Computational Fluid Dynamics Driven Mass Transfer Modelling of CO_2 Corrosion in Complex Flow Geometries

Udayraj Yuvraj Thorat

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Mechanical Engineering

November 2023

Publication Statement

I confirm that the work submitted is my own and that appropriate credit has been given where reference has been made to the work of others.

Papers contributing to this thesis:

Thorat, U., Jones, M., Woollam, R., Owen, J., Barker, R., Thompson, H. and de Boer, G. 2024. Computational fluid dynamics driven mass transfer model for the prediction of CO2 corrosion in pipelines. Journal of Pipeline Science and Engineering. 4(1), p.100148.

Acknowledgements

I would like to express my deepest gratitude to my primary supervisor Dr. Gregory de Boer, whose guidance, wisdom, and unwavering support have been instrumental throughout this doctoral journey. His expertise and mentorship have been invaluable, shaping not only the trajectory of my research but also my growth as a scholar.

I am equally thankful to my secondary supervisors, Professor Harvey Thompson and Professor Richard Barker. Their insights, constructive feedback, and encouragement have significantly enriched the depth of my research.

I would like to thank Dr Joshua Owen and Richard Woollam for their support and feedback during my research work. I would also like to thank Michael Jones for his valuable advice on numerical modelling.

To my parents, I express my profound appreciation for your enduring love and support throughout my academic career. Your belief in me since childhood has been a constant source of motivation. I would like to thank my brother Utkarsh for his support and encouragement throughout my life.

I would like to honour my late uncle, Mohan Thorat, who was not merely a source of inspiration but a pillar of strength who always offered words of wisdom during difficult times.

I would like to thank my friends for being there for me during my PhD journey.

I dedicate this thesis to my parents!

Abstract

Oil and gas industries, during the extraction of oil from wells, find water along with carbon dioxide (CO_2), hydrogen sulphides (H_2S), and organic acids. The presence of these species affects the integrity of pipelines used for transportation due to the electrochemical reactions on the metal surface. This phenomenon of degradation is known as carbon dioxide (sweet) corrosion when the presence of CO_2 is much higher than H_2S. Several studies have shown a comprehensive understanding of the mechanism enabling the development of mechanistic predictive tools with the help of empirical correlations of viscous sublayer thickness, turbulent diffusivity, and reaction rate kinetics. These empirical correlations are valid for fully developed flow situations and constrain the corrosion rate predictions in developing and disturbed flow conditions. This limits the scope of current prediction tools in the literature and generates a need to develop a flow and mass transfer coupled model applicable to a broader range of operating conditions. Hence, this is the subject of investigation in this research work.

A computational fluid dynamics (CFD) driven mass transfer model is developed to predict CO_2 corrosion in pipelines. The model involves accurate predictions of viscous sublayer thickness and turbulent diffusivity in a horizontal pipe using CFD. These predictions then drive the 1-dimensional mass transfer model to predict CO_2 corrosion. These predictions were then verified with the experimental dataset available in the literature for pH 4 to 6, velocity 1 to 10 m/s, partial pressure of CO_2 (pCO_2) of 1 bar and temperature of 20°C. A verification with an experimental dataset highlighted the robustness of the CFD-driven model, as the predicted values are well within the range of experimental data.

Machine learning models such as Artificial Neural Network (ANN), Gaussian Process Regression (GPR), Random Forest (RF), and Support Vector Regression (SVR) are applied to predict corrosion rate based on input variables such as pH, velocity, temperature and pCO_2. Machine learning-enabled surrogate modelling is used to determine the sensitivity of electrochemical reaction rate constants and then calculate a reliable set of electrochemical reaction rate constants. Random Latin Hypercube (RLH) sampling was then applied to a range of electrochemical reaction rate constants obtained from the literature. A dimensionality reduction technique, Principal Component Analysis (PCA), is then applied to check if the initial design variables can be reduced. An optimal machine learning model is selected from ANN, GPR, RF and SVR based on evaluation metrics. A set of optimal electrochemical reaction rate constants was then used to compare them against the experimental dataset, and predictions were obtained using the current set of electrochemical reaction rate constants.

When it comes to the predictions of corrosion rates in complex flow situations, few prediction tools are available in the literature. A CFD-driven mass transfer model for the prediction of CO_2 corrosion in complex flow situations is developed in the current study, which accurately couples the CFD model with the mass transfer model by setting the benchmark for complex flow corrosion modelling. The complex flow situation model considered here is a 2D expansion/constriction pipe in which expansion and constriction domains are connected gradually. This CFD model is coupled with the 1D mass transfer model that calculates the corrosion rate over a surface.

CFD-driven mass transfer model for the prediction of CO_2 corrosion in horizontal pipelines predicted corrosion rates reasonably well for pH 5 and pH 6. However, for pH 4 it was found that the corrosion rate predictions were sensitive to the choice of electrochemical reaction rate constants. A systematic approach to finding an optimal set of electrochemical reaction rate constants for pH 4 with the help of supervised machine learning models showed that the GPR model consistently provided lower RMSE values compared to other models. An optimal set of electrochemical reaction rate constants provided the lowest RMSE value of 0.28 between predicted corrosion rates and experimental corrosion rates, showing the robustness of this approach. The current model has shown its capability to predict VSL conditions and corrosion rates in complex flow geometry situations.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACO	Ant Colony Optimization
AELH	Audze-Eglais objective function
ANN	Artificial Neural Network
b	Tafel slope (V)
BAS	Beetle Antennae Search
BPNN	Back Propagation Neural Network
$CaCO_3^{2-}$	Calcium carbonate
CCD	Central Composite Designs
$cCO_{2,ref}$	Reference concentration of species (mol/m^3)
c_j	Concentration of species j (mol/m^3)
CR	Corrosion rate (mm/yr)
CSA	Cuckoo Search Algorithm
D_j	Molecular diffusivity of species j (m^2/s)
DL	Deep Learning
DOE	Design of Experiments
DP	Dynamic Programming
DROpt	Dimensional Reduction-based Optimisation
D_t	Turbulent diffusivity (m^2/s)
EIS	Electrochemical Impedance Spectroscopy
E_{rev}	Reversible potential (V)
F	Faraday constant (C/mol)
FAC	Flow Accelerated Corrosion
FAES	Failure Analysis Expert System
$FeCO_3$	Iron carbonate
GA	Genetic Algorithms
GDP	Growth Domestic product
$g_i(x)$	Inequality constraints
GP	Gaussian Process
GPR	Gaussian Process Regression
GridSearchCV	Grid search cross-validation
ΔH	Activation energy (J/mol)
H^+	Hydrogen ions

H_2CO_3	Carbonic acid
HCO_3^-	Bicarbonate ions
$h_j(x)$	Equality constraints
$i_{0,ref}$	Reference exchange current density (A/m^2)
IFE	Institute for Energy Technology
IP	Integer Programming
I_T	Turbulence intensity
k	Turbulent kinetic energy (m^2/s^3)
$k(x,x')$	Covariance function
K_b	Backward reaction rate constant
K_f	Forward reaction rate constant
LHS	Latin Hypercube sampling
LP	Linear Programming
L_T	Turbulent length scale
M	Measure of uniformity
m(x)	Mean function
MAE	Mean Averaged Error
MLP	Multilayer Perceptrons
MLSM	Moving Least Squares Method
MSE	Mean Squared Error
NID	Network Interpretation Diagrams
N_j	Flux of species j (mol/m^2·s)
NSGA-II	Non-dominated Sorting Genetic Algorithm
nx	Normal x wall component
ny	Normal y wall component
OFAT	One Factor at a Time
OLH	Optimal Latin Hypercube
PCA	Principal Component Analysis
pCO_2	Partial pressure of CO_2 (bar)
PEMFC	Proton Exchange Membrane Fuel Cell
permGA	Permutation Genetic Algorithm
P_k	Production term
PSA	Pressure Swing Adsorption
PSO	Particle Swarm Optimization

RAAE	Relative Average Absolute Error
RANS	Reynolds Averaged Navier Stokes
RBFNN	Radial Basis Function Neural Networks
Re	Reynolds number
RF	Random forest
R_j	Production/reduction of species j
RLH	Random Latin Hypercube
RMAE	Relative Maximum Absolute Error
RMSE	Root Mean Square Error
S	Characteristic magnitude of mean velocity gradients
SA	Simulated Annealing
Sc_t	Turbulent Schmidt number
S_{ij}	Mean strain-rate tensor
SNN	Shallow Neural Networks
SOpt	Surrogate-assisted Optimisation
SSA	Sparrow Search Algorithm
T	Temperature (K)
T	Time (s)
T_{ref}	Reference temperature (°C)
tx	Tangential x wall component
ty	Tangential y wall component
u	Averaged velocity (m/s)
$u^{'}$	Instantaneous velocity (m/s)
u^+	Dimensionless velocity
u_j	mobility of species j
u_t	Tangential velocity (m/s)
u_τ	Frictional velocity (m/s)
VSL	Viscous sublayer
WOA	Whale Optimisation Algorithm
x and y	Cartesian coordinate
y^+	Dimensionless wall distance
ε	Viscous dissipation rate
θ	Closeness of fit
μ_t	Turbulent viscosity (Pa·s)

τ_w	Wall shear stress of the fluid
χ	Design space
Ω_{ij}	Mean rotation-rate tensor

Chapter 1: Introduction

1.1 Introduction

Corrosion is a natural process of degradation of materials as a result of chemical and electrochemical reactions with the environment (Ibrahimi et al., 2021). This process transforms pure metal into chemically stable oxide. The cost of corrosion in terms of global impact was estimated to be £2 trillion, equivalent to 3.4% of the global growth domestic product (GDP) in 2013. The United States (U.S.) refinery's capacity was 27% of the global refinery capacity towards the end of 2014 (Abbas, 2016). Table 1.1 shows the share of each of the industry categories in the U.S. affected by CO_2 corrosion (Koch et al., 2002).

Table 1.1. Corrosion costs share for the United States industry categories (Koch et al., 2002)

Category	Share (%)
Utilities	35
Transportation	21
Infrastructure	16
Government	15
Production and Manufacturing	13

The production and manufacturing industry shared around 13% of total corrosion costs, accounting for £14.65 billion (Koch et al., 2002), which was doubled by 2014 (Abbas, 2016). Figure 1.1 shows the pie chart of costs associated with the production and manufacturing industry.

Figure 1.1 Pie chart of corrosion costs for the production and manufacturing industry(Koch et al., 2002).

As shown in Figure 1.1, the total cost cuts for the oil and gas industry are around 39%, including oil and gas exploration and production, petrochemical, petroleum refining and chemical, and pharmaceutics. Oil and gas industries transport crude oil from the wellhead to the refinery for separation and processing. Pipeline flows are the most economical method for transporting oil/water mixtures. Such pipelines comprise steel for practical engineering implementation and cost-related reasons (Choi et al., 2010). However, several causes result in pipeline failure, as shown in Table 1.2.

Table 1.2 Causes of pipeline failure in the oil and gas industry (Kermani and Harrop, 1996).

Causes of failure	The proportion of failures (%)
Corrosion	33
Fatigue	18
Mechanical damage/overload	14
Brittle fracture	9
Fabrication defects	9
Welding defects	7
Others	10

Out of overall pipeline failures reported, 33% of total failures are caused by corrosion where CO_2 corrosion accounted for the highest proportion of failure shown in Table 1.3 (Kermani and Harrop, 1996). Subsequent to CO_2 corrosion, H_2S corrosion and preferential weld corrosion each contribute 18%, followed by pitting corrosion at 12% and erosion-corrosion at 9%. Galvanic corrosion represents 6% of failures, while crevice, impingement and stress

corrosion each contribute 3% to the overall pipeline failures due to corrosion in the oil and gas industry. Preferential weld corrosion occurs near welds due to factors such as residual stresses, differences in metallurgical properties, and environmental conditions, including effects of flow, scale formation and pH of the solution (Hebert et al., 2019). Impingement corrosion is the degradation of the metal surface due to the impact of fluid with a high velocity, resulting in enhanced mass transfer of corrosive species to the metal surface and removal of corrosion products from the metal surface (Efird, 2000).

It is essential to have an efficient design after considering the factors that affect CO_2 corrosion to minimise capital as well as operational costs. To this end, it is necessary to develop models to predict accurate corrosion rates to determine pipeline longevity, reduce extensive degradation, and improve effective operation. This will aid the development of efficient transportation system design. The need for a predictive tool of CO_2 corrosion is directly related to the health, safety, and economic aspects of the industry, where the monetary impact is in billions of pounds.

Table 1.3 Pipeline failure in the oil and gas industry due to corrosion (Kermani and Harrop, 1996).

Types of corrosion	The proportion of failures (%)
CO_2 corrosion	28
H_2S corrosion	18
Preferential weld	18
Pitting	12
Erosion-corrosion	9
Galvanic	6
Crevice	3
Impingement	3
Stress corrosion	3

Over the last two decades, numerous studies have been carried out to build an accurate model for the prediction of CO_2 corrosion in pipelines (Gardner et al., 2019). There are several empirical (Halvorsen and Sontvedt, 1999; Olsen, 2003), semi-empirical (de Waard and Milliams, 1975; de Waard et al., 2003), elementary mechanistic (Gray et al., 1989; Nesic et al., 1995; Nesic et al., 2001; Nordsveen et al., 2003; Nešić et al., 2009), and comprehensive mechanistic models (Nešić et al., 2019; Kahyarian and Nesic, 2020) that are now available for the prediction of CO_2 corrosion in single-phase pipeline flow. The elementary mechanistic model (Gray et al., 1989) encompasses the simplified fundamental corrosion processes

focused on the fundamental electrochemical reactions occurring at the metal surfaces (Nešić et al., 2009). The comprehensive mechanistic models maintain mass and charge transfer balances while incorporating the effect of homogeneous chemical reactions. The mechanistic models available have extended their scope by adding complexity, such as the effect of multiphase flow (Nešić et al., 2019), the impact of protective corrosion product formation (Nesic et al., 2001; Nordsveen et al., 2003), and other corrosive species into these models (Zheng, 2015; Kahyarian and Nesic, 2020). Some approaches include combinations of experimental work with theoretical modelling for investigating the impact of surface roughness (Al-Khateeb et al., 2018) and rapid expansion geometry with the help of computational fluid dynamics (CFD)(Owen et al., 2019).

