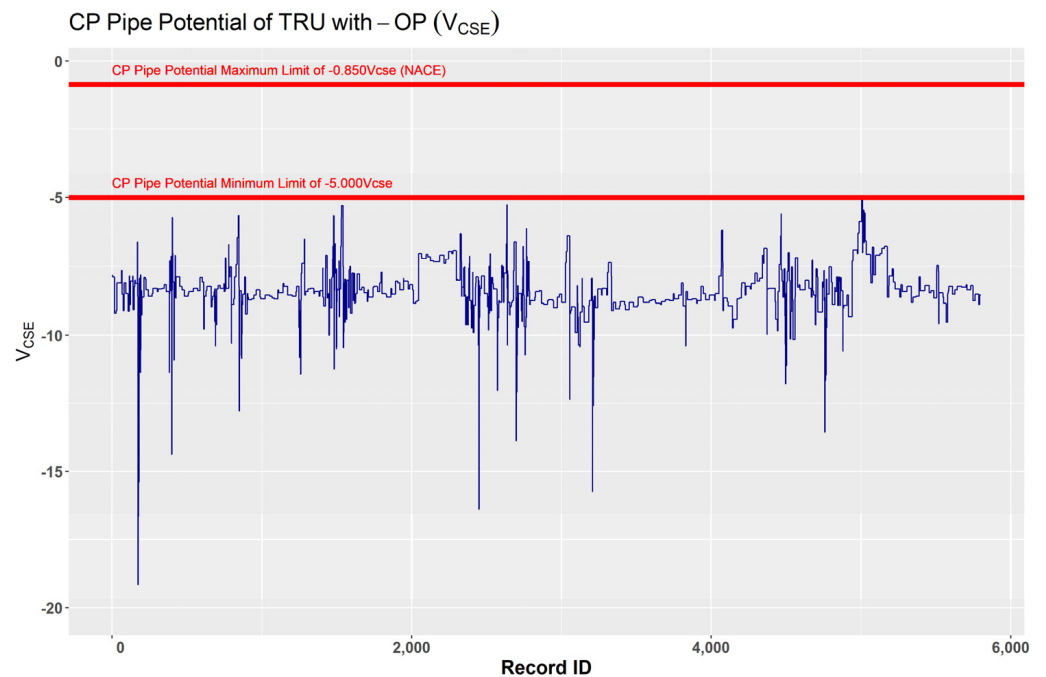


Figure 6. CP pipe potential below operating window.

4.1.4. Stray Current

Figure 8 illustrates high-magnitude spikes in the CP pipe potential, which indicate stray current. The stray current event will be flagged if the numerical difference between two sequential rows exceeds the selected stray current setpoint.

The stray current can cause the CP pipe potential to vary between the OP and UP setpoints, and further trend analysis is required to determine the actual CP pipe potential by filtering out the “noise” present in the signal.

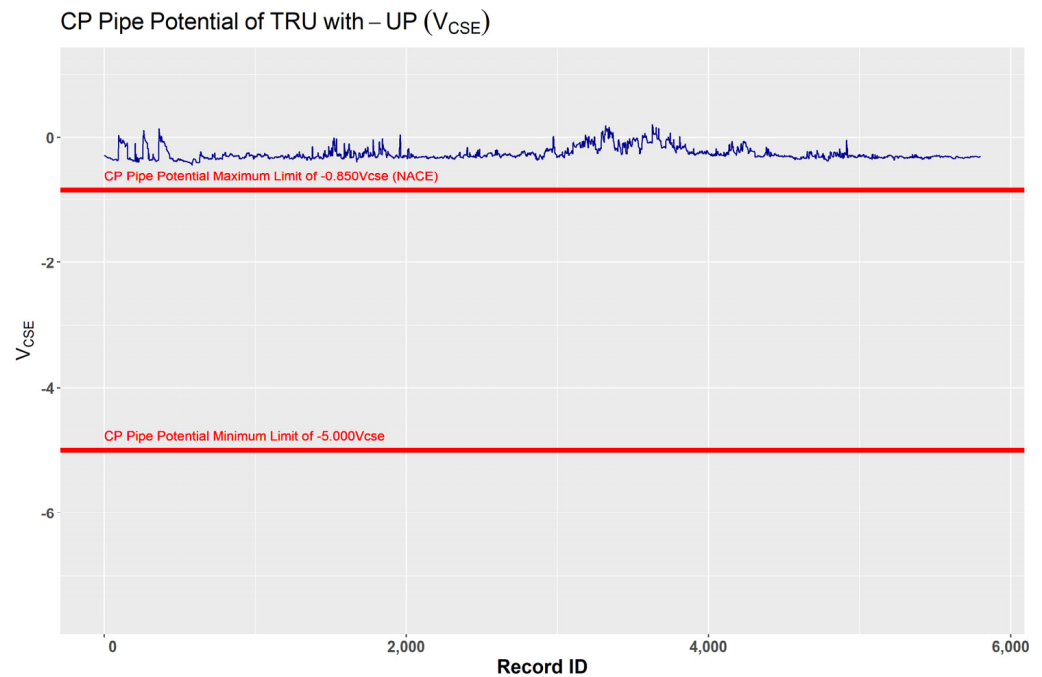


Figure 7. CP pipe potential above operating window.