CFD was only known for its use in high-technology engineering areas of astronautics and aeronautics; now, it is a quickly adopted technique for solving highly complex problems in general engineering practice. In aerospace and aeronautical industries, CFD is used to analyse wing/tail structures or fuselages, optimise fuel delivery and engine cooling systems, and the critical components' initial or complete design process (Isaac, 2013). CFD has many advantages in automotive engineering, such as shortening cycles, optimising existing components, studying the vehicle's external aerodynamics, and improving the in-car environment (Dhaubhadel, 1996). Medical researchers rely primarily on simulation techniques to predict the behaviour of blood flow through the veins and arteries of the human body (Djukic et al., 2022). The information challenging to obtain from the experiments can be provided with the help of computational simulations by allowing variations of problems to be studied parametrically (Veersteg and Malalasekara, 2016). Governments, corporations, and research institutes are more actively seeking ways to reach environmental regulatory requirements by reducing waste (Wójtowicz-Wróbel et al., 2023) and maintaining acceptable production levels to meet increased market demands. CFD has been beneficial for predicting the flow field required for corrosion prediction in pipelines (Subraveti et al., 2019) and elbows (Keating and Nesic, 2000) for single-phase flow simulations. In addition, CFD provided valuable insights into understanding the mechanism of slug flow-induced CO_2 corrosions (Zheng et al., 2007; Lv et al., 2020).

Underpinning the current state-of-the-art modelling CO_2 corrosion in pipeline flows, empirical measurements of reaction rate kinetics, viscous sublayer thickness, and turbulent diffusivity restrict predictions to simple geometries, steady-state and single-phase flows. This investigation aims to further the current capabilities of CO_2 corrosion prediction in pipeline flows by deriving and implementing a novel method for accurately incorporating CFD with mass transfer modelling. Current predictive tools in the literature are valid for fully developed flow conditions as the empirical correlations used for the prediction of flow conditions cannot be applied to the flow situations where the flow is developing and/or disturbed (Nordsveen et al., 2003; Kahyarian and Nesic, 2020). This limitation can be overcome with the capability of CFD modelling to predict corrosion rates when the flow is in developing and disturbed condition (Li and Woollam, 2012). There is a research gap in the literature related to the mass transfer modelling of CO_2 corrosion in complex flow geometries. In addition, the current study has implemented machine learning technologies to develop a CFD-enabled machine learning predictive tool. Machine learning tools with fast computing speed and the ability to handle large datasets have been used to develop predictive tools in the corrosion science field (Aghaaminiha et al., 2021). The current study has implemented machine learning techniques

to predict corrosion rates based on operating parameters and to predict optimal electrochemical reaction rate constants.

The main driving force for this study is to develop a robust predictive tool for a better understanding of corrosion in complex flow situations.

1.2 Aim and Objectives

The aim of this research is to build a robust mass transfer model with the accurate coupling of the CFD model to predict CO_2 corrosion in complex flow situations.

Below are the objectives of this research project,

- Carry out a literature review to obtain knowledge of the latest research activities associated with the field of CO_2 corrosion, plan the project activities and direct the research's focus.
- Investigate the current mechanistic models for the CO_2 corrosion prediction and understanding the limitations of those models.
- Develop a methodology for the prediction of CO_2 corrosion with the aid of CFD in horizontal straight pipelines.
- Find out the most influential factors that affect CO_2 corrosion. Plot corrosion rate profiles against factors, for example, pH, pCO_2, temperature and velocity.
- Develop a CFD-enabled machine learning model to predict corrosion rates based on operating parameters such as pH, partial pressure of CO_2, Temperature and velocity. Study the performance of supervised machine learning models and choose the appropriate model based on cross-validation.
- Develop machine learning-based surrogate modelling to find optimal electrochemical reaction rate constants.
- Develop a methodology for the CO_2 corrosion in complex flow situations by accurately coupling CFD with a mass transfer model.

1.3 Thesis Structure

The first two chapters of the thesis provide the theory, background information and a literature review related to CO_2 corrosion, CFD, surrogate modelling, machine learning techniques and optimisation techniques. The third chapter explains the methodology for the mass transfer modelling of CO_2 corrosion and its coupling with CFD to provide a robust tool. The fourth chapter describes the methodology for machine learning enabled surrogate modelling that includes the analysis of the most significant factors, such as pH, velocity, temperature and partial pressure of CO_2 (pCO_2). The fifth chapter provides a detailed overview of machine learning-enabled surrogate modelling to find optimal electrochemical reaction rate constants. The sixth chapter explains the mass transfer modelling of CO_2 corrosion in complex flow situations. The seventh chapter discusses the outcomes and the conclusion of the research study.

Chapter 2: Literature Review

2.1   CO₂ Corrosion

Carbon and low alloy steel pipelines transport oil/gas/water from wells to the station where crude oil is separated from water. This distance between the wells and the station ranges between 10 to 100 kilometres (Martínez-Palou et al., 2011). During the transportation of the oil/water mixture, interaction occurs between metal and electroactive species due to the chemical surface properties of metal and the electrochemical properties of dilute species, where the term dilute refers to the relatively small number of species compared to the total oil/water mixture in fluid flows. Their concentration field is calculated using the Nernst-Planck equation shown in section 2.3. This phenomenon is described in this section.

2.1.1   Water Chemistry

In CO_2 corrosion process, gaseous carbon dioxide dissolves in water, which then hydrates to form carbonic acid (H_2CO_3). This reaction is rapid and reversible. Carbonic acid is a weak acid, and the dissociation of carbonic acid forms bicarbonate ions (HCO_3^-) and hydrogen ions (H^+). The bicarbonate ions then dissociate to form carbonate ions (CO_3^{2-}) and hydrogen ions. Water dissociates to form hydroxide ions (OH^-) and hydrogen ions. These chemical reactions are shown in Table 2.1, and Table 2.2 shows the empirical equations for the equilibrium constants where K_f is the forward reaction rate constant, and K_b is the backward reaction rate constant ($K = K_f/K_b$).

Table 2.1 Chemistry of CO_2 dissociation in water.

Description	Reaction	Equilibrium Constant	
Carbon dioxide dissolution	$CO_{2(g)} \rightleftharpoons CO_{2(aq)}$	$K_{sol} = \dfrac{CO_{2(aq)}}{pCO_{2(g)}}$	(2.1)
Carbon dioxide hydration	$CO_{2(aq)} + H_2O_{(l)} \rightleftharpoons H_2CO_{3(aq)}$	$K_{hyd} = \dfrac{[H_2CO_3]}{\left[CO_{2(aq)}\right]}$	(2.2)
Dissolution of carbonic acid	$H_2CO_{3(aq)} \rightleftharpoons HCO_{3\ aq}^- + H^+_{(aq)}$	$K_{ca} = \dfrac{[HCO_3^-][H^+]}{[H_2CO_3]}$	(2.3)
Dissolution of bicarbonate ion	$HCO_{3\ aq}^- \rightleftharpoons CO_{3\ (aq)}^{2-} + H^+_{(aq)}$	$K_{bi} = \dfrac{[CO_3^{2-}][H^+]}{[HCO_3^-]}$	(2.4)
Dissociation of water	$H_2O_{(l)} \rightleftharpoons OH^-_{(aq)} + H^+_{(aq)}$	$K_w = [OH^-][H^+]$	(2.5)

Table 2.2. Empirical equations for the equilibrium constants.

Equilibrium constants	Reference	
$K_{sol} = \dfrac{14.5}{1.00258} \times 10^{-(2.27+0.00565 \cdot T_f - 8.06 \times 10^{-6} \cdot T_f^2 + 0.075 \cdot I)} \dfrac{\text{molar}}{\text{bar}}$	(Oddo and Tomson, 1982)	(2.6)
$K_{hy} = 0.00258$	(Palmer and Van Eldik, 1983)	(2.7)
$K_{f,hy} = 10^{329.85 - 110.541 \times \log_{10}(T_k) - (\frac{17265.4}{T_k})} \text{s}^{-1}$	(Palmer and Van Eldik, 1983)	(2.8)
$K_{ca} = 387.6 \times 10^{-(6.41 - 1.594 \times 10^{-3} \cdot T_f + 8.52 \times 10^{-6} \cdot T_f^2 - 3.07 \times 10^{-5} \cdot p - 0.4772 \cdot I^{0.5} + 0.118 \cdot I)} \text{ molar}$	(Oddo and Tomson, 1982)	(2.9)
$K_{f,ca} = 10^{5.71 + 0.0526 \cdot T_c - 2.94 \times 10^{-4} \cdot T_c^2 + 7.91 \times 10^{-7} \cdot T_c^3} \text{s}^{-1}$	(Bamford & Tiffer, 1972)	(2.10)
$K_{bi} = 10^{-(10.61 - 4.97 \times 10^{-3} \cdot T_f + 1.331 \times 10^{-5} \cdot T_f^2 - 2.624 \times 10^{-5} \cdot p - 1.166 \cdot I^{0.5} + 0.3466 \cdot I)} \text{ molar}$	(Oddo and Tomson, 1982)	(2.11)
$K_{f,bi} = 10^9 \text{s}^{-1}$	(Nordsveen et al., 2003)	(2.12)
$K_{wa} = 10^{-(29.3868 - 0.0737549 \cdot T_k + 7.47881 \times 10^{-5} \cdot T_k^2)} (\text{molar})^2$	(Kharaka et al., 1988)	(2.13)
$K_{b,wa} = 7.85 * 10^{10} \ (\text{molar})^{-1}(\text{s})^{-1}$	(Delahay, 1952)	(2.14)

2.1.2 Electrochemical Reactions

The general form of electrochemical reaction is given by Eq. (2.15) (Owen, 2018),

$$M \rightarrow M^{n+} + ne^- \tag{2.15}$$

where n is the number of electrons and M is metal. The metal ions are transported from the anode to the cathode through an electrolyte. The anode is the metal part where oxidation occurs, whereas the cathode is the part where reduction occurs. The anode and cathode can be next to each other or separated by distance. Figure 2.1 shows the primary corrosion cell.

Figure 2.1 Basic corrosion cell to show the anode, cathode and flow of electrons.

The overall reaction that takes place in a CO_2 corrosion process is given below,

$$Fe_{(s)} + CO_{2(g)} + H_2O_{(l)} \rightarrow FeCO_{3(s)} + H_{2(g)} \tag{2.16}$$

The anodic dissolution of iron is as follows,

$$Fe_{(s)} \rightarrow Fe^{2+}{}_{(aq)} + 2e^- \tag{2.17}$$

The CO_2 presence increases the rate at which hydrogen ions are reduced, thus increasing the corrosion rate of the surface. CO_2 corrosion is a mass transfer-limited process; hence, the rate of hydrogen evolution depends on the rate at which it is transported from the bulk to the metal surface (Nesic et al., 2001). As the pH value increases above 4, the mass transfer controlled current reduces, and then the presence of carbonic acid (H_2CO_3) allows hydrogen evolution. The cathodic reactions of hydrogen evolution are given as follows,

$$2H^+{}_{(aq)} + 2e^- \rightleftharpoons 2H_{2(g)} \tag{2.18}$$

$$2H_2CO_{3(aq)} + 2e^- \rightleftharpoons 2HCO_{3\ aq}^- + 2H_{2(g)}$$

2.2 Types of CO₂ Corrosion

2.2.1 Localised Corrosion

Localised corrosion is the type of corrosion that occurs over smaller regions of the metal surface. A smaller surface area of the metal is attacked faster by a corrosive environment than the total surface area of the metal (Alawadhi, 2009). The surrounding area affected by localised corrosion either remains unaffected or is subject to general corrosion (Kermani and Morshed, 2003). Non-uniform formation of corrosion product scale on the metal surface is one of the vital factors that contribute to localised CO_2 corrosion. There are three different variants of localized corrosion: pitting, mesa attack and flow-induced localized corrosion (Kermani and Morshed, 2003). Figure 2.2 shows the different types of localised corrosion.

(a)   (b)

(c)

Figure 2.2 Different types of localised corrosion in pipelines. a) Pitting attack, b) Mesa-type attack, c) Flow-induced localized corrosion (Perez, 2013).

Pitting occurs at moderate to low velocities when the temperature ranges are around the dew point. The pitting depends on the temperature and CO_2 partial pressure. As the temperature and partial pressure of CO_2 increases pitting increases. Pitting is likely around 80-90 degrees Celsius in sweet gas (H_2S) wells (Kermani and Morshed, 2003).

The mesa-type attack is often associated with medium flow velocity conditions. Corrosion due to mesa type of attack results in extensive flat-bottomed localised damage with sharp steps at the edges of the metal surface (Dugstad, 1998). Corrosion due to mesa attack is most likely to occur at the surface with unstable carbonate films (Al-Moubaraki and Obot, 2021). Local spalling of carbonate scales is also a significant factor in mesa attack. Metal below the carbonate scales gets exposed due to local spalling, which then corrodes, and surface films are again reformed. The spalling of the carbonate scales is due to intrinsic growth stresses in the scale.

Flow-induced localised corrosion, above critical flow velocities, is initiated from where pits and/or mesa attack sites are already created. These pits and/or mesa attack sites result in local turbulence, which aids in disseminating flow-induced localised corrosion attack. The critical flow conditions – flow with enhanced wall shear stress may hinder the formation of scale growth on the metal surface after the destruction by local turbulence, along with stresses developed during the scale growth (Halvorsen and Sontvedt, 1999).

2.2.2   Uniform or General Corrosion

Uniform or general corrosion is the type of corrosion in which the attack is consistent over the entire surface area of a material (Liu et al., 1994). It is also the most common type of corrosion observed in industries. In addition, it is one of the widely studied and well-understood types of corrosion (Kahyarian et al., 2016).

2.2.3   Galvanic Corrosion

Galvanic corrosion is experienced when two dissimilar metals come in contact with each other in the presence of an electrolyte (Ma et al., 2023), formation of precipitates on the metal surface (Chang et al., 2014; Sainz-Rosales et al., 2022), and due to the turbulence in disturbed flow (Heitz, 1996; Li and Woollam, 2012). Figure 2.3 shows the formation of anodic and cathodic sites during erosion-corrosion in disturbed flow conditions. The stages of formation of galvanic cells show that copper base materials develop anodes in high-intensity regions. In contrast, non or low-alloyed ferrous materials develop anodes in more stagnant flow regions (Heitz, 1996).

Figure 2.3 Galvanic or electrochemical cell formation during erosion-corrosion in a disturbed turbulent flow situation(Heitz, 1996).

2.2.4   Key Factors Influencing Corrosion

Different factors affect the corrosion rate of the metal surface. These factors include operating conditions, fluid chemistry, hydrocarbon, inhibitors' presence and material properties shown in Figure 2.4 (Nordsveen et al., 2003; Kermani and Morshed, 2003; Abbas, 2016; Owen et al., 2019). Out of all the factors that affect the CO_2 corrosion, the most significant factors such as temperature, partial pressure of CO_2 (pCO_2), and the velocity of flow is discussed in this section. Figure 2.4 shows the parameters affecting corrosion in the oil and gas industries.

Figure 2.4 Parameters affecting corrosion in oil and gas industries (Kermani and Morshed, 2003; Owen, 2018).

Effect of Temperature

The corrosion rate increases with an increase in temperature at a given partial pressure of CO_2 (Nesic et al., 1995). This results from the speed-up of physicochemical processes involved in corrosion when temperature increases (Nešić et al., 2019). Figure 2.5 shows the variation

in corrosion rate for X65 steel, a type of high-strength, low-alloy steel, as a function of temperature and pH values (Nešić et al., 2019).

Figure 2.5 Effect of temperature on the CO_2 corrosion rate in a flow loop at velocity (v) = 2 m/s, diameter (d) = 0.0254 m, C_{Fe}^{2+}< 1 ppm, 1% wt. of NaCl, total partial pressure p_{total}= 1 bar. Solid lines represent the corrosion rate from the (Nešić et al., 2019) mechanistic model, and points represent the results from the experimental data for X65 steel (Nesic et al., 1995).

Above a specific temperature range, it is agreed that the protective layers of iron carbonate ($FeCO_3$) are formed, which results in a decrease in the corrosion rate as temperature increases (Nordsveen et al., 2003). These protective layers create a physical barrier between surrounding fluid and metal, thus reducing the rate of species' transport to and from the metal surface (Nesic et al., 2001). The temperature of an aqueous solution significantly impacts the formation of protective carbonate layers. At temperatures below 60°C, the protective layers may not form as the solubility of $FeCO_3$ is high, and the precipitation rate is slower than at higher temperatures (Nesic et al., 2001). At this temperature range, CO_2 corrosion is a function of temperature, pH, and partial pressure of CO_2, and the metallurgy of the steel (Hunnik et al., 1996). As the temperature increases above 60°C, the solubility of iron carbonate decreases, resulting in the enhancement of the protectiveness of the iron carbonate layer, and the corrosion rate increases up to a temperature of 80°C, and after that, with the increase in the temperature, corrosion rate decreases.

Effect of Partial Pressure of CO₂

The partial pressure of CO_2 is significant in both the absence and the formed protective film conditions. de Waard and Milliams (1975), in a study, provided a relationship between the partial pressure of CO_2 (p_{CO_2}) and corrosion rate at temperatures 15, 25, and 60°C is given below,

$$\text{Corrosion rate} = \text{constant} \times p_{CO_2}{}^C$$

where C is 0.67. Some researchers found that the value of C is between 0.5 and 0.8 (Wang, 1999).

The effect of partial pressure of CO_2 is studied in the absence and presence of carbonate scales. As the partial pressure of CO_2 increases in the absence of carbonate scales, the solution's pH decreases and the carbonic acid reduction rate increases. Thus, the corrosion rate increases (Nesic and Lunde, 1994). Sun and Nesic (2004) found that under protective film-forming conditions, high partial pressure of CO_2 at pH below 5.2 reduces the corrosion rate. This was due to fewer cathodic sites, which increased bicarbonate and carbonate ions (Sun and Nesic, 2004).

Effect of Flow Velocity

Flow contributes to the transport of species entirely or partially to the surface due to turbulence. In CO_2 corrosion, reduction of H^+ ions are mass transfer controlled and at higher velocities, H_2CO_3 reduction is influenced by the interaction between a chemical reaction and mass transfer. This shows that the flow velocity has a significant impact on the CO_2 corrosion (Nesic et al., 2004). Flow velocity affects the CO_2 corrosion in two distinct ways depending on the conditions favourable for forming a protective layer. In conditions unfavourable for the formation of the protective layer, flow enhances the transport of species to and from the metal surface, increasing the corrosion rate (Nordsveen et al., 2003). Contrarily, when the conditions are favourable for the formation of the protective layer, which usually occurs at higher pH values, the flow interferes with the formation of protective layers or removes them from the surface, increasing the corrosion rate (Nesic et al., 2003). Figure 2.6 depicts the impact of velocity on the corrosion rate in a straight pipeline and rotating cylinder electrode (RCE) (Nesic et al., 1995).

Figure 2.6 Variation of corrosion rate for X65 steel with respect to velocity at pH = 4, Temperature T = 20, 1% wt. of NaCl total partial pressure p_{total}= 1 bar (Nesic et al., 1995).

2.2.5 CO₂ Corrosion Prediction Models

Several oil and gas companies have created different models for the prediction of CO_2 corrosion rates. Few of these models are based on empirical correlations with either laboratory or field data (Nyborg, 2010. Some of these models are the mechanistic model, which includes homogeneous chemical reactions and electrochemical reactions that occur in CO_2 corrosion of steel. Nyborg (2010) in their study compared all the available models for the CO_2 corrosion prediction. This section provides an overview of the CO_2 corrosion prediction models.

de Waard Model

de Waard and Milliams (1975) developed a CO_2 corrosion model that was dependent on the temperature and partial pressure of CO_2. This model was updated later by adding correction factors (de Waard et al., 1991; de Waard et al., 1995). de Waard et al. (1995) created an updated version of de Waard and Milliams (1975) model, which has a correction factor for corrosion product scales. The model could not provide a good account for forming protective films above the temperature of $80°C - 90°C$. This model also included a factor for oil wetting where crude oil systems were available. This system did not apply to condensate systems (de Waard et al., 1991), as water separation is much easier than crude oil systems. The model requires input for pH when the formation water chemistry is assumed. Considering all these limitations, this model did not apply to the broader range of the data; however, it was a good starting point for the empirical prediction models (Nyborg, 2010).

Norsok M-506 Model

Norwegian oil companies Statoil, Norsok Hydro and Saga Petroleum (Nyborg, 2010) developed an empirical model for the prediction of CO_2 corrosion. This model is fitted to data similar to that of de Waard et al. (1995), with some other experiments at higher temperatures between 100 °C and 150 °C and corrosion rates for temperatures below 20 °C set equal to corrosion rates at $20°C$. This model considers the effect of the corrosion product layer at high temperatures and high pH efficaciously compared with other corrosion prediction models. In addition to this, the model is more sensitive to variation in pH than the de Waard et al. (1995) model. The model includes options for calculating the pH value for three conditions and wall shear stress based on the pipe diameter and production rates. The model does not include any effect of oil wetting on the corrosion rate (Nyborg, 2002).

However, the model has several limitations, such as under-predicted corrosion rates for the total content of organic acids over 100 ppm and pCO_2 less than 0.5 bar offered no feasibility to determine the critical flow velocity as the correlation was purely empirical.

Hydrocor Model

Shell developed the Hydrocor model (Kopliku and Gunaltun, 2006) to combine fluid flow and corrosion modelling. This model includes the multiphase flow's effect and an oil-wetting factor considered for crude oil systems. For liquid velocity above 1.5 m/s and water cut below 40%, it is assumed that the surface is wetted by oil, hence no corrosion. As the water cut increased above 40%, the wettability conditions would change from oil-wet to water-wet; hence, the corrosion risk would be higher. This model includes the predictions for top-of-line corrosion, H_2S corrosion and organic acid corrosion. The fluid flow model includes calculating velocity, pressure and temperature profiles along a pipeline.

The remarkable thing about this model was the prediction of H_2S corrosion. It uses the pitting factor between 0.7 and 6 for the regions where sulphide dominates based on the ratio of $\frac{pCO_2}{pH_2S} < 20$. Some of the limitations of this model include the assumption of weak protection from corrosion product layers and the neglect of the effect of gas condensate systems.

Corplus Model

Total developed a model (Nyborg, 2002) by merging the Cormed tool created by Crolet and Bonis (1991) and the Lipucor model created by Total (Gunaltun, 1996). The model adopts the pH calculation from the Cormed tool and gives the same potential corrosivity index as the Cormed tool in no protective films or oil-wetting conditions. First, the fluid flow calculations are obtained from the Lipucor model. It then provides a corrosion rate as the potential corrosivity index multiplied by a water-wetting factor, giving lower predictions of corrosion rates for a critical velocity of 0.5 m/s.

The model has a provision for the input of concentration of calcium carbonate ($CaCO_3$). When the concentration value of $CaCO_3$ provided by the user leads to $CaCO_3$ super-saturation, the model shows a warning to verify the magnitude of $CaCO_3$ concentration. If the user ignores this warning message, the program corrects this value by calculating a pH of a lower magnitude (Nyborg, 2010).

KSC Model

Institute for Energy Technology (IFE), Norway developed a KSC model by combining the electrochemical and transport models (Nesic et al., 1995). This mechanistic model includes homogeneous chemical reactions, electrochemical reactions at the surface, diffusion of species to and from the bulk solution, and species diffusion through porous protective films. This was the first step in building comprehensive mechanistic models. This model is considered to have a strong influence on protective corrosion films. Hence, the model predicts low corrosion rates for higher values of temperature and pH.

Multicorp Model

Multicorp model is a mechanistic model based on the KSC model, and it includes homogenous chemical reactions, electrochemical reactions at the surface, and species transport due to diffusion, convection and electro-migration. This model is an updated version of the KSC model with the coupling of the multiphase flow model. It also includes the effect of iron carbonate films and oil wetting. The effects of organic acids such as acetic acid and H_2S are also included in the model. The results from this model are verified against the data available from the laboratory and field (Nesic et al., 2005; Nyborg, 2010).

The corrosion prediction model is a 1D model which assumes uniform corrosion across the length of the pipe. This model is applied to situations where the flow is fully developed. It uses empirical correlations to predict viscous sublayer thickness and turbulent diffusivity (Davies, 1972). In addition, it uses the empirical constants for the calculation of chemical and electrochemical reaction rates, which are found to be inconsistent in different mathematical prediction tools for the prediction of corrosion rates (Nordsveen et al., 2003). These limitations have restricted the use of mechanistic models for predicting corrosion rates in complex flow situations.

Recent Prediction Models

Recent prediction models include the theoretical investigation based on the comprehensive mechanistic model integrated with a Pitzer-type specific interaction model (Kahyarian and Nesic, 2020). This model considers the system's non-ideal behaviour and the hydrogen ion reduction as the only cathodic reaction. The predictions from this model were compared against the experimental data on X65 mild steel and 316L stainless steel for pH range 4-6 and partial pressure of CO_2 from 1 to 15 bar and found a reasonable agreement between predicted data and experimental data.

All the models summarised here are either empirical, semi-empirical, mechanistic or comprehensive mechanistic models in nature. Empirical and semi-empirical models are fitted to large data sets, and the predictions are limited to a specific range of parameters (de Waard and Milliams, 1975; Olsen, 2003). Whereas current mechanistic and comprehensive mechanistic models are based on empirical correlations for chemical and electrochemical reaction rate constants, viscous sublayer thickness and turbulent diffusivity predictions (Nordsveen et al., 2003; Kahyarian and Nesic, 2020). The mechanistic models available have extended their scope by adding complexity, such as the effect of multiphase flow (Nešić et al., 2019), the impact of corrosion product layer (Nesic et al., 2001; Nordsveen et al., 2003), and other corrosive species into these models (Zheng, 2015; Kahyarian and Nesic, 2020). Some approaches include combinations of experimental work with theoretical modelling for investigating the impact of surface roughness (Al-Khateeb et al., 2018) and rapid expansion geometry (Owen et al., 2019).

2.3 Mass Transfer Modelling

2.3.1 Governing Equation for Species Transport

The governing equation (species conservation) for the transport of diluted species j in a fluid medium is given by Eq. (2.21),

$$\frac{\partial c_j}{\partial t} = -\nabla \cdot \boldsymbol{N}_j + R_j \tag{2.21}$$

where c_j the concentration of species j, R_j is the production/reduction of species j due to homogeneous chemical reactions, $\boldsymbol{N}_j$ is the flux of species j, and t is time.

The flux of species $\boldsymbol{N}_j$ consists of diffusion, electro-migration, and convection. The flux of species is given by,

$$\boldsymbol{N}_j = -D_j \nabla c_j + c_j \boldsymbol{u} - z_j F c_j u_j \nabla \phi \tag{2.22}$$

where D_j is the molecular diffusivity of species j, z_j is a charge of species j, u_j mobility of species j, F is Faraday constant (96485 C/mol), $\boldsymbol{u}$ is the fluid velocity in bulk (m/s) and $\boldsymbol{u}'$ is the instantaneous velocity in m/s.

The fluid velocity $\boldsymbol{u}$ has two components: averaged velocity $\bar{\boldsymbol{u}}$ and instantaneous velocity $\boldsymbol{u}'$. The averaged velocity component $\bar{\boldsymbol{u}}$ does not contribute to species transport in the 1D mass transfer model (Nesic et al., 2001).

The turbulent diffusivity (D_t) the term is used to account for $\boldsymbol{u}'$ is defined by,

$$D_t = \frac{\mu_t}{\rho Sc_t} \tag{2.23}$$

where μ_t is turbulent viscosity term, and Sc_t is turbulent Schmidt number with an approximate value of 0.5 to 0.9 (Tominaga and Stathopoulos, 2007). Sc_t is dimensionless quantity that characterizes rate of momentum diffusion to mass diffusion in turbulent flow conditions.

For 1D modelling, the average velocity component $\boldsymbol{u}$ does not contribute to species transport and hence is ignored. The flux due to electro-migration is ignored in the current 1D model. The flux of species $\boldsymbol{N_j}$ for 1D is given as,

$$\boldsymbol{N_j} = -(D_j + D_t)\nabla c_j \tag{2.24}$$

Equation becomes,

$$\frac{\partial c_j}{\partial t} = \nabla \cdot \left[(D_j + D_t)\nabla c_j\right] + R_j \tag{2.25}$$

Figure 2.7 shows the transport of species in 1D mass transfer modelling.

Figure 2.7. Transfer of species in mass transfer modelling.

For 2D mass transfer modelling, average velocity component $\boldsymbol{u}$ contributes to the transport of species; hence, the equation for the flux of species N_j becomes,

$$\boldsymbol{N_j} = -(D_j + D_t)\nabla c_j + c_j\boldsymbol{u} \tag{2.26}$$

Eq. (2.21) becomes,

$$\frac{\partial c_j}{\partial t} = \nabla \cdot \left[(D_j + D_t) \nabla c_j \right] + \nabla \cdot \left(c_j \boldsymbol{u} \right) + R_j \tag{2.27}$$

This equation is used to solve for the transport of species in complex flow situations where the flow is disturbed and not fully developed. The term turbulent diffusivity D_t and average velocity component $\boldsymbol{u}$ are obtained from fluid flow modelling using computational fluid dynamics. The following section focuses on the general review of computational fluid dynamics and its applications in corrosion modelling.