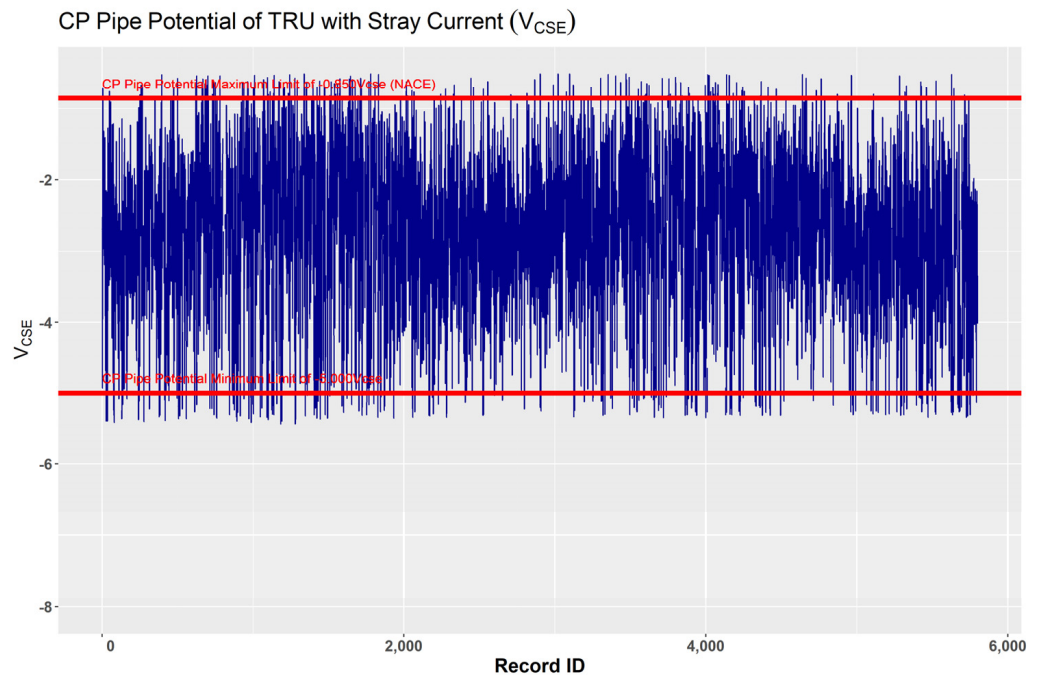


Figure 8. CP Pipe Potential with Stray Current.

4.2. Raw Data from an FDU

The raw data from an FDU with measurement points rectifier voltage, current, and drainage current and the CP pipe potential is plotted to evaluate the change in rectifier operation with the presence of stray current (caused by a DC transit system).

The line graphs in Figure 9 show that the output voltage initially decreases if the output current and drainage current increase. After a short period, the output voltage significantly spikes up, while the output current decreases sharply. The change in the rectifier output magnitude signifies an attempt to regulate the CP pipe potential within a specific control band or at a specific setpoint when a stray current is present.

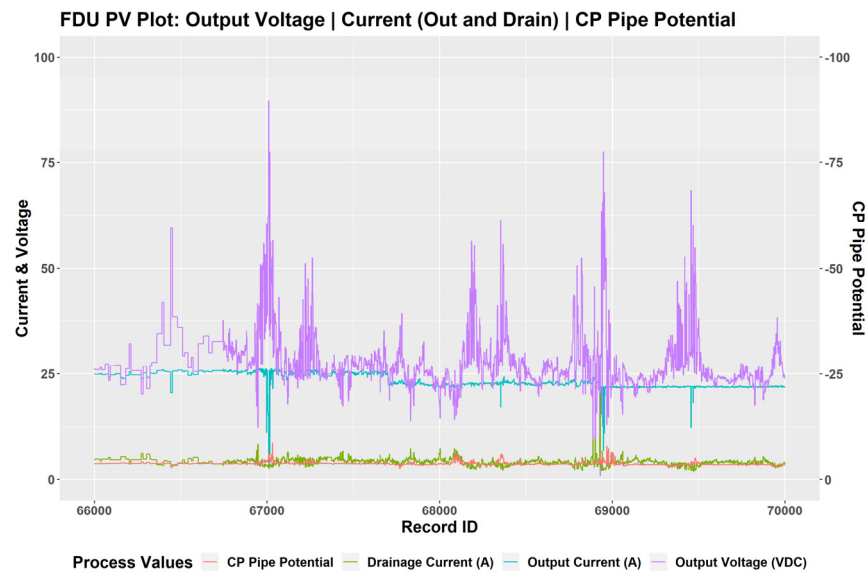


Figure 9. ICCP unit variables (FDU with stray current).

4.3. ICCP Unit Variable Correlation

Variable correlation signifies the relationship between two variables as a value between -1 and $+1$ [37], and the correlation results for a TRU in Figure 10 illustrate the change in correlation once stray current is present.

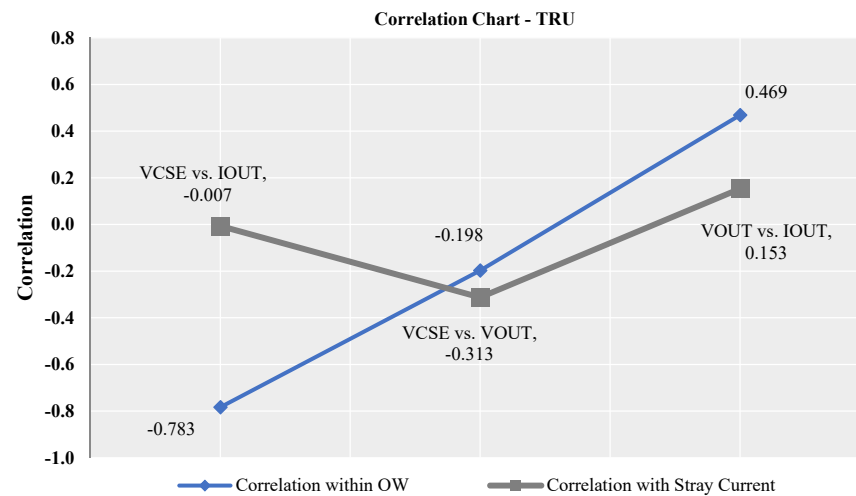


Figure 10. TRU correlation result comparison.

The correlation results for an FDU presented the same findings as for the TRU, i.e., change in correlation if the stray current is present. The change in correlation indicates that predictive modeling with stray currents can potentially affect the prediction approach and accuracy.

4.4. Downstream Effect of Dynamic Stray Current at TPs

A pipeline section was evaluated to determine the effect of stray current at downstream TPs. The pipeline section consisted of one FDU and four TPs. The TPs are 1, 2, 5, and 8 km from the FDU. The effects of the stray current (for this specific pipeline section) decay along the pipeline based on the magnitude of the CP pipe potential measured at the FDU and each TP. However, the CP pipe potential waveform repeats at the TPs, as illustrated in Figure 11.

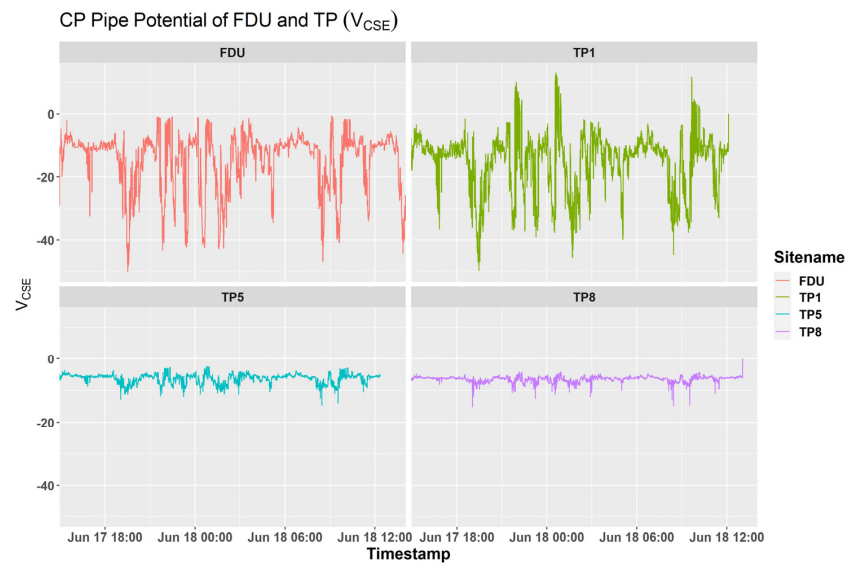


Figure 11. Downstream effect of stray current.

4.5. Trend Component Analysis

The line graphs in Figure 11 are the noisy CP pipe potentials, and plotting the trend component of a time-series object (Figure 12) enables visual evaluation of the CP pipe potential trend for the specific time window. Trend decomposition was enabled using the forecast package in R:

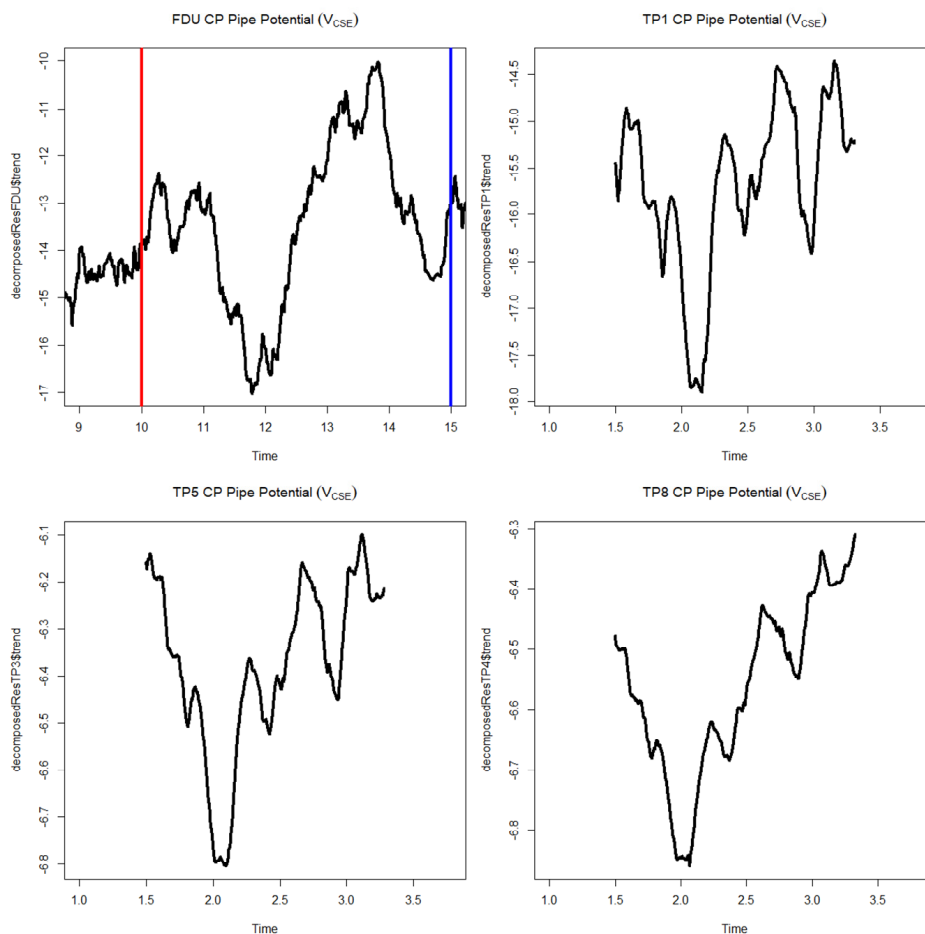


Figure 12. Downstream effect of stray current using time-series trend component.

5. Predictive Modeling Results

The predictive modeling results are discussed in the sections following.

5.1. Variables

This paper uses the variables defined in Table 2 for predictive modeling.

Table 2. Predictive modeling variables.

Variable Names for ML Models		
Variable	Description	Units
V _{CSE}	Instant-on pipe-to-soil potential	V _{CSE}
I _{OUT}	Rectifier output current	A
V _{OUT}	Rectifier output voltage	V
I _{DRAIN}	Rectifier drainage current	A

5.2. ML Datasets

Training and test datasets were created from the original ICCP unit and TP raw data sets with a 70%:30% ratio for training and test datasets.

5.3. CP Pipe Potential Prediction Using Linear Regression

The first section of the predictive modeling process aims to predict the CP pipe potential of a TRU operating at a steady state, a malfunctioning FDU, and an FDU operating with a stray current. Multiple linear regression techniques were selected from the caret package in R to predict the CP pipe potential and inform the selection of the most accurate model (based on the model RMSE results). The predictors consisted of V_{Out} + I_{Out} for the TRU and V_{Out} + I_{Out} + I_{drain} for the FDU.

5.3.1. TRU Operating at Steady State

Figure 13 presents the RMSE results for the different ML models evaluated, and the untuned random forest (RF) model presented the best prediction accuracy (RMSE of 0.153).

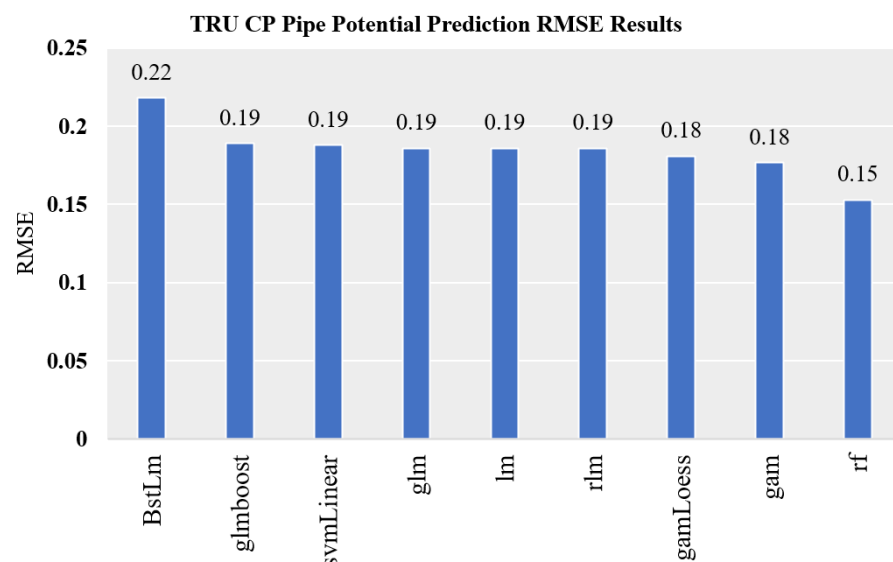


Figure 13. TRU CP pipe potential prediction results (RMSE).