2.3.2 Charge Conservation

The motion of charged particles causes current in the electrolytic solution and is expressed by,

$$i = F \sum_j z_j \boldsymbol{N}_j \tag{2.28}$$

According to the law of charge conservation,

$$\nabla \cdot i = 0 \tag{2.29}$$

Substituting Eq. (2.28) into Eq. (2.29)

$$\nabla \cdot (k \nabla \phi) + F \sum_j z_j \nabla \cdot \left(D_j \nabla c_j \right) = 0 \tag{2.30}$$

For a uniform value of electric conductivity k and in the absence of concentration gradients, the equation is reduced to Laplace's equation given below,

$$\nabla^2 \phi = 0 \tag{2.31}$$

2.4 Computational Fluid Dynamics

2.4.1 Introduction

CFD is a branch of fluid mechanics that solves governing equations for fluid flow using computational resources. With the help of numerical analysis and algorithms, it solves problems that involve fluid flows. The obtained solution is a collection of pointwise field solutions space points when carried out at distinct time levels. CFD integrates the disciplines of fluid mechanics and mathematics with computer science, where computer programmers code the equations that represent the physical laws that govern fluid flow. When an analytical or theoretical solution is impossible, CFD approximates a solution of the governing equations. The traditional way of dealing with the problem is to have an experimental method and an analytical solution to study problems in fluid dynamics and heat transfer problems. However, with the arrival of digital computers, CFD seems a reliable approach for many engineers when the fluid flows are incredibly complex (Yeoh and Tu, 2010).

Some challenges of using CFD include turbulence modelling and discretisation errors (Li and Nielsen, 2011). For a stable numerical procedure, five significant sources of error are insufficient spatial and temporal discretization convergence, computer round-off error,

insufficient convergence of an iterative procedure and computer programming errors (Balint, 2001; Oberkampf & Trucano, 2002).

2.4.2 Reynolds Number

The flow in a pipeline is characterised as laminar, transitional, or turbulent on the basis of Reynolds number (Re) and is given by Eq. (2.32),

$$Re = \frac{\rho u D}{\mu}$$

(2.32)

where ρ is the fluid density (kg/m^3), u is the velocity of the fluid (m/s), D is the diameter of the pipe (m), and μ is the dynamic viscosity ($Pa \cdot s$).

For *Re* < 2000, flow is assumed to be laminar, 2000< *Re* < 4000 flow is transitional, and *Re* > 4000 flow is turbulent. The flow field in most of the mechanistic models is assumed as turbulent (Nesic et al., 1995; Nesic et al., 2001; Nordsveen et al., 2003; Nešić et al., 2009; Zheng, 2015; Nešić et al., 2019; Kahyarian and Nesic, 2020) as the length of the pipes used for transportation is enormous, and the distance between the oil wells and the station where the separation takes place is in hundreds of kilometres resulting into the fully developed turbulent flow (Nešić, 2007).

2.4.3 Strategy for CFD Modelling

CFD problems are usually solved in three stages, as shown in Figure 2.8. These three stages are the pre-processing stage, numerical solution, and post-processing. In the first stage of CFD, a pre-processing step, the geometry of the given problem is created using the inbuilt techniques. This procedure is followed by prescribing the boundary conditions along the boundary. The governing equations are solved over the volume in 3D and on an area for 2D geometries created. The mesh is generated by breaking this volume into smaller cells or units called grids. The set of equations chosen for the problem is part of the physical flow model. The second stage of CFD consists of the governing equations solved over the geometry volume. For steady-state solutions, the equations are solved iteratively. A step-in time is taken for time-dependent problems, and the equation is numerically solved by giving a solution for the specified time steps. The third and final stage of the CFD process is post-processing. This stage involves the analysis of solutions for dynamic problems with the aid of visualization and animation. In addition to this, the results obtained are compared with any experimental data or available theoretical or analytical solutions. All these stages are described in detail below in Figure 2.8.

Figure 2.8 Stages of CFD modelling (Yeoh and Tu, 2010).

Pre-processing

This is the work that must be completed before running the simulation. It can be reduced to four general areas: geometry definition, volume division, selection of model and boundary condition definition. These are areas described below in detail,

- Geometry definition -

The first and foremost pre-processing step is defining a given problem's computational domain. The crucial point while creating a computational domain is to permit flow to develop fully when it reaches the outlet boundary.

- Volume division -

The second stage is mesh generation or volume division. The discrete values of the flow temperature, velocity, pressure, and other transport parameters are determined in each of these cells obtained numerically. Several cells in the mesh within the computational domain strongly affect the CFD solution. Several factors affect the CFD solution, such as the type of mesh, the adequacy of the techniques chosen for the physics of the problem, and the order of the accuracy of the numerical method. The common orders of accuracy encountered in numerical simulations are first-order, second-order and higher-order accuracies. The order of accuracy is determined by fitting a line to the error vs grid spacing plot, also known as the

convergence plot, followed by calculating the slope of it. It is important to note that the mesh near the wall must be dense to capture the boundary layer development and resolve the small-scale turbulence structures discussed in the next section (Versteeg, 2007).

Selection of Physics and Fluid Properties –

This step deals with the computational and applied mathematical side of CFD. Figure 2.9 gives an overall procedure for selecting appropriate physics for the model based on the categories it falls into.

The final solution gets affected if an inappropriate approximating governing equation is chosen. Therefore, the errors introduced due to simplifications or assumptions are classified as physical approximation errors (Yeoh and Tu, 2010).

Figure 2.9 Selection procedure for appropriate physics (Eslahpazir et al., 2019).

Specification of Boundary Conditions –

To replicate the physical representation of the fluid flow in a CFD problem, it is essential to define all the conditions. A general fluid flow problem consists of an inlet, outlet, and wall as boundary conditions. The fluid's velocity or pressure can be ascertained for inflow boundary conditions. At the outflow boundary, the specified relative pressure will typically be imposed. At the walls, no-slip condition, i.e. zero velocity, is considered. Figure 2.10 shows the boundary conditions for an internal flow problem.

Figure 2.10 Initial values and boundary conditions (Yeoh and Tu, 2010).

There are two boundary conditions: Dirichlet or a direct boundary condition and Neumann or natural boundary condition. The Dirichlet boundary condition is the boundary condition type where the unknown value is specified. In contrast, on the Neumann boundary condition, the value of a derivative of the unknown will be specified. In the above example, the prescription of velocity at the inlet is considered a Dirichlet boundary condition.

The Dirichlet boundary conditions – specifying the velocity on a boundary is given by $u = f(x, y, z)$ at an inlet where u can be a scalar, vector or tensor.

The Neumann boundary conditions – specifying the velocity gradient at a fully developed flow boundary is given by, $\frac{\partial u}{\partial n} = 0$, for (x, y, z) over a domain.

Numerical Solution- Solving the Governing Equations of CFD Modelling

This stage deals with the discretization of the model over the entire domain. The process by which a continuous function or expression is approximated by using an analogous discrete alternative is known as discretization. The continuous functions are the governing partial differential equations for fluid flow problems given in the next section. Analytical solutions of partial differential equations would provide the variation of the dependent variables throughout the domain. Numerical solutions provide answers only at the discrete points in the domain, known as grid points.

The three most popular methods are finite difference, finite volume, and finite element methods. These methods contain variations specific to the application area. Some other discretization methods include high-resolution, boundary element, and spectral methods (Yeoh and Tu, 2010).

Finite Difference Method

The finite difference method is the oldest among the most popular methods. Using Taylor series expansion, 1st and 2nd partial derivatives are replaced by algebraic difference quotients. This method is not used as commonly as the finite volume or finite element method due to its geometric limitation on the applications. In addition, it does not always guarantee

the conservation of mass, momentum, and energy on non-uniform grids. It is also considered the easiest method to code compared to other methods.

Finite Volume Method

The governing equations are solved in the finite volume method over smaller finite control volumes. The fluxes across the volumes are conserved as the governing equations are cast traditionally over each control volume. This method is developed from the finite difference method. It discretizes the solution domain using computational mesh. The finite volume method deals directly with the integral form of N-S equations as an integral form of N-S equations does not require mathematical continuity. The critical advantage of FVM is that the integral conservation will be satisfied over the control volume. This method is more appropriate for applications with strong discontinuities, such as sudden pressure and density changes in the presence of shock waves.

Finite Element Method

The finite element method is more stable than the finite volume and finite difference methods (Yeoh and Tu, 2010). This is a residual method in which a residual equation is weighted and integrated over the entire domain. The domain is partitioned into many mesh elements. Hence, actual integration takes place over each element in the mesh. In terms of applications, tests, validation, and literature, the finite element method is robust in all numerical discretization methods.

There are other discretization methods apart from the ones mentioned above. These methods can be used with functions of varying order. As the order of the functions increases from linear to quadratic, quadratic to cubic, cubic to quartic, etc., computational costs increase. The higher-order functions solve more terms and coefficients in the governing equations, requiring more sample or interpolation points to resolve them correctly. The section below describes the governing equations used for CFD.

Governing Equations for CFD Modelling

The governing equations of fluid dynamics represent mathematical statements of physics conservation laws below.

Conservation of Mass
Conservation of mass law states that the mass is neither created nor destroyed. The net mass flux into the boundary of fixed control volume must equal the increase in mass inside that volume. For a 3-dimensional compressible flow, the continuity equation is given by,

$$\nabla \cdot \boldsymbol{u} = 0 \tag{2.33}$$

where **u** is the velocity vector (m/s) and ∇ is the gradient operator.

Conservation of Momentum
Conservation of momentum law asserts that the sum of total forces acting on the fluid element is equal to the product of mass and the acceleration of the element. For a 3-dimensional steady-state, incompressible flow, the conservation of momentum equation is given as,

Eq. (2.34) below represents the physical significance of the Navier-Stokes equation.

$$\rho(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} \;=\; -\nabla p \;+\; \nabla \cdot \big((\mu + \mu_t)(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T)\big) \tag{2.34}$$

The term on the left-hand side is convection, also known as advection or inertia. The first term on the right-hand side (RHS) is a pressure gradient (Pa), known as source/sink, which provides energy to the flow. The second term on the RHS describes the effect of viscosity $(\mathrm{Pa} \cdot \mathrm{s})$ on the transport of fluid. μ_t is turbulent dynamic viscosity in $(\mathrm{Pa} \cdot \mathrm{s})$.

Post-processing Stage

This stage involves the post-processing of the results obtained from the CFD simulation. It includes three steps – verification, validation and interpretation of results. There is a distinction between verification and validation, as one refers to "solving the equations right" (verification) and the other to "solving the right equations" (validation) (Roache, 1998).

In the verification stage, the accuracy of CFD model implementation is determined by quantifying the error, including round-off error, iterative convergence error and discretization error. Different levels of machine accuracies (single precision, double precision) are used to quantify the round-off error. Investigating the effect of systematic variation of truncation error related to the selection of first or second-order discretization schemes and convergence criteria is used to quantify iterative convergence error. The discretization error is quantified with the help of mesh convergence study, often related to the investigation of systematic refinement of space and time meshes.

The validation stage involves the quantification of input uncertainty and physical model uncertainty. As the name suggests, the input uncertainty is related to the sensitivity analysis of inputs when varied individually (sensitivity) or all together (uncertainty). On the other hand, physical model uncertainty is quantified by comparing the results from CFD simulation to the results of the high-quality experiments. The graph of target quantity from CFD and experimental data is used to support the validation of CFD simulation results. The more robust validation of the CFD model is considered when the difference between the CFD simulation results and experimental results is lower than the validation uncertainty (Roache, 1998).

2.4.4   Turbulence Modelling

Turbulence appears in the swirling fluid structures, also known as eddies, which have a massive range of lengths and time scales. These eddies interact in a complex way. These turbulent eddies could be as large as in metres or as small as micrometres. The largest eddies interact with the flow to extract the energy from it and then stretch to create smaller eddies, producing even smaller eddies. During this process, kinetic energy transfer is known as the energy cascade. This process is shown in Figure 2.11.

Figure 2.11 Turbulent eddies visualisation (Aburebaiya, 2020).

The turbulent flow calculation methods are grouped into three categories: Direct numerical simulation (DNS), large-eddy simulation (LES), and turbulent models for Reynolds-averaged Navier-Stokes equations (RANS). Direct numerical simulation (DNS) computes all turbulent velocity fluctuations and mean flow. Fine spatial grids are required to solve the unsteady Navier-Stokes equations. These fine spatial grids must resolve the most minor length scales of motion, also known as Kolmogorov length scales. The time steps to resolve the period of fluctuations are sufficiently small to match the characteristic time scale to the swirling motion of the most miniature eddy in turbulent flow; for example, Eggels et al. (1994) used a time-step of $0.0002t^*$, where t^* is the ratio of the diameter of the pipe and friction velocity. The total computational time required for studying fully turbulent flow for Re = 7000 was 160 CPU hours. Hence, these simulations are very costly in terms of computing resources (Versteeg and Malalasekara, 1995). The large-eddy simulation tracks the behaviour of the large eddies. The smaller eddies are rejected, and only larger eddies are allowed by using space filtering of the unsteady Navier-Stokes equations before the simulation. A sub-grid scale model includes the effects on the resolved flow due to unresolved and smallest eddies. The computer resources required for this method are enormous (Yeoh and Tu, 2010). Reynolds-Averaged Navier-Stokes (RANS) equations focus on the mean flow and the effect of turbulence on the properties of the mean flow (Versteeg and Malalasekara, 1995). The Navier-Stokes equations are time-averaged before the application of computational methods. The velocity component is decomposed as,

$$u = \bar{u} + u'$$
(2.35)

Where $\bar{u}$ and u' are the mean and fluctuating components of velocity.

For scalar quantity (ϕ) such as pressure, energy, or concentration of species, decomposition is given as,

$$\phi = \bar{\phi} + \phi'$$
(2.36)

Due to the interaction between various turbulent fluctuations, extra terms appear in Reynolds-averaged flow equations. The classical models are used to model these extra terms (Versteeg, 2007). The Reynolds-Averaged-Navier-Stokes equations:

$$\rho \frac{\partial \overline{\boldsymbol{u}}}{\partial t} + \rho \overline{\boldsymbol{u}} \cdot \nabla \, \overline{\boldsymbol{u}} + \nabla \cdot (\overline{\rho \, \boldsymbol{u}' \otimes \boldsymbol{u}'}) = -\nabla P + \nabla \cdot \mu (\nabla \overline{\boldsymbol{u}} + (\nabla \overline{\boldsymbol{u}})^T) \tag{2.37}$$

where $\otimes$ is the vector product. An additional term is on the left-hand side compared with Eq. (2.37). This term is known as Reynolds stress tensor, and it represents interaction occurring between the fluctuating components of the velocity field (Versteeg, 2007). This term is modelled to close the above equation by turbulence modelling as they consider the effects of turbulence on the mean flow. The available turbulence models differ based on the Reynolds stress tensor being modelled (Versteeg, 2007).