5.3.2. FDU (Malfunction)

A multiple linear regression approach for a malfunctioning FDU presented an RMSE of 26.95, which is unacceptably high to predict the CP pipe potential.

5.3.3. FDU (Stray Current)

Analysis of an FDU with stray current considered using ICCP data at different sampling rates and time intervals. The results from Figure 14 suggest an RMSE improvement from 2.06 to 0.68 since the noise is eliminated with a decreased instrument polling rate from 30 seconds to two minutes. Furthermore, the data period was also extended to 3 months for training and testing.

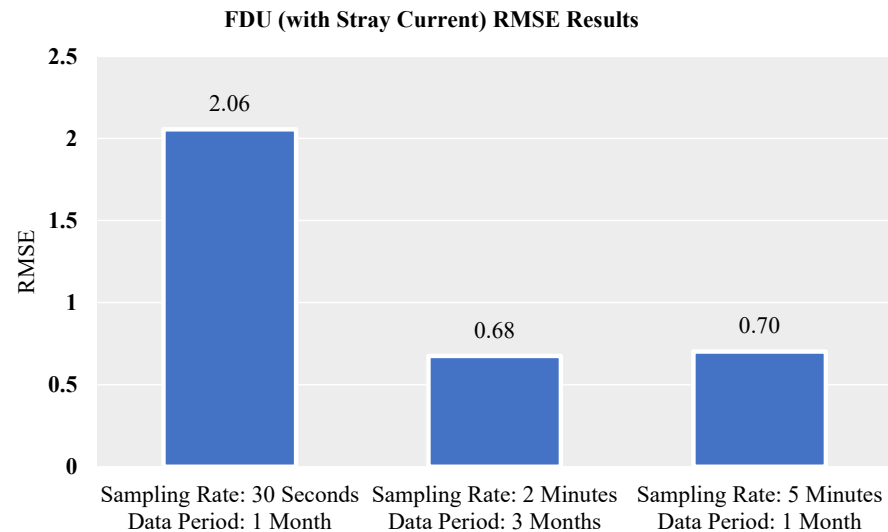


Figure 14. FDU prediction results (stray current) at multiple time components.

By selecting an optimal tuning parameter for kNN, the RMSE error was reduced to 1.898557 by selecting a grid-tuning parameter of 11, which is an improvement that can be implemented.

5.3.4. Downstream Test Post CP Pipe Potential Estimation

The multiple linear regression coefficients are required for the supplying ICCP unit to estimate the CP pipe potential of downstream TPs. The following coefficients were obtained for the supplying ICCP unit on a 19 km pipeline section:

$$V_{CSE} = -4.082 - 1.703I_{OUT} + 0.382V_{OUT} \quad (2)$$

Since the output current of the rectifier predominantly shifts the CP pipe potential up or down (where no stray current exists), the downstream TP CP pipe potential can be estimated by substituting a unique output current coefficient for each downstream TP in (2). Each TP's unique output current coefficients were calculated using historical CP pipe potentials and manipulating (2).

The results presented a mean CP potential difference between the true and estimated TP CP pipe potentials of 0.0014 V_{CSE} for the pipeline section. Figure 15 illustrates the estimated versus actual CP pipe potentials.

5.3.5. CP Pipe Potential State Prediction Using Classification

This paper also considered a classification ML approach to improve prediction accuracy and uses the defined state labels (OP, UP, P) to predict the CP pipe potential state. The RF, SVM, and NN techniques were evaluated, and the RF model presented the best accuracy. The prediction accuracy results in Figure 16 consider an FDU operating at steady state and an FDU with stray current.

An increase is observed in the prediction accuracy for an FDU operating with stray current using the classification approach (compared to the linear regression approach).

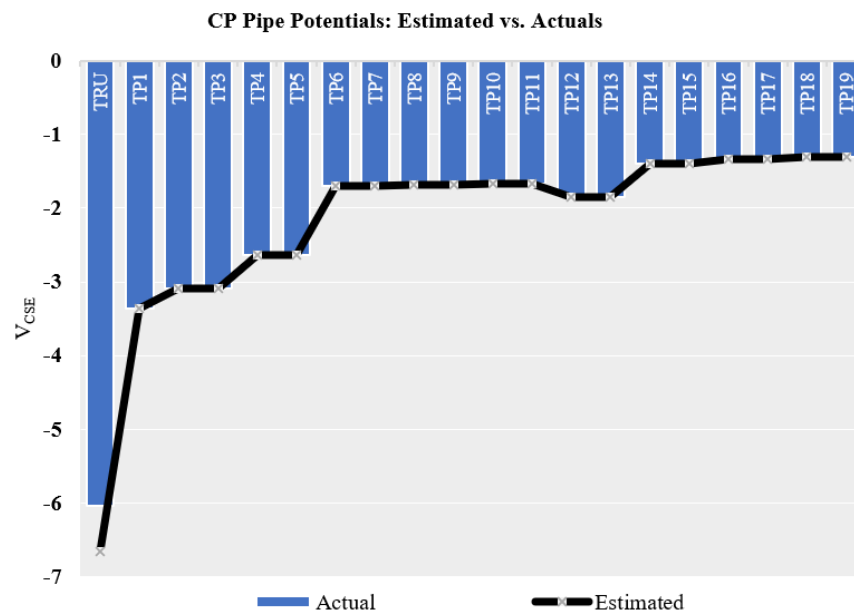


Figure 15. Pipeline section CP pipe potential estimation results.