Turbulence is assumed to be pure diffusive, which is one of the most common ways to model it (Tu et al., 2018). The varying part of the additional term (Reynolds stress) is then shown as,

$$\rho(\overline{\boldsymbol{u}' \otimes \boldsymbol{u}'}) - \frac{\rho}{3} trace((\overline{\boldsymbol{u}' \otimes \boldsymbol{u}'}))\boldsymbol{I} = -\mu_T (\nabla \overline{\boldsymbol{u}} + (\nabla \overline{\boldsymbol{u}})^T) \tag{2.38}$$

where the second term on LHS is a spherical part written as,

$$\frac{\rho}{3} trace((\overline{\boldsymbol{u}' \otimes \boldsymbol{u}'}))\boldsymbol{I} = \frac{2}{3}\rho k \tag{2.39}$$

where k is the turbulent kinetic energy.

The Standard $k - \varepsilon$ Model

The Standard $k - \varepsilon$ model has shown good performance for different flow situations and hence is recognised as a popular turbulence model. This is one of the oldest models which solves two extra transport equations for k and ε (Launder and Spalding, 1983)

The transport equation for turbulent kinetic energy is given as,

$$\rho \frac{\partial k}{\partial t} + \rho \boldsymbol{u} \cdot \nabla k = \nabla \cdot \left(\left(\mu + \frac{\mu_T}{\sigma_k} \right) \nabla k \right) + P_k - \rho \varepsilon \tag{2.40}$$

P_k is a production term that can be written as,

$$P_k = \mu_T \left(\nabla \boldsymbol{u} : (\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T) - \frac{2}{3}(\nabla . \boldsymbol{u})^2 \right) - \frac{2}{3}\rho k \nabla . \boldsymbol{u} \tag{2.41}$$

The transport equations for dissipation rate are given as,

$$\rho \frac{\partial \varepsilon}{\partial t} + \rho \boldsymbol{u} . \nabla \varepsilon = \nabla \cdot \left(\left(\mu + \frac{\mu_T}{\sigma_\varepsilon} \right) \nabla \varepsilon \right) + C_{\varepsilon 1} \frac{\varepsilon}{k} P_k - C_{\varepsilon 2} \rho \frac{\varepsilon^2}{k} \tag{2.42}$$

The values of the constants used in the above equation are $C_\mu = 0.09, C_{\varepsilon 1} = 1.44, C_{\varepsilon 2} = 1.92, \sigma_k = 1.00$ and $\sigma_\varepsilon = 1.30$. These standard values are used for homogeneous systems (Wilcox 2013).

The turbulent variables are obtained from the turbulent length scale L_T and turbulence intensity I_T when the inlet data is unavailable. A value of I_T can be specified between 0.001 (0.1%) as low intensity and 0.1 (10%) as high intensity for fully turbulent flow. The turbulence length scale measures the size of unresolved eddies (Versteeg, 2007). For a fully developed flow in a pipe,

$$L_T = 0.07 \, D \tag{2.43}$$

The values of the turbulence length scale L_T and turbulence intensity I_T are used to obtain turbulent kinetic energy k and viscous dissipation rate ε,

$$k = \frac{3}{2}(uI_T)^2 \tag{2.44}$$

$$\varepsilon = \frac{3}{2}C_\mu^{\frac{3}{4}}\frac{k^{\frac{3}{2}}}{L_T} \tag{2.45}$$

The standard $k - \varepsilon$ model provides good results while predicting boundary layers, thin layers and duct flows (Tu et al., 2018). However, it performs poorly when the problem involves flow separation (Veersteg and Malalasekara, 2016).

The $k - \omega$ Model

Wilcox (Wilcox, 1998) proposed the $k - \omega$ model, which has shown the capability to handle near-wall regions. This model solves two transport equations for turbulence kinetic energy k and specific dissipation rate ω.

$$\rho\frac{\partial k}{\partial t} + \rho\boldsymbol{u}.\nabla k = P_k - \rho\beta^*k\omega + \nabla\cdot\left(\left(\mu+\frac{\mu_T}{\sigma_k}\right)\nabla k\right) \tag{2.46}$$

$$\rho\frac{\partial \omega}{\partial t} + \rho\boldsymbol{u}.\nabla\omega = \alpha\frac{\omega}{k}P_k - \rho\omega^2\beta + \nabla\cdot\left(\left(\mu+\frac{\mu_T}{\sigma_k}\right)\nabla k\right) \tag{2.47}$$

where the eddy viscosity is calculated by,

$$\mu_T = \rho\frac{k}{\omega} \tag{2.48}$$

and values are,

$$\alpha = \frac{13}{25} \quad \beta = \beta_0 f_\beta \quad \beta^* = \beta_0^* f_\beta \quad \sigma = \frac{1}{2} \quad \sigma^* = \frac{1}{2}$$

$$\beta_0 = \frac{13}{125}, \quad f_\beta = \frac{1+70\chi_\omega}{1+80\chi_\omega}, \quad \chi_\omega = \left|\frac{\Omega_{ij}\Omega_{jk}S_{ki}}{(\beta_0^*\omega)^3}\right| \quad \beta_0^* = \frac{9}{100},$$

$$f_{\beta^*} = \begin{cases} 1 & \chi_k \le 0 \\ \dfrac{1+680\chi_k{}^2}{1+400\chi_k{}^2} & \chi_k > 0 \end{cases}, \quad \chi_k = \frac{1}{\omega^3}(\nabla k\cdot\nabla\omega)$$

The mean rotation-rate tensor Ω_{ij},

$$\Omega_{ij} = \frac{1}{2}(\nabla\boldsymbol{u} - (\nabla\boldsymbol{u})^T) \tag{2.49}$$

The mean strain-rate tensor S_{ij},

$$S_{ij} = \frac{1}{2}(\nabla\mathbf{u} + (\nabla\mathbf{u})^T) \tag{2.50}$$

$$\varepsilon = \beta^* \omega k \tag{2.51}$$

and

$$l_{mix} = \frac{\sqrt{k}}{\omega} \tag{2.52}$$

A combined model of the $k - \omega$ model and standard $k - \varepsilon$ model was proposed by Menter (1994) and discussed in detail in the following subsection.

The Shear Stress Transport $k - \omega$ Turbulence Model

The Shear Stress Transport (SST) $k - \omega$ model was introduced by Menter (1994). It combines the precise behaviour of the $k - \omega$ model in the region near the wall with the robustness of the $k - \varepsilon$ model away from the wall. In this turbulence model, the k-equation and the Reynolds stress computation are the same as the $k - \omega$ model. It differs in the calculation of a ω equation as the $k - \varepsilon$ equation is transformed by putting $\varepsilon = k\omega$.

The equations used in the SST $k - \omega$ models are given by (COMSOL, 2016),

$$\rho \frac{\partial k}{\partial t} + \rho \cdot \nabla k = P - \rho \beta_0^* k\omega + \nabla \cdot ((\mu + \sigma_k \mu_T)\nabla k) \tag{2.53}$$

$$\rho \frac{\partial \omega}{\partial t} + \rho \boldsymbol{u} \cdot \nabla \omega = \frac{\rho \gamma}{\mu_T} P - \rho \beta \omega^2 + \nabla \cdot ((\mu + \sigma_\omega \mu_T)\nabla \omega)$$
$$+ 2(1 - f_{v1}) \frac{\rho \sigma_{\omega^2}}{\omega} \nabla \omega \cdot \nabla k \tag{2.54}$$

$$P = min(P_k, 10\rho \beta_0^* k\omega) \tag{2.55}$$

where, P_k is the production term given as,

$$P_k = \mu_T \left(\nabla \boldsymbol{u} : (\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T) - \frac{2}{3}(\nabla \cdot \boldsymbol{u})^2 \right) - \frac{2}{3} \rho k \nabla \cdot \boldsymbol{u} \tag{2.56}$$

The turbulent viscosity (μ_T) is given by,

$$\mu_T = \frac{\rho a_1 k}{max(a_1 \omega, S f_{v2})} \tag{2.57}$$

where k is the turbulent kinetic energy (m^2/s), ω is the specific turbulent dissipation rate (1/s), a_1 is a model constant with a value of 0.31, S is the mean strain-rate tensor, and f_{v2} is a blending function.

$$S = \sqrt{2S_{ij}S_{ij}} \tag{2.58}$$

The constants in the model are defined through interpolation of inner and outer values,

$$\phi = f_{v1}\phi_1 + (1 - f_{v1})\phi_2 \text{ for } \phi = \beta, \gamma, \sigma_k, \sigma_\omega \tag{2.59}$$

where f_{v1} and f_{v2} are interpolation functions defined as,

$$f_{v1} = tanh(\theta_1^4) \tag{2.60}$$

$$\theta_1 = min \left[max \left(\frac{\sqrt{k}}{\beta_0^* \omega l_\omega}, \frac{500\mu}{\rho \omega l_\omega^2} \right), \frac{4\rho \sigma_{\omega^2} k}{CD_{k\omega} l_\omega^2} \right] \tag{2.61}$$

$$CD_{k\omega} = max \left(\frac{2\rho \sigma_{\omega^2}}{\omega} \nabla \omega \cdot \nabla k, 10^{-10} \right) \tag{2.62}$$

and,

$$f_{v2} = tanh(\theta_2^2) \tag{2.63}$$

$$\theta_2 = max\left(\frac{2\sqrt{k}}{\beta_0^* \omega l_\omega}, \frac{500\mu}{\rho \omega l_\omega^2}\right) \tag{2.64}$$

where l_ω is the distance closest to the wall. The default values of constants are $\beta_1 = 0.075, \beta_2 = 0.0828, \gamma_1 = \frac{5}{9}, \gamma_2 = 0.44, \sigma_{k1} = 0.85, \sigma_{k2} = 1.0, \sigma_{\omega1} = 0.5, \sigma_{\omega2} = 0.856, \beta_0^* = 0.09, a_1 = 0.31$

As described above, the SST model is a low-Reynolds model; hence, the equations are integrated through the boundary layer to the wall. This allows for a no-slip condition at the wall. The velocity **u** becomes zero at the wall, so *k* must be zero at the wall.

Similarly, the boundary condition for ω is

$$\lim_{l_\omega \to 0} \omega = \frac{6\mu}{\rho \beta_1 l_\omega^2} \tag{2.65}$$

At the wall, to avoid the singularity ω is not solved in the cells. Hence, its value is given below,

$$\lim_{l_\omega \to 0} \omega = \frac{6\mu}{\rho \beta_0 l_\omega^2} \tag{2.66}$$

To achieve an accurate solution, it is required,

$$l_\omega^+ = \frac{\rho u_\tau l_\omega}{\mu} \sim 1 \tag{2.67}$$

For turbulence models with wall functions, COMSOL Multiphysics (COMSOL, 2016) assumes that the thickness of the buffer layer is minimal and hence uses an approximation in this region.

The Spalart-Allmaras Model

The Spalart-Allmaras model is a one-equation model designed for aerospace applications involving wall-bounded flows (Spalart and Allmaras, 1994). This model is a low-Reynolds-number model. It has been calibrated for use in airfoil and turbine blade applications, for which it provides satisfactory results; however, one of the nonphysical properties of this model is that in a uniform free stream, it usually predicts zero decay rate for the eddy viscosity (Wilcox, 1998).

This model solves for undamped eddy kinematic viscosity, $\tilde{v}$ given below,

$$\frac{\partial \tilde{v}}{\partial t} + \boldsymbol{u} \cdot \nabla \tilde{v} = c_{b1}\tilde{S}\tilde{v} - c_{wl}f_w \left(\frac{\tilde{v}}{l_w}\right)^2 + \frac{1}{\sigma}\nabla \cdot \left((v + \tilde{v})\nabla\tilde{v}\right) + \frac{c_{b1}}{\sigma}\nabla\tilde{v} \cdot \nabla\tilde{v} \tag{2.68}$$

The auxiliary variables are,

$$c_{wl} = \frac{c_{b1}}{k_v^2} + \frac{1 + c_{b2}}{\sigma} \tag{2.69}$$

$$\chi = \frac{\tilde{v}}{v} \tag{2.70}$$

$$f_{v1} = \frac{\chi^3}{\chi^3 + c_{v1}^3} \tag{2.71}$$

$$f_{v2} = 1 - \frac{\chi}{1 + \chi f_{v1}} \tag{2.72}$$

$$f_w = g \left(\frac{1 + c_{w3}^6}{g^6 + c_{w3}^6} \right)^{\frac{1}{6}} \tag{2.73}$$

$$g = r + c_{w2}(r^6 - r) \tag{2.74}$$

$$r = min \left(\frac{\tilde{v}}{\tilde{S} k_v^2 l_w^2}, 10 \right) \tag{2.75}$$

$$\tilde{S} = max \left(\Omega + C_{Rot} \, min(0, S - \Omega) + \frac{\tilde{v}}{k_v^2 l_w^2} f_{v2}, 0.3\Omega \right) \tag{2.76}$$

$$S = \sqrt{2 S_{ij} S_{ij}} \tag{2.77}$$

$$\Omega = \sqrt{2 \Omega_{ij} \Omega_{ij}} \tag{2.78}$$

$$S_{ij} = (\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T) \tag{2.79}$$

$$\Omega_{ij} = \frac{1}{2} (\nabla \boldsymbol{u} - (\nabla \boldsymbol{u})^T) \tag{2.80}$$

where S_{ij} are the mean strain rate, Ω_{ij} are the rotation rate tensors, v is the kinematic viscosity, l_w is the distance to the closest wall.

The turbulent viscosity (μ_T) is given by,

$$\mu_T = \rho \tilde{v} f_{v1} \tag{2.81}$$

The parameter values are $c_{b1} = 0.1355, c_{b2} = 0.622, c_{v2} = 7.1, \sigma = \frac{2}{3}, \ c_{w2} = 0.3, c_{w3} = 2, k_v = 0.41, C_{rot} = 2.0$.

2.4.5 Turbulent Boundary Layer Profile

The no-slip wall condition is essential in turbulent flows since tangential velocity fluctuations near the flat plate surface are decreased by viscous damping (Stanković et al., 2014). The large gradients in mean velocity cause a rapid increase in turbulence as turbulent kinetic energy is produced away from the wall (Veersteg and Malalasekara, 2016). Figure 2.12 shows the velocity profile over the length of a flat plate. The laminar boundary layer develops when the flow with uniform velocity hits the leading edge of the flat plate. As discussed above, the velocity of the fluid is zero at the surface of a flat plate and increases linearly in the viscous sublayer near the surface. Adjacent to the viscous sublayer, the buffer layer exists where turbulent stresses dominate over viscous stresses. The average flow velocity in that region is proportional to the log of the distance to the surface. This is called the log-law region.

Figure 2.12 Understanding the formation of the boundary layer on a flat plate (not to scale).

The formation of near-wall regions for pipe flow is plotted in semi-log coordinates shown in Figure 2.13. The dimensional distance from the wall represents the horizontal axis y^+ and the vertical axis is represented by u^+. The y^+ and u^+ values are obtained using,

$$y^+ = \frac{y u_\tau \rho}{\mu} \tag{2.82}$$

and

$$u^+ = \frac{u}{u_\tau} \tag{2.83}$$

$$u_\tau = \sqrt{\frac{\tau_w}{\rho}} \tag{2.84}$$

where, u_τ frictional velocity and τ_w is the wall shear stress of fluid.

The velocity profile is primarily divided into parts: inner and outer layers. The inner layer is divided into three regions: viscous sublayer up to $y^+ = 5$, buffer layer $5 < y^+ < 30$ and fully turbulent region, also known as a log-law region for $y^+ > 30$. The upper limit of the axes depends on the Reynolds number of the flow.

Figure 2.13 Subdivisions of regions near-wall in a pipe (not to scale).

2.5   CFD Modelling in the Area of Corrosion

There has been a significant increase in the use of CFD modelling for corrosion or erosion-corrosion prediction models over the last couple of decades. In elbows, CFD is used to predict hydrodynamic flow fields and then coupled to the erosion-corrosion model to determine the significance of bend orientation (Keating and Nesic, 2000) and flow velocity (Bozzini et al., 2003) in multiphase flows. Keating and Nesic (2000) used a standard two-equation $k - \varepsilon$ model and Lam and Bremhorst (1981) low Reynolds modification to predict mass transfer in axisymmetric sudden expansion geometry. However, upon verification of the results with the experimental data, it was found that Lam and Bremhorst (1981) near-wall model showed poor agreement and the highest deviation was observed at the reattachment point. Although there was a significant deviation in the mass transfer results, the authors ignored this issue, expecting that there would not be any separation and reattachment in bends. On the other hand, some studies did not resolve flow near the wall in bends; instead, they used the standard two-equation $k - \varepsilon$ model and standard wall functions for the near-wall treatment (Bozzini et al., 2003; El-Gammal et al., 2010).

Flow accelerated corrosion (FAC) has been studied using both experimental and CFD techniques at Re = 40000 (El-Gammal et al., 2010) to explore the importance of parameters such as surface roughness, wall shear stress, upstream turbulence, and the role of geometry on the prediction of local mass transfer coefficients (Pietralik, 2012). El-Gammal et al. (2010) performed experiments to quantify the time evolution of wall wear patterns and CFD to simulate flow hydrodynamics in a 90-degree elbow. CFD simulations aided in obtaining wall shear stress along the elbow wall and are presented in terms of the coefficient of friction shown in Figure 2.14. It was found that the increased levels of skin friction along the elbow intrados resulted in maximum wear in the entrance region.

Figure 2.14 Contours plot of skin friction coefficient along the elbow wall between a fluid and the surface (diameter = 25.4mm) (El-Gammal et al., 2010).

Zhang and Cheng (2010) studied the FAC behaviour of X65 pipeline steel in CO_2 − saturated formation waters with the help of electrochemical measurements and CFD simulations in which micro-electrodes were installed on an impingement jet test system. Their study modelled steady-state – incompressible fluid flow with the standard $k − \varepsilon$ turbulence model to determine the flow distribution at different impact angles and velocities. The effect of impact angle on steel corrosion was then attributed to the flow and shear distribution on electrode surfaces. In some studies, CFD approximated the correlations for the mass transfer coefficient when no empirical equations were available at high Reynolds numbers (Pietralik, 2012). This shows one of the remarkable capabilities of CFD in the area of corrosion.

In pipeline flows, CFD simulations were used to understand the mechanism of slug flow-induced CO_2 corrosion (Zheng et al., 2007; Lv et al., 2020). The shear stress and mass transfer coefficient in the upward slug flow increased as the superficial velocities of the gas and liquid phase increased (Lv et al., 2020). This increased the 316L SS corrosion rate. These studies only focused on typical slug flows while ignoring the corrosion that could occur in transition processes where superficial velocities of liquid and gas vary.

CFD located the most probable corrosion sites based on water volume fraction (El-Batsh et al., 2012) and water accumulating regions (Hu and Cheng, 2016). El-Batsh et al. (2012) modelled a two-phase liquid-liquid flow using a mixture model (Manninen et al., 1996) and an SST $k − \omega$ turbulence model to resolve viscous sublayer in complex pipeline geometry that included six elbows. The model showed reasonable accuracy upon comparison with the experimental fluid flow data and concentration distribution. The reduction in elbow thickness was linked with the volume fraction distribution of water near the wall obtained using CFD. Hu and Cheng (2016) modelled two-phase oil-water flow in straight and inclined pipelines using the volume of fluid (VOF) method, realizable $k − \varepsilon$ turbulence model and standard wall functions. They obtained the volume fraction of water and shear stress distribution across the domain and coupled it with the empirical corrosion prediction model (Kanwar, 1994). The correlation for the prediction of corrosion rate is given as

$$CR = k \, pCO_2{}^c \tau^b \tag{2.85}$$

where CR is corrosion rate (mm/yr), k is the constant equal to 15.5 ± 0.5, pCO_2 is the partial pressure of CO_2 (MPa), τ is wall shear stress (N/m^2), c and b are the constants equal to 0.83 ± 0.07 and 0.1, respectively.

It is important to note that this empirical model is valid for oil content below 60% and temperatures less than 60℃. For oil content above 70%, another empirical equation was used for the corrosion prediction (Jepson et al., 1996),

$$CR = 31.15 \left(\frac{\Delta P}{L}\right)^{0.3} v^{1.6} pCO_2{}^{0.8} e^{\left(-\frac{2671}{T}\right)} \tag{2.86}$$

A similar approach was adopted by Li et al. (2016) to predict CO_2 corrosion rates using CFD based on the semi-empirical corrosion model of Jepson et al. (1996). CFD simulations were carried out to obtain pressure drop gradient terms for different oil contents required in Eq. (2.86), and results were compared with the experimental data. It was found that there was a relatively good agreement between the predictions and experimental data; however, due to the semi-empirical nature of the model, the authors mentioned that there is specific applicability related to corrosion environments and operating conditions (Li et al., 2016).

Some studies coupled the mechanistic model with CFD to predict corrosion rates in a pipeline (Wang, 1999; Li and Woollam, 2012; Prasad et al., 2018; Hu et al., 2018). Li and Woollam (2012) used CFD and a mechanistic model to predict corrosion rates in the sharp bend geometry of disturbed flow. A steady-state incompressible flow was modelled with the help of RANS and the Abe-Kondoh-Nagano $k-\varepsilon$ turbulence model (Abe et al., 1994) to resolve the viscous sublayer. The study was remarkable in highlighting the cathodic and anodic regions based on the potential distribution inside a bend, as shown in Figure 2.15. The cathodic regions were developed downstream of the bend, where the higher potential and current density values were seen. They found higher corrosion rates near the bend as the mass transfer experienced was maximum.

Figure 2.15 Current and potential distribution inside a bend. Surface – electric potential (V) and Arrows - Current density (A) (Li and Woollam, 2012).

Chang et al. (2014) modelled a water droplet on the metal surface surrounded by air, known as Evan's drop (Evans, 1960) and a solid deposit surrounded by water in a pipeline to model under-deposit corrosion. In their mathematical model, the cathodic and anodic sites were not defined prior and were the result of the formation of precipitates. However, the limitation of this model is that it only considers aerated electrolyte systems, and it needs to be explored for the systems relevant to the oil and gas industries, where deaerated electrolytes play a significant role. A similar methodology was developed by Sainz-Rosales et al. (2022) to study Evan's drop (Evans, 1960) with the help of tertiary current distribution to obtain oxygen concentration distribution between the centre of the droplet and its periphery, which set up potential differences across it. The value of potential difference was verified against the experiment data of the kelvin probe technique (Chen and Mansfeld, 1997).

Prasad et al. (2018) coupled the mechanistic model (Sanchez-Caldera et al., 1988) and CFD to predict FAC by considering the reduction in wall thickness. In their study, the mass transfer coefficient (MTC) was computed using CFD in bend and orifice geometries. The SST $k - \omega$ turbulence model was used to resolve the viscous sublayer in their study. The thinning of the wall was calculated based on the mass transfer coefficient obtained from CFD simulations. They observed a significant difference between the CFD-predicted and experimental wall thickness values linked to how the ferrous iron formation was modelled.

Wang (1999) coupled the fluid flow model with the 1D numerical model for the prediction of CO_2 corrosion in pipelines. The corrosion rate predictions were based on mixed potential theory, and the straight pipe correlation of Berger and Hau (1977) was used for overall mass transfer coefficients in a simplified electrochemical model. Although an attempt was made to couple the flow and electrochemical models, the corrosion rate predictions showed discrepancies upon comparison with experimental data. It was assumed that VSL exists for $y^+ < 5$, and the discrepancies were not explored in detail. Some of the recent research

36

includes the prediction of CO_2 corrosion in the straight pipe and jet impingement using CFD, which over-predicted uniform corrosion rate for pH 4 (Hu et al., 2018). This study used the AKN $k - \varepsilon$ turbulence model (Abe et al., 1994) for the mass transfer model's fluid flow simulations with different electrochemical reaction rate constants.

<u>Research Gap</u>

The literature review has shown that there is a lack of methodology which accurately predicts VSL and turbulent diffusivity profile and then subsequently couples it to the mass transfer model of CO_2 corrosion rate. In addition to this, the predictive tools available are applicable for fully developed flow, resulting in uniform corrosion. The current study is focused on the prediction of VSL and turbulent diffusivity profile with the help of CFD and then using these predictions to drive the 1D mass transfer model for the prediction of CO_2 corrosion rate. The reliable 1D mass transfer model for the prediction of CO_2 the corrosion rate is then applied to construct CFD driven mass transfer model for the prediction of corrosion rate in gradual constriction pipe that will account for the variation in viscous sublayer thickness and turbulent diffusivity profile along the length of the pipe.

2.6   Design Optimisation

Design optimisation is finding the optimum design without violating certain constraints. The objective function is either minimised or maximised. In the current study, a difference between experimental corrosion rates (Nesic et al., 1995) and predicted corrosion rates (Thorat et al.,2024) is considered as an objective function ($f(\mathbf{x})$), where **x** is a design variable. The 12 design variables used in the current study are electrochemical reaction rate constants, which include reversible potential (E_{rev}), Tafel slope (b), activation energy (ΔH), power constant (a_1), reference exchange current density $\left(i_{0,ref}\right)$ and reference concentration of species ($cCO_{2,ref}$). A design optimisation methodology is implemented in this study to find an optimal set of electrochemical reaction rate constants to build a robust mass transfer model.

The standard optimisation problem is given below,

Minimise:

$$\text{The cost function: } f(\boldsymbol{x}) = f(x_1, x_2, \dots, x_n) \tag{2.87}$$

Subject to:

$$g_i(\boldsymbol{x}) \leq 0; \ i = 1 \ to \ m \tag{2.88}$$
$$h_j(\boldsymbol{x}) = 0; \ j = 1 \ to \ p \tag{2.89}$$

The design space is,

$$x_{lower} \leq x_i \leq x_{upper} \tag{2.90}$$

where, $g_i(x)$ are the inequality constraints, $h_j(x)$ are the equality constraints and x_{lower} and x_{upper} represent lower and upper limits for each of the design variables.

Figure 2.16 below gives the framework of the surrogate-based optimisation implemented in the current study.

Figure 2.16 The framework of surrogate-based optimisation to find out optimal parameters for operating conditions and electrochemical reaction rate constants.

2.6.1   Design of Experiments

Sampling the design space can be directly linked to obtaining maximum information from a limited number of samples, as the simulations or experiments are expensive or time-consuming (Yondo et al., 2018). The techniques that are used for the planning of the experiments include one factor at a time (OFAT), trial and error approach (Montgomery, 2013), and design of experiments (DoE) (Fisher, 1937). Yondo et al. (2018) classified DoE sampling techniques into classical DoE (Montgomery, 2013; Antony, 2014) and Modern DoE (Sacks et al., 1989; Koehler and Owen, 1996; Crary, 2002; Bursztyn and Steinberg, 2006) shown in Figure 2.17.

Figure 2.17 Classification of design of experiment techniques (Yondo et al., 2018).

Classical DoE includes factorial designs (Mukerjee and Wu, 2006; Montgomery, 2013), 'Box-Wilson' Central Composite Designs (CCD) (Box and Wilson, 1951), optimal designs (Kiefer and Wolfowitz, 1959), and orthogonal arrays experiments (Taguchi Methods) (Rao, 1947; Taguchi, 1959; Taguchi, 1989).

Figure 2.18 shows the sketch of the classical design of experiments. There are two types of factorial designs: full and fractional (Yondo et al., 2018). A factorial design sets points at the hypercube vertices, and as the number of factors increases, the sample points spread (Mukerjee and Wu, 2006; Montgomery, 2013). This methodology is considered complex for the design variables greater than three, e.g., 3-levels (3^k), k is the number of design variables (Yondo et al., 2018). CCDs (Box and Wilson, 1951) are developed to fit full quadratic objective functions with the advantage of having adaptive sampling features. The optimal designs are assessed based on the variance matrix included in the surrogate model (Kiefer and Wolfowitz, 1959). Orthogonal arrays (Rao, 1947; Taguchi, 1959; Taguchi, 1989) are fractional factorial designs with the advantage of focusing on controllable factors' impact on the objective function responses (Yondo et al., 2018). These designs are represented using cubes; each dimension depicts a factor, and the levels of each factor are displayed along that dimension (Yondo et al., 2018).

Figure 2.18 Sketch of classical design of experiments (Veldhuis et al., 2016).

Modern DoE includes random sampling (Gentle, 2003), quasi-random designs (Lemieux, 2009; Sobol' et al., 2011), projections-based designs (Kacker et al., 1991), uniform designs (Fang and Li, 2006), and some miscellaneous design techniques that cannot be categorized in either of these two classifications.

Random sampling generates sampling points with the likelihood of appearing in any sample with the same probability. Simple random sampling and Monte Carlo methods such as Gibbs sampling, Markov Chain Monte Carlo – MCMC Samplings, Metropolis-Hastings Sampling, etc., are popular (Kim et al., 2000). The disadvantage of this technique is that it needs to have a deep knowledge of the entire design space without missing data (Yondo et al., 2018). Quasi-random designs (Lemieux, 2009; Sobol' et al., 2011), often known as low discrepancy sequences generators, provide a high level of uniformity in a multidimensional space (Yondo et al., 2018). Uniform designs (Fang and Li, 2006) are space-filling designs based on the measure of uniformity $\mathcal{M}$. For the entire design space χ, a design $\xi \in \chi$ is a uniform design if the value of $\mathcal{M}$ is minimised over ξ (Yondo et al., 2018).