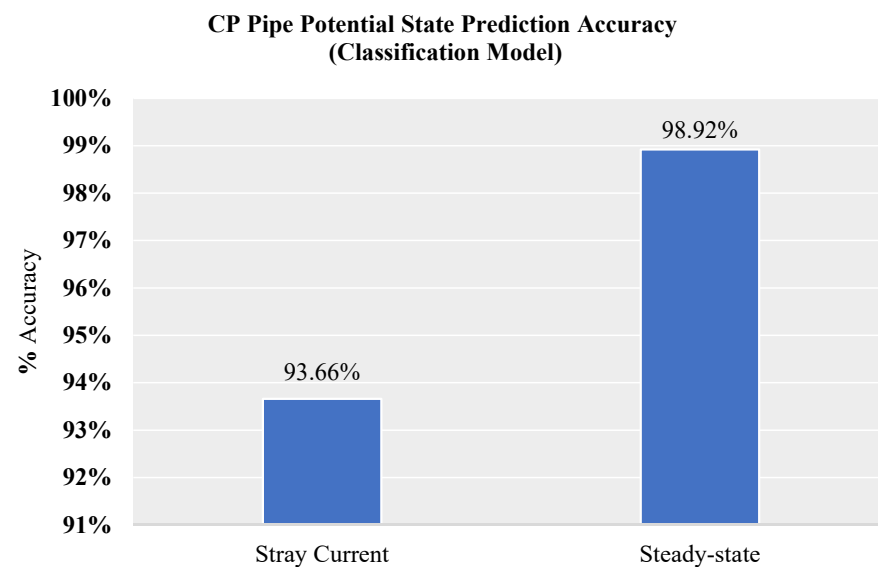


Figure 16. FDU CP pipe potential state prediction results.

5.3.6. Maintenance Matrix

This paper presents a simplified maintenance matrix (Table 3) to remedy the ICCP unit state based on the CP pipe potential. The time limit serves as the maximum time an ICCP unit can operate a specific state and forms the basis for maintenance time estimation.

5.3.7. Maintenance Prediction

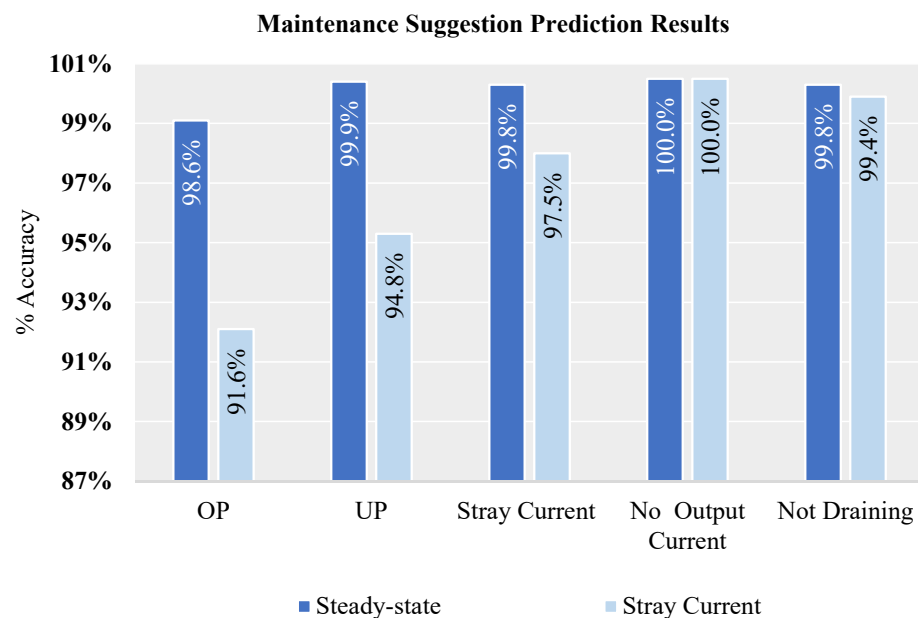
Based on the presented maintenance matrix in Table 3, the RF, SVM, and NN classification techniques were evaluated to predict the required maintenance activity. The RF classification model presented the best prediction accuracy results and is illustrated in Figure 17.

However, the prediction results do not include a time component, and three approaches are presented in the next section.

Table 3. Maintenance matrix.

Maintenance Matrix			
Fault	Overall Risk Level	Time Limit (h)	Required Remedial Action
Overprotection	1	24	M
Overprotection	2	16	A,M
Overprotection	3	8	A,M
Overprotection	4	4	I,A,M
Underprotection	1	24	M
Underprotection	2	16	A,M
Underprotection	3	8	A,M
Underprotection	4	4	I,A,M
Stray-current interference	1	N/A	I,A,M
Rectifier not supplying current	1	8	I,A
Rectifier not draining current	1	8	I,A

Legend: M, monitor; A, adjust; I, investigate.

**Figure 17.** Maintenance suggestion prediction results.

5.3.8. Maintenance Time Suggestion

This paper presents three maintenance time estimation methods: Kaplan–Meier survival analysis, cycle time, and time-series trend component analysis.

5.3.9. Survival Analysis Using Kaplan-Meier Curve

Survival analysis is a statistical tool to determine the probability of surviving an event with a probability of 0.5. A review of the CP pipe potentials over a time frame presents a time component related to the time-to-state (OP, UP, P).

Using the Kaplan–Meier curve or the median survival time, assumptions can be made about future CP pipe potentials if the ICCP unit should continue operating as-is. These assumptions can aid in determining the time-to-maintenance.

Figure 18 plots the median survival time as approximately 75,000 s to a state of “Not-Protected” using a survival probability of 0.5.

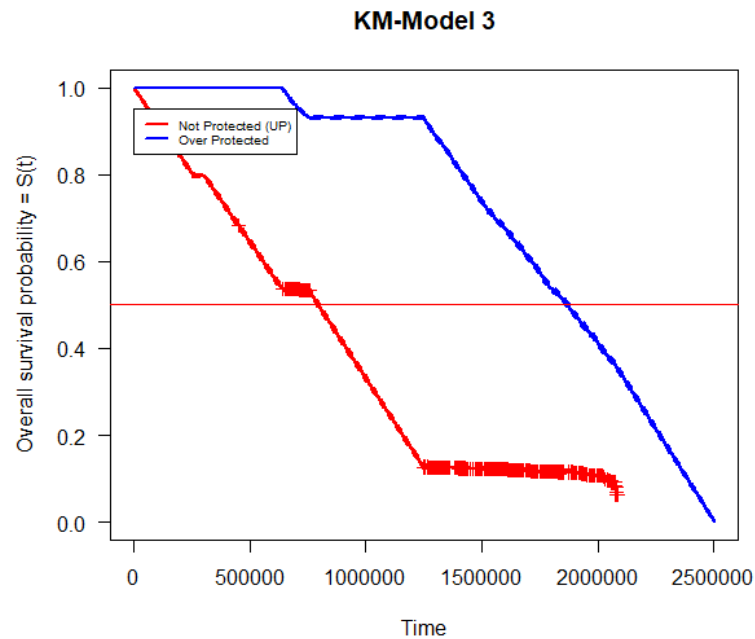


Figure 18. Kaplan-Meier survival curve.

5.3.10. Cycle Time Method

The cycle time method considers the time limit specified in the maintenance matrix and decreases once a specific state (OP, UP, P) is active.

The timestamp of the raw data received is pivotal to estimate the maintenance time, considering the time limit and the entire state-time active. Once the cycle time reaches zero, the time is reset, and the cycle repeats. In practice, the time should only reset once maintenance has been performed.

Table 4 illustrates a typical table of events and the estimated time to maintenance. Combinations of a reset event can be introduced to reset the maintenance timer.

Table 4. Cycle time method for maintenance time suggestion.

Time-To-Event Analysis Using 40 h Event Cycle			
Event Time	Time Difference	Remaining Cycle Time (h)	Estimated Maintenance Time
2020/04/01 01:00:00	0.00	40.00	2020/04/02 17:00:00
2020/04/01 02:00:00	1.00	39.00	2020/04/02 17:00:00
2020/04/01 07:00:00	5.00	34.00	2020/04/02 17:00:00
2020/04/01 14:00:00	7.00	27.00	2020/04/02 17:00:00
2020/04/01 23:00:00	9.00	18.00	2020/04/02 17:00:00
2020/04/02 02:00:00	3.00	15.00	2020/04/02 17:00:00
2020/04/02 06:00:00	4.00	11.00	2020/04/02 17:00:00
2020/04/02 09:00:00	3.00	8.00	2020/04/02 17:00:00
2020/04/02 13:00:00	4.00	4.00	2020/04/02 17:00:00
2020/04/02 17:00:00	4.00	0.00	2020/04/02 17:00:00
2020/04/02 18:00:00	0.00	40.00	2020/04/04 10:00:00
2020/04/02 19:00:00	1.00	39.00	2020/04/04 10:00:00
2020/04/02 20:00:00	1.00	38.00	2020/04/04 10:00:00

5.3.11. Trend Component of Time-Series Object

The forecast package in R enables the decomposition of time-series objects. The trend component of the time-series object enables evaluating the CP pipe potential trend at different frequencies.

After-the-fact analysis of the CP pipe potential can inform planned maintenance activities, such as ICCP unit adjustment due to seasonal weather changes or operational patterns and adjustment based on the CP pipe potential estimation of downstream TPs.

Figure 19 indicates the quarterly change in CP pipe potential over two years. The CP pipe potential experiences seasonal changes (due to the change in soil resistivity and temperature). Trend component analysis can potentially prevent significant CP pipe potential changes if planned maintenance is executed to adjust the supplying ICCP unit output accordingly.

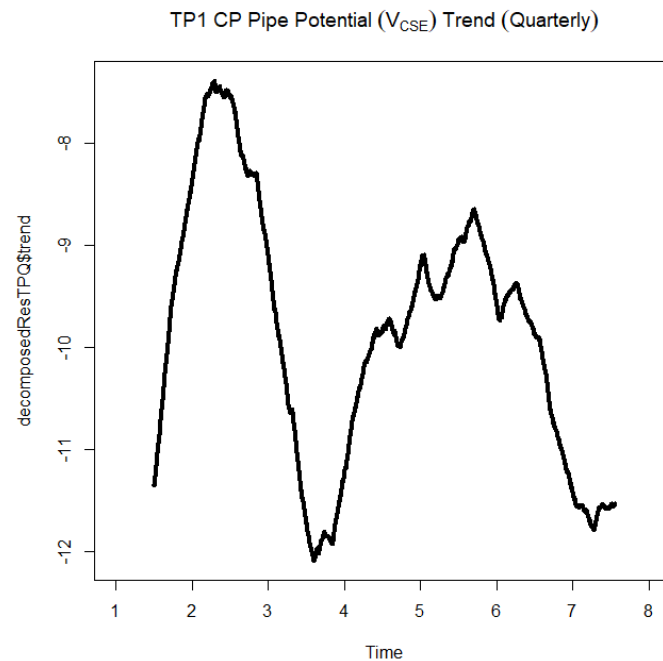


Figure 19. Quarterly CP pipe potential trend.

CP pipe potentials can also be forecasted based on historical data using the forecast package for different frequencies, as illustrated in Figure 20.

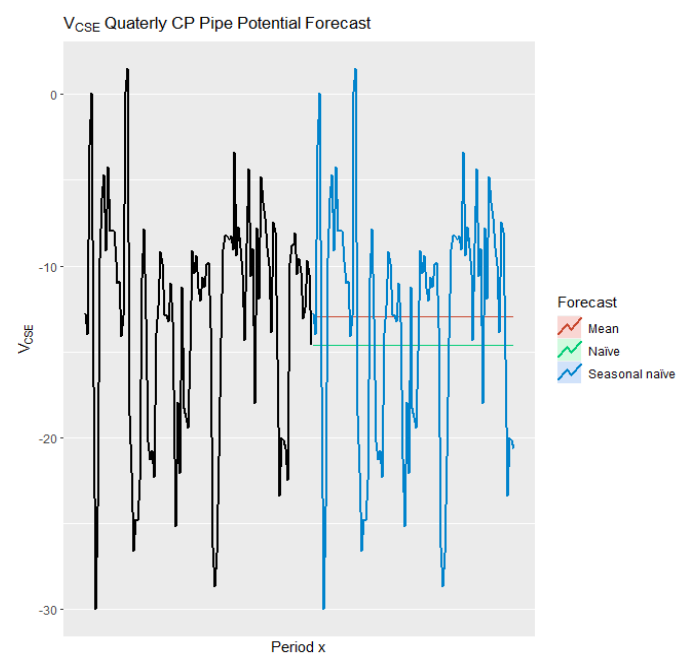


Figure 20. Quarterly forecasted CP pipe potential.

5.4. Descriptive Statistics

Descriptive statistics enable past performance evaluation of an ICCP unit or TP based on the CP pipe potential. The descriptive statistics report on the CP pipe potential's conformance to the OW as time and percentage statistics and are illustrated in Table 5.

Table 5. Descriptive statistics.