Latin Hypercube Sampling (LHS)

Latin Hypercube sampling (LHS) (McKay et al., 1979) comes under the miscellaneous design sampling techniques (Yondo et al., 2018), in which **N** design variables are divided into equal **P** levels. For LHS DoE, it is made sure that there is only one sample point for each level. LHS DoE is classified into random Latin Hypercube (RLH) and optimal Latin Hypercube (OLH) sampling. This classification depends on how the sampling points are distributed in the design space (Bates et al., 2004). The RLH method utilizes randomness to obtain each sample point (Bates et al., 2004). OLH samples used in this study are generated with the help of the permutation genetic algorithm (permGA) and based on the methodology developed by (Bates et al., 2004). This methodology solves a combinatorial optimisation problem of minimisation of the

potential energy of repulsive forces for a set of sample points with a unit mass (Bates et al., 2004). The magnitude of these repulsive forces depends on the square of the distance between the points and is given by,

$$\sum_{p=1}^{P} \sum_{q=p+1}^{P} \frac{1}{r_{pq}^2} \rightarrow min \qquad (2.91)$$

where r is the distance between sample points p and q.

The function given in Equation (2.91) is known as the Audze-Eglais objective function (AELH) (Bates et al., 2004). The coordinates method is used to encode AELH. The formulation is done in such a way that the first X numbers are a random sequence between 1 and X, repeating this up to N without the repetitions of numbers in each sequence. Hence, the rule of LHS of having a point in each level is secured. The process of using genetic operators with a permutation encoding starts with Mutation, as given below,

$$[5\ \mathbf{1}\ 4\ 2\ \mathbf{3}] \Longrightarrow [5\ \mathbf{3}\ 4\ 2\ \mathbf{1}]$$

In the above step, two numbers are exchanged. A crossover step follows this, applied to all the sequences of X numbers up to N design variables. Different methods can be used for crossover, such as simple crossover, cycle crossover, and inversion. Simple crossover involves the selection of a crossover point up to which a permutation is copied from the first parent, followed by scanning the second parent from the beginning.

$$\text{Parent } 1 = [\mathbf{5}\ \mathbf{2}\ 1\ 4\ 3] \quad \Longrightarrow \quad \text{Child } 1 = [\mathbf{5}\ \mathbf{2}\ 4\ 1\ 3]$$

$$+$$

$$\text{Parent } 2 = [\mathbf{4}\ \mathbf{1}\ 3\ 5\ 2] \quad \Longrightarrow \quad \text{Child } 2 = [\mathbf{4}\ \mathbf{1}\ 5\ 2\ 3]$$

Cycle crossover maintains the element's positions in the sequence of parents as each value and its position belongs to one of the parents (Bates et al., 2004). The implementation of cycle crossover is given below (Bates et al., 2004),

$$\text{Parent } 1 = [1\ 3\ 9\ 7\ 5\ 4\ 6\ 2\ 8]$$

$$\text{Parent } 2 = [4\ 6\ 2\ 1\ 7\ 8\ 9\ 3\ 5]$$

The process starts with the selection of the first value of parent 1 followed by identifying the value at the same position in parent 2 and then finding the position of that value in parent 1. Thus,

$$\text{Child } 1 = [1 * * * * 4\ * * *]$$

This provides a value of 9 from parent 2 at the 4th position. Thus,

$$\text{Child } 1 = [1 * * * * 4 * * 8]$$

This rule is repeated, and the following values for child 1 are 5 and 7. As the value of 7 is selected, it requires the selection of value 1 from parent 1. Since a value of 1 is used, the cycle is completed.

$$\text{Child } 1 = [1 * * 7\ 5\ 4 * * 8]$$

The final values of child 1 are,

$$\text{Child } 1 = [16\ 2\ 7\ 5\ 4\ 9\ 3\ 8]$$

This process is applied to obtain child 2 as given below,

$$\text{Child } 2 = [4\ 3\ 9\ 1\ 7\ 8\ 6\ 2\ 5l]$$

The inversion crossover method consists of choosing two random points in a parent between which the points are inverted. This is shown below, the cut-off points are 2 and 6, marked by '|'.

$$\text{Parent } 1 = [4\ 2\ |\ 9\ 1\ 7\ 6\ |\ 8\ 5\ 3] \quad \Longrightarrow \quad \text{Child } 1 = [4\ 2\ |\ 6\ 7\ 1\ 9|\ 8\ 5\ 3]$$

The OLH sampling method is used to solve a problem of 4 design variables and 100 sample points in the current study.

2.6.2 Surrogate Models

The use of CFD to solve problems in aerospace, automotive, medical and oil/gas industries has seen a massive increase in recent years (Yeoh and Tu, 2010b). Although advanced computational resources such as high-performance computer clusters and cloud computing resources are available, high-fidelity simulations require hundreds of hours to complete (Yondo et al.,2018). Considering the amount of computational time and computational power required to solve high-fidelity simulations, it is irrational to depend solely on these simulations (Yondo et al., 2018). Box and Draper (1987) introduced surrogate models, also known as metamodels or emulators or response surface models, to overcome the computational restrictions of these simulations. Surrogate models often replicate the response of computationally expensive (high-fidelity) simulations (Yondo et al., 2018). Ahmed and Qin (2009) categorized surrogate models into black-box and physics-based models. The black-box approach models were again classified according to two criteria that involved classifying them based on design variables used: either parametric (Gunst, 1996; Forrester, 2008; Montgomery, 2013) or nonparametric (Wasserman, 1993) and based on a technique used for the construction of surrogate model; either interpolation or regression. This section reviews some of the techniques used for obtaining the response surface.

Artificial Neural Network (ANN)

Artificial neural networks (ANN) are one of the most popular machine-learning techniques for creating data-driven models (Gosmann et al., 2022). In the architecture of ANN shown in Figure 2.19, the input layer consisting of input variables (x_i) is connected to a hidden layer with some weights(w_{ij}). A weighted sum $(F_j = \sum_i w_{ij}x_i)$ of the inputs from the input layer is calculated at each node in the first hidden layer. An activation function is then used to transform F_j and is an output from each node. This output from each node in the first hidden layer becomes the input for the second hidden layer. This process is repeated until the output layer is reached. To minimise the mean squared error between the predicted and target values, weights that connect the nodes are adjusted. Various types of ANN currently exist; out of those, radial basis function neural networks (RBFNN) (Wasserman, 1993) and multilayer perceptrons (MLP) (White, 1992) have been found to have better learning capabilities (Yondo et al., 2018). Figure 2.19 shows the structure of multilayer perceptron neural networks.

Figure 2.19 Structure of multilayer perceptron neural networks (Zhang et al., 2019).

RBFNN, shown in Figure 2.20, consists of three layers of neurons: an input layer, a hidden layer, and an output layer. The weights of this nonlinear surrogate model are computed using nonlinear least squares methods, hybrid methods or maximum likelihood estimation (Myers, 1990; Forrester, 2008; Yondo et al., 2018). The formulation of RBFNN is given below,

$$x_i \mapsto y_i = f(x^{(i)})| \, i = 1,2,3, \dots, n \tag{2.92}$$

where $x = \left[x^{(1)}, x^{(2)}, \dots, x^{(n)}\right]^T$ and $y = [y(x^{(1)}), y(x^{(2)}), \dots, y(x^{(n)})]^T$

A linear predictor of the RBF approximation is (Yondo et al., 2018),

$$\hat{y}_{RBF}(x) = w^T \phi = \sum_{i=1}^{n_c} w^{(i)} \phi(||x - c^{(i)}||) \tag{2.93}$$

where $w^{(i)}$ are the weights, $c^{(i)}$ the ith of the n_c basis function centres and ϕ the n_c- vector containing the values of the basis functions ϕ. These basis functions ϕ are evaluated at the Euclidean distances between the prediction x and centres $c^{(i)}$ of the basis functions (Yondo et al., 2018).

Figure 2.20 Single hidden layer neural network.

Moving Least Squares Method

The moving least squares method (MLSM) is a generalized form of the conventional polynomial weighted least squares method (Choi et al., 2001; Toropov et al., 2005; Loweth et al., 2011). The weights in MLSM are dependent on the Euclidian distance between a sampling point x_i and a prediction point x where the metamodel is evaluated unlike the traditional least squares methods where the weights are constant. The magnitude of the weight function changes as the prediction point x changes its location to provide the moving least squares approximation $\check{f}(x)$ of the original function $f(x)$ (Toropov et al., 2005). The numerical noise in the sampling data can be handled effectively by adjusting the "closeness of fit" (Toropov et al., 2005; Loweth et al., 2011). The "closeness of fit" is adjusted by changing the weight decay parameter. The Gaussian weight decay function, one of the popular weight decay functions (Toropov et al., 2005), is given below,

$$w_i = exp(-\theta r_i^2) \tag{2.94}$$

where θ is the closeness of fit and r_i is the normalised distance between the i-th sampling point and the current point.

Gaussian Process Regression (GPR)

GPR is a probabilistic method that comes under non-parametric regression types of regression models capable of carrying out nonlinear input/output mappings. Rasmussen and Williams (2006) provided an extensive overview of the Gaussian process (GP). They stated that a GP is a generalization of the Gaussian probability distribution. A probability distribution represents

a stochastic process for random variables, either scalar or vectors, in the case of multivariate distributions (Snelson, 2007).

A function $f(\mathbf{x_i})$ is assumed as a GP even before considering the data, where $\mathbf{x_i}$ is an input vector known as a regressor. For y_i, a scalar output value of the function $f(\mathbf{x_i})$ also called targets at locations i corrupted with Gaussian noise ϵ_i given by,

$$y_i = f(\boldsymbol{x_i}) + \epsilon_i \qquad (2.95)$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma_n^2) \qquad (2.96)$$

where $\boldsymbol{x_i}$ are input vectors ($\boldsymbol{x_i} \in \mathbb{R}^{D \times 1}$) and y_i are scalar outputs ($y_i \in \mathbb{R}$).

A mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x'})$ can be used to fully define a GP in a function-space view (Rasmussen and Williams, 2006),

$$f(\boldsymbol{x}) = \mathcal{N}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x'})) \qquad (2.97)$$

A data set $D = \{X, \boldsymbol{y}\}$ represents all the available data for N number of sample points.

Where $X = [x_1, x_2, .., x_N]^T \in \mathbb{R}^{D \times 1}$ and $y = [y_1, y_2, .., y_N]^T \in \mathbb{R}^N$.

The vector of actual values of function evaluation is given by $\boldsymbol{f}$, and the vector for actual values of prediction points is written as $\boldsymbol{f}_*$. A probabilistic joint Gaussian distribution is assumed for the unknown function values given by,

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} m_f \\ m_* \end{bmatrix}, \begin{bmatrix} k_{ff} & k_{f*} \\ k_{*f} & k_{**} \end{bmatrix} \right) \qquad (2.98)$$

The mean function value $m(\cdot)$ is considered less critical for simplicity and hence assumed to be zero for noise-free applications (Zhang et al., 2022). A general understanding of taking the mean value equal to zero is that the prediction value is assumed to be zero in the regions of sample space where the data is absent (António and Rodrigues, 2021).

The covariance function $K(\cdot, \cdot)$ has two regressors as arguments and returns a scalar value given below,

$$K(\cdot, \cdot): \mathbb{R}^D \times \mathbb{R}^D \longmapsto \mathbb{R}^D \qquad (2.99)$$

For the given data, the model prediction $\boldsymbol{f}_*$ can be derived as conditionally distributed,

$$\boldsymbol{f}_* | X, \boldsymbol{f} \sim \mathcal{N}(\mu_*, \textstyle\sum_*) \qquad (2.100)$$

where μ_* is the regression output for the input x_* predicted by GPR calculated using,

$$\mu_* = k_{*f} k_{ff}^{-1} \boldsymbol{f} \qquad (2.101)$$

The covariance matrix $\sum_*$ quantifies the uncertainty in the predictions as given by,

$$\textstyle\sum_* = k_{**} - k_{*f} k_{ff}^{-1} k_{f*} \qquad (2.102)$$

This probability distribution is called posterior distribution. For the test data corrupted with Gaussian noise, the inference procedure is extended by adding the noise variance term:

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{ff} + \sigma^2{}_n I & k_{f*} \\ k_{*f} & k_{**} \end{bmatrix}\right) \qquad (2.103)$$

The posterior distribution is then given as,

$$\boldsymbol{f}_*|X, \boldsymbol{f} \sim \mathcal{N}(\mu_*, \Sigma_*) \qquad (2.104)$$

$$\mu_* = k_{*f}(k_{ff} + \sigma^2{}_n I)^{-1}\boldsymbol{f} \qquad (2.105)$$

$$\Sigma_* = k_{**} - k_{*f}(k_{ff} + \sigma^2{}_n I)^{-1}k_{f*} \qquad (2.106)$$

This distribution can be used to evaluate function predictions at a test/query point x_*. Figure 2.21 shows the process of inference.

Figure 2.21 Process of inference for a) Prior distribution with dotted values showing generated y values. b) Posterior distribution for the three random functions (Rasmussen and Williams, 2006).

Support Vector Regression (SVR)

Support vector regression is considered one of the most powerful algorithms for supervised machine learning (Aghaaminiha et al., 2021). An alternative loss function, shown in Figure 2.22, is introduced in the support vector machine algorithm to transform it into SVR (Brereton and Lloyd, 2010). This is done by modifying the loss function to include a distance measure (Yondo et al., 2018). Loss functions used in SVR are Quadratic, Laplace, Huber and ε-insensitive (Vapnik, 1999).

Figure 2.22 Loss functions (Brereton and Lloyd, 2010).

SVR in a general form is given below,

$$\hat{y}(x) = \mu + \sum_{i=1}^{n} w^{(i)} + \psi^{(i)}(x, x^{(i)}) \tag{2.107}$$

where μ is the base term, $w^{(i)}$ are weights and $\psi^{(i)}$ are the basis functions.

A linear regression function with the form shown below,

$$\hat{f}(x) = \mu + \langle w, x \rangle = \mu + w^T x \tag{2.108}$$

with $w \in \chi$ and $\mu \in \mathbb{R}$.