Unit	CP Descriptive Statistics					
	% P	% UP	% OP	Time P	Time OP	Time UP
				(min)	(min)	(min)
FDU	4.8	0.05	95.14	2006.5	39731.41	22.09
TP1	3.11	2.6	94.29	41.83	1269.34	35
TP5	12.38	0.1	87.52	165.15	1168.02	1.33

6. Discussion of Results

This paper presented a predictive state and maintenance framework for ICCP units, using the CP pipe potential and guided by the NACE SP0169-2013 CP criteria for instant-on potentials. The framework considers preliminary data analysis activities such as data preparation and transformation to ensure the data are in a format that enables prediction. State labels were assigned to each row in the datasets based on the CP criteria defined in the methodology.

Using R Studio, an exploratory data analysis for ICCP units suggests that the correlation between the rectifier measurements changes once a stray current is present. The stray current waveform repeats in decaying fashion for TPs along the pipeline section evaluated. Based on this finding, two predictive modeling approaches were evaluated: a linear regression technique to predict the CP pipe potential and a classification approach to predict the CP pipe potential states (OP, UP, P).

The multiple linear regression results suggested that the CP pipe potential prediction is achievable using the rectifier output current, voltage, and drainage current (FDU only) as predictors. The best RMSE achieved was 0.153 (or 0.153 V_{CSE}) when a TRU was operating at a steady state. Evaluation of an FDU malfunctioning resulted in an RMSE of 26.95, while an FDU with stray current presented the best RMSE of 0.67 when the data sampling rate was changed from 30 s to 2 min and the interval was increased from 30 days to 4 months. The RF classification approach presented an improved accuracy when predicting the CP pipe potential state for an FDU with a stray current (up to 93.66%).

Estimating the CP pipe potential of downstream TPs (without stray current) requires the output current coefficients for each downstream TP. This paper determines the coefficients from historical CP data and substitutes them in the linear regression equation for the supplying ICCP unit. The difference between the actual and estimated TP CP pipe potentials was 0.0014 V_{CSE} . Where the stray current is present, the coefficients need to be determined at shorter intervals to improve the accuracy.

A maintenance matrix is presented in this paper based on the CP pipe potential states (OP, UP, P), which considers a risk component, a time limit, and suggested remedial actions based on the ICCP unit state. The maintenance activity prediction followed a classification approach and presented the best average accuracy of 96.2% for an FDU with a stray current and 99% for an FDU at a steady state. The maintenance activity's time component consists of the Kaplan–Meier survival analysis, cycle time analysis, and trend component analysis of time-series objects.

7. Conclusions

The predictive maintenance framework presented in this paper can predict the CP pipe potential or the CP pipe potential state and inform the required maintenance activity considering the risk and duration of an event.

By modeling the linear regression coefficients, the pipeline CP segment can be monitored by substituting operating parameters of the specific supplying rectifier output settings (with limited real-time data received).

Selection of the proposed maintenance matrix will depend on the CP data availability (historical versus real-time), the maintenance frequency (long-term versus short-term), and the risk level of the relevant ICCP unit or test post (time-based model based on real-time data and operational state). Based on the data analysis, the selected maintenance model can be selected.

Descriptive statistics enable evaluating the past performance of ICCP units and TP based on the CP pipe potential conformance to the defined OW. Conformance is represented as time and percentage statistics and can typically be used for daily CP monitoring where remote monitoring is in place.

8. Recommendations and Future Work

For an ML model in this paper, to improve and sustain the prediction accuracy, it is recommended that the model is continuously trained with new data. The selection of the ML technique should consider the RMSE and computational overhead. Although a pipeline model can be established using periodic CP data, it is recommended that continuous CP data are used for learning and predicting. The trade-off between continuous and periodic data is based on the accuracy of the prediction outcome.

The sampling rate and data intervals for training data also require consideration for each ICCP unit or pipeline section. Each ICCP unit will present its own unique linear regression coefficients, and it is recommended that a baseline ML analysis is performed for each unit ICCP individually.

Furthermore, high RMSE results can potentially indicate equipment malfunction and be considered a status indicator for each ICCP unit. The suggested maintenance matrix considered the same risk level and time limit for all CP equipment. It is recommended that the risk levels and time limits are adjusted for each ICCP unit individually.

This study only considers a pipeline section with one ICCP unit, and future work includes modeling two or more ICCP units on a pipeline section. Furthermore, additional CP equipment such as AC mitigation stations, cross-bonds, and ER probes can be incorporated into the modeling process. The paper did not consider the IR-drop in the CP pipe potential measurement, and this can be included in future work. Additional predictors can be considered for the predictive modeling (for example, the rectifier output frequency, pipe AC potential, coupon current and voltage, and resistance probes).

Author Contributions: Conceptualization, W.D.; formal analysis, E.R. and W.D.; investigation, E.R. and W.D.; methodology, E.R.; supervision, W.D.; validation, E.R.; writing—original draft, E.R.; writing—review and editing, W.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for the study can be made available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, P.; Wang, L.; Zhu, W.; Wang, J. Statistical Analysis of Underground Pipeline Typical Accidents and Numerical Simulation Research. In Proceedings of the 2nd International Conference on Data Science and Business Analytics, Changsha, China, 21–23 September 2018. [CrossRef]
2. Lochner, P.; Laurie, S.; Bundy, S.; Green, C. Strategic Environmental Assessment for the Development of a Phased Gas Pipeline Network in South Africa FINAL SEA REPORT. Available online: <https://gasnetwork.csir.co.za/final-sea-reports/> (accessed on 15 July 2021).

3. Geck, C. The World Factbook. *Charlest. Advis.* **2017**, *19*, 58–60. [CrossRef]
4. Njomane, L.; Telukdarie, A. Corrosion management: A case study on South African oil and gas company. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Paris, France, 26–27 July 2018.
5. Callister, W.D. Materials science and engineering: An introduction (2nd edition). *Mater. Des.* **1991**, *12*, 59. [CrossRef]
6. Koch, G.; Varney, J.; Thopson, N.; Moghissi, O.; Gould, M.; Payer, J. *International Measures of Prevention, Application, and Economics of Corrosion Technologies Study*; NACE International: Houston, TX, USA, 2016.
7. Marshall, E.G.P.; Parker, E. Fundamentals of Corrosion. In *Pipeline Corrosion and Cathodic Protection*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 1999; pp. 146–149.
8. Datta, K.; Fraser, D.R. A corrosion risk assessment model for underground piping. In Proceedings of the Annual Reliability and Maintainability Symposium, Fort Worth, TX, USA, 26–29 January 2009. [CrossRef]
9. Construction Mentor. Backfilling Pipe Trenches. 2020. Available online: <https://constructionmentor.net/backfilling-pipe-trenches/> (accessed on 2 May 2020).
10. Marshall, E.G.P.; Parker, E. Cathodic Protection of Steel in Soil. In *Pipeline Corrosion and Cathodic Protection*, 3rd ed.; Marshall, E.G.P., Parker, E., Eds.; Elsevier: Amsterdam, The Netherlands, 1999; pp. 150–152.
11. PHMSA. *Pipeline Safety Regulations—49 Cfr Part 192*; U.S. Department of Transportation, Pipeline and Hazardous Materials Safety Administration: Oklahoma City, OK, USA, 2015; pp. 60–85.
12. NACE International. *NACE Standard SP0169—Standard Practice Control of External Corrosion on Underground or Submerged Metallic Piping Systems*; NACE International: Houston, TX, USA, 2013.
13. Engineering, C. Cathodic Protection Rectifier & Electrical Systems. 2019. Available online: <https://www.cathtect.com/cathodic-protection-rectifiers-electrical-systems/> (accessed on 6 September 2020).
14. GCP German Cathodic Protection. 2020. Available online: <http://www.gcp.de/index.php/en/datasheets.html> (accessed on 6 September 2020).
15. Peabody, A.W. *Control of Pipeline Corrosion*, 2nd ed; NACE International: Houston, TX, USA, 2001.
16. Guyer, J.P. *Introduction To Cathodic Protection. Pipes Pipelines Int.*; The Clubhouse Press: El Macero, CA, USA, 2009.
17. O'Connor, P.D.T.; Kleyner, A. *Practical Reliability Engineering, Fifth Edition*; Wiley: Hoboken, NJ, USA, 2011.
18. ISO. *Condition Monitoring and Diagnostics of Machines—General Guidelines*; ISO 17359:2018; ISO: Geneva, Switzerland, 2006.
19. Yu, Z.; Hao, J.; Liu, L.; Wang, Z. Monitoring Experiment of Electromagnetic Interference Effects Caused by Geomagnetic Storms on Buried Pipelines in China. *IEEE Access* **2019**, *7*, 14603–14610. [CrossRef]
20. Holtsbaum, W.B. *Cathodic Protection Survey Procedures*, 2nd ed.; NACE International: Houston, TX, USA, 2012.
21. Inzelt, G.; Lewenstam, A.; Scholz, F. *Handbook of Reference Electrodes*; Springer: Berlin/Heidelberg, Germany, 2013.
22. Park, R.M. A guide to understanding reference electrode readings. *Mater. Perform.* **2009**, *48*, 32–36.
23. Tamhane, D.; Thalapil, J.; Banerjee, S.; Tallur, S. Smart Cathodic Protection System for Real-Time Quantitative Assessment of Corrosion of Sacrificial Anode Based on Electro-Mechanical Impedance (EMI). *IEEE Access* **2021**, *9*, 12230–12240. [CrossRef]
24. Kumar, R.; Dewal, M. Multi-Supervisory Control and Data Display. *Int. J. Comput. Appl.* **2010**, *2*, 1–5. [CrossRef]
25. Runkler, T.A. *Data Analytics*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2016.
26. Abate, F.; di Caro, D.; di Leo, G.; Pietrosanto, A. A Networked Control System for Gas Pipeline Cathodic Protection. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 873–882. [CrossRef]
27. Martin, K. A review by discussion of condition monitoring and fault diagnosis in machine tools. *Int. J. Mach. Tools Manuf.* **1994**, *34*, 527–551. [CrossRef]
28. ISO. ISO 13374-2:2007(en) Condition Monitoring and Diagnostics of Machines—Data Processing, Communication and Presentation—Part 2: Data Processing. 2002. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:13374:-2:ed-1:v2:en> (accessed on 30 September 2020).
29. Medjaher, K.; Zerhouni, N. Framework for a hybrid prognostics. *Chem. Eng. Trans.* **2013**, *33*, 91–96. [CrossRef]
30. Ramasubramanian, K.; Singh, A. *Machine Learning Using R*; Springer: Berlin/Heidelberg, Germany, 2016.
31. Viti, A.; Terzi, A.; Bertolaccini, L. A practical overview on probability distributions. *J. Thorac. Dis.* **2015**, *7*, E7–E10. [CrossRef] [PubMed]
32. Ragab, A.; Ouali, M.-S.; Yacout, S.; Osman, H. Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation. *J. Intell. Manuf.* **2014**, *27*, 943–958. [CrossRef]
33. Lipovetsky, S. Introduction to Data Science: Data Analysis and Prediction Algorithms with R. *Technometrics* **2020**, *62*, 1–282. [CrossRef]
34. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 1–21. [CrossRef] [PubMed]
35. Sahu, H.; Shirma, S.; Gondhalakar, S. A Brief Overview on Data Mining Survey. *Int. J. Comput. Technol. Electron. Eng.* **2011**, *1*, 114–121.
36. Berry, M.; Mohamed, A.; Yap, B.W. *Supervised and Unsupervised Learning for Data Science*; Unsupervised and Semi-Supervised Learning Book Series; Springer International Publishing: Berlin/Heidelberg, Germany, 2019. [CrossRef]
37. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
38. NACE International. *NACE CP 2—Cathodic Protection Technician Training Manual*; NACE International: Houston, TX, USA, 2013.
39. NACE International. *NACE CP 3—Cathodic Protection Technologist Training Manual*; NACE International: Houston, TX, USA, 2005.

40. NACE International. *TM0497-2018-SG Measurement Techniques Related to Criteria for Cathodic Protection on Underground or Submerged Metallic Piping Systems*; NACE International: Houston, TX, USA, 2018.
41. Calhoun, P.; Luo, W.; McPherson, D.; Peirce, K. Layer Two Tunneling Protocol (L2TP) Differentiated Services Extension. 2002. Available online: <https://www.ietf.org/rfc/rfc3308.txt> (accessed on 10 August 2021).
42. Peratta, A.; Baynham, J.; Adey, R.; Pimenta, G.F. Intelligent remote monitoring system for cathodic protection of transmission pipelines. In *NACE—International Corrosion Conference Series*; NACE International: Houston, TX, USA, 2009.
43. Lebold, M.; Reichard, K.; Byington, C.S.; Orsagh, R. OSA-CBM architecture development with emphasis on XML implementations. *Maint. Reliab. Conf.* **2002**.
44. Kuhn, M. The caret Package. 2019. Available online: <http://topepo.github.io/caret/models-clustered-by-tag-similarity.html> (accessed on 25 October 2020).