Support vectors approximate the predictor while the sample points within the ε-insensitive tube are discarded to build an accurate SVR surrogate (Yondo et al., 2018). This problem is often considered a constrained convex quadratic optimization problem, which requires,

$$\text{Minimise } \frac{1}{2} \| w^2 \| \tag{2.109}$$

$$\text{Such that } \begin{cases} y_i - w.x^{(i)} - \mu \le \varepsilon \\ w.x^{(i)} + \mu - y_i \le \varepsilon \end{cases} \tag{2.110}$$

A linear SVR regression with the introduction of Lagrange multipliers is given as,

$$\hat{y}(x) = \mu + \sum_{i=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)})(x^{(i)}.x) \tag{2.111}$$

For a non-linear relationship between the input variables and the output, the transformation (ϕ) of input variables to higher dimensions is carried out to make them linearly separable (Aghaaminiha et al., 2021). The SVR for non-linear regression can be expressed as,

$$\hat{y}(x) = \mu + \sum_{i=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)})\psi^{(i)} \tag{2.112}$$

where $\psi = \phi.\phi$ a continuous, symmetric and positive definite basis function known as Mercer kernel.

Random Forest

Random forest (RF) is an ensemble machine-learning technique used for regression and classification (Quinlan, 1987). This technique consists of multitudes of decision trees in which smaller subsets (leaf) are created around input variables by splitting (branch) a dataset. As a result, these subsets will have minimal variances in the outcome values. This splitting is carried out until the terminal condition is met, which could be either the number of splits or a cut-off value for the standard deviation for each subset. Following this, once the terminal condition is met, the average value in the leaf is the predicted outcome for that set of inputs (Aghaaminiha et al., 2021). The random forest has two hyperparameters: the number of trees and the maximum number of features that can be split.

2.6.3   Surrogate Model Quality Assessment

The quality of the surrogate model can be assessed using statistical techniques such as the Root Mean Square Error (RMSE), the Relative Maximum Absolute Error (RMAE), the Relative Average Absolute Error (RAAE) and the R^2. This subsection gives an overview of the techniques used for the quality assessment of the surrogate model.

The Root Mean Squared Error

The Root Mean Squared Error (RMSE) is calculated using,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2.113}$$

where n is the number of measurements, $\hat{y}_i$ is the predicted value and y_i is the actual value.

RMSE is often used in regression analysis to measure the average deviation between actual and predicted values. This metric helps know the typical magnitude of errors. RMSE values are sensitive to outliers as the squared differences increase the impact of more significant errors. The lower the RMSE value, the better the model's performance. RMSE value equal to 0 is considered the best.

The relative maximum absolute error

The Relative Maximum Absolute Error is given by,

$$\text{RMAE} = \frac{max(|\hat{y}_1 - y_1|, |\hat{y}_2 - y_2|, |\hat{y}_3 - y_3|, \ldots, |\hat{y}_n - y_n|)}{\sigma} \tag{2.114}$$

where σ is the standard deviation of the observed values.

It gives the relative measure of the maximum absolute error between the actual and predicted values. This metric helps understand the worst-case scenario in terms of the accuracy of the prediction model, and the values range from 0 to 1, where 0 is the ideal value.

The Relative Average Absolute Error

The relative average absolute error is obtained from,

$$\text{RAAE} = \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{n\sigma}$$

(2.115)

RAAE gives a relative measure of the average absolute error between actual and predicted values. This metric helps compare the accuracy of predictions for the datasets with different scales.

The Correlation Coefficient R^2

The R^2 is given by,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

(2.116)

where $\bar{y}$ is the average of observed values.

The value between 0 and 1, where 1 is the ideal value, shows the proportion of variance in the target variable. It represents the goodness of the fit of the model.

In summary, the choice of metric depends on the specific goals and analysis requirements. The RMSE metric is considered the best choice for finding out the overall prediction accuracy of the model. RMAE can be used to find extreme variations in the dataset. RAAE can be helpful if the interest is in finding out the relative error in capturing the scale of the target variable. R^2 metric can be used to evaluate the model's fit.

In this work, RMSE and R^2 values are used to find prediction accuracy and the model's overall fit, respectively.

2.6.4   Optimization Techniques

An optimization technique can be applied to find the global minimum after generating the response surface from the metamodel. In the current research, optimization techniques are used to find an optimal value of corrosion rate based on pH, pCO_2, T and u and the optimal set of electrochemical reaction rate constants to tune the mass transfer model of corrosion rate.

The classical optimization techniques are divided into two types: optimality criteria methods and search methods. A function must satisfy optimality criteria at its minimum point in optimality criteria methods. In search methods, also known as direct methods, the optimum design is estimated initially and then iteratively improved to satisfy the optimality criteria. Some of the line search algorithms include Nelder-Mead, Fibonacci, golden section, Hooke and Jeeves', Powell's, gradient descent, and coordinate descent methods to solve unconstrained optimization problems (Venkateswarlu and Jujjavarapu, 2020).

Optimization techniques are classified into heuristic and non-heuristic optimization techniques. Heuristic optimization techniques, also called metaheuristic optimization techniques, are used to find near-optimal solutions to complex problems when traditional optimization techniques are ineffective (Gandomi et al., 2013). These techniques provide approximate solutions to problems with high dimensionality, nonlinearity and many

constraints. These techniques involve Genetic Algorithms (GA) (Goldberg and Richardson, 1987), Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995), Ant Colony Optimization (ACO) (Dorigo et al., 2006) and Simulated Annealing (SA)(Bertsimas and Tsitsiklis, 1993). Non-heuristic or deterministic optimization techniques find the optimal solution by exploring the entire design space. These techniques include Linear Programming (LP) (Bertsimas and Tsitsiklis, 1997), Integer Programming (IP) (Li and Sun, 2006), Dynamic Programming (DP) (Sakoe and Chiba, 1978) and Branch and Bound technique (Brusco and Stahl, 2005).

Optimisation techniques can also be classified into gradient-based solvers and non-gradient-based solvers. Gradient-based solvers find the minimum or maximum of the objective function by evaluating its gradient (derivative). In this case, the gradient is a vector of partial derivatives concerning each design variable. The optimal solution is approached iteratively by following a direction of negative gradient. On the other hand, non-gradient-based solvers require function evaluation only. Some examples of non-gradient-based solvers include the Nelder-Mead simplex method, the Fibonacci method, simulated annealing and the genetic algorithm.

Efficient design in terms of optimal time, reduced computational burden and human resources is of paramount importance in many industries. Although computational resources are available, it is absurd to be completely dependent on expensive simulations, also known as high-fidelity simulations, to find the optimal design. Surrogate models, with the help of the DOEs, replace these high-fidelity simulations to predict the objective function at the unknown sample points. This process reduces the time required for analysis using high-fidelity simulations, allowing parametric optimization and sensitivity analysis (Yondo et al., 2018).

Nelder-Mead method

The Nelder-Mead method is one of the popular methods applied to unconstrained nonlinear optimization problems (Nelder and Mead, 1965). It uses heuristic ideas to search n^{th} dimensional space. The advantages of this method are that it does not require derivatives of the objective function, and the objective function does not need to be smooth. It uses a geometrical shape called Simplex, a polygon with n+1 vertices (Bagherian et al., 2021). Where n is the number of design variables. For 2 design variables (n=2), the simplex will have 3 vertices, shown in Figure 2.23.

Figure 2.23 A Simplex with 2 design variables.

The process starts with the initialization of n-dimensional Simplex (guess vertex) with n+1 vertices. The worst vertex in i^{th} iteration is replaced using the steps such as reflection(x_r), expansion(x_e), contraction (inside(x_c), outside(x_{cc})), reduction, and convergence. The expressions for these points are given below,

$$x_r = \bar{x} + \hat{\rho}(\bar{x} - x_{n+1}) \tag{2.117}$$

$$x_e = \bar{x} + \chi(x_r - \bar{x}) \tag{2.118}$$

$$x_c = \bar{x} + \beta(x_r - \bar{x}) \tag{2.119}$$

$$x_{cc} = \bar{x} - \beta(\bar{x} - x_{n+1}) \tag{2.120}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.121}$$

where, $\hat{\rho} = 1$ is the coefficient of reflection, $\chi = 2$ is the coefficient of expansion, $\beta = 0.5$ is the coefficient of contraction, and $\bar{x}$ is the centroid.

This method is shown in the flow chart in Figure 2.24.

Figure 2.24 Nelder-Mead simplex algorithm for unconstrained optimization (Bagherian et al., 2021).

Hooke-Jeeves Method

The Hooke-Jeeves method (Hooke and Jeeves, 1961) is a pattern search method that evaluates the objective function at specific points in the design space. This method involves an exploratory search step and a pattern move. An exploratory search step finds a direction from the current point in which the value of the objective function is improved. Pattern move accelerates the search in the best direction. The Hooke-Jeeves algorithm requires five parameters, a starting point vector x^0, a perturbation step size vector P_0, the perturbation

tolerance limit vector T_l , the step size reduction parameter η, and the acceleration factor a. The following steps are carried out in the Hooke-Jeeves method,

At the starting point, the vector x^0, the value of objective function $f(x^0)$ is calculated. Consider, $x^0 = x^{best}$ and $f(x^0) = f(x^{best})$. A starting point vector x^0 is then combined with a perturbation step size vector P_0 to calculate the value of objective function f(x^1) and $x^1 = x^0 + P_0$. If $f(x^1) < f(x^{best})$ then assume x^1 as x^{best} and $f(x^1)$ as $f(x^{best})$. If $f(x^1) > f(x^{best})$ then change the direction by taking $x^1 = x^0 - P_0$ following the evaluation of the objective function and its comparison with the $f(x^{best})$. The pattern move is then carried out using x^0 and x^{best} to create a new point x^2 in improving direction.

$$x^2 = x^0 + a(x^{best} - x^0) \tag{2.122}$$

A typical value of the acceleration factor is equal to 2. Following this, an exploratory step is a repeater for a new point x^2 until the values in the perturbation size vector fall below the perturbation tolerance limit vector T_l.

Powell's Method

Powell's method is a type of iterating method used for unconstrained optimization problems (Powell, 1964). This method does not require derivatives of the objective function, which is a part of quasi-Newton methods. The method finds minima of objective function f(x_i), where is x_i is the vector with n variables at iteration i. The steps of Powell's method are given below,

The method starts with setting an Initial vector P_0 = initial $x_{i=1}$ with an initial base vector U_k for k = 1,2,3…n. Each U_k the vector consists of a kth term that can be set to one and the remaining terms to zero. The next step is to find a value of a_k that minimises function $f(P_{k-1} + a_k U_k)$ for k = 1 to n. Following this, the searching vector at the kth iteration P_k is set as,

$$P_k = P_{k-1} + a_k U_k \tag{2.123}$$

The one-dimensional minimisation for function f is obtained from the above step. Ju and Hsieh (2022) suggested that Brent's method with the golden section search can efficiently obtain one-dimensional optimisation. The value of i is incremented by 1 and set $U_j = U_{j+1}$ for j = 1 to n-1, set $U_n = P_n - P_0$. The value of a is then found to minimise the one-dimensional function of $f(P_0 + aU_n)$ (Ju and Hseih, 2022). The variable vector is then updated to $x_i = P_0 + aU_n$. Following this set $P_0 = x_i$ and repeat the above steps.

2.6.5 Dimensionality Reduction Technique – Principal Component Analysis (PCA)

PCA is a technique used to reduce the dimensionality of a large dataset. This method transforms a large set of variables or features into a smaller one without losing the information in the large set.

The method begins with standardising the continuous initial variables to ensure they contribute equally to the analysis. This step aids in reducing the dominance of initial variables with significant differences over the ones with slight differences in value. This is followed by computing the covariance matrix (ndv × ndv) to understand how the input data set variables vary from the mean to each other (Jollife and Cadima, 2016). The covariance matrix has the covariance entries associated with the pairs of the initial variables.

If the covariance sign is positive, then the two variables increase or decrease together, which means they are correlated. If covariance is negative, one variable increases when the other variable decreases (Jollife and Cadima, 2016). The next step involves computing the eigenvalues and eigenvectors from the covariance matrix to determine the principal components of the data. The principal components are new variables constructed as linear combinations, so principal components (new variables) are not correlated, squeezing most of the information into the initial components (Tapeh & Naser, 2023).

Considering the advantages of PCA, it has been used in this work to check if the dimensionality of the dataset can be reduced for both the studies of finding optimal operating parameters and electrochemical reaction rate constants. Previously, in several studies, PCA has been used to find dimensionality reduction-based optimisation.

2.7   Applications of Machine Learning Modelling in Corrosion

Several studies in the literature used machine learning modelling in the area of corrosion science (Xu et al., 2023). Machine learning models with the enhanced abilities of high-speed computing have been used to make necessary real-time corrosion decision support in oil and gas industries, as shown in the framework in Figure 2.25 (Xu et al., 2023). The data collected from the oil and gas gathering and transportation system is inspected for outlier recognition, followed by correlation analysis in which redundant input variables were removed. The identified input and output variables are then used to predict the corrosion rate from the constructed machine-learning model. Following this, the machine learning model's performance is evaluated based on the metrics. The best machine learning model is then used for the decision supporting corrosion mitigation (Xu et al., 2023).

Figure 2.25 Framework for implementing artificial intelligence (AI) in corrosion decisions (Xu et al., 2023).

ANN (Bassam et al., 2009; Weckman et al., 2010; De Masi et al., 2015; Chou et al., 2017; Pai et al., 2020), RF (Ossai, 2019; Aghaaminiha et al., 2021; Ben Seghier et al., 2022), SVM (Lee et

al., 2013; Ji et al., 2015; Hatami et al., 2016; Chou et al., 2017), GP (Zhang et al., 2022) and Deep Learning (DL) (Ossai, 2019) are some of the studies that focused on the application of machine learning in the corrosion prediction. Out of these models, ANN models, often referred to as Shallow Neural Networks (SNN), are the most popular machine learning of CO_2 corrosion.

Bassam et al. (2009) developed an ANN model to determine the type of Electrochemical Impedance Spectroscopy (EIS) from inhibitor concentrations of three datasets. ANN models predicted the type of corrosion using hyperbolic sigmoidal transfer functions; however, a more efficient ANN could have been developed. Weckman et al. (2010) developed the ANN model for a better understanding of CO_2 corrosion inhibition performance by extracting the knowledge from the ANN model. The knowledge extraction techniques included Network Interpretation Diagrams (NID), Garson's algorithm, a family of curves, sensitivity analysis, and TREPAN-plus (Weckman et al., 2010). However, Weckman et al.'s model was not tuned for the hyperparameters, and the predicted data's cross-validation was missing. Castellanos et al. (2011) developed a Failure Analysis Expert System (FAES) using a multilayer perceptron structure and Back Propagation Neural Network (BPNN) approach to understanding failure mechanisms such as external corrosion, internal corrosion, erosion, material selection and vandalism. Khalajestani and Bahaari (2014) developed a four-layer multilayer perceptron model to predict interacting local thinning areas caused by erosion or corrosion in elbows. The model's accuracy was verified using mean squared error (MSE) metrics. ANN is also used to predict experimental values such as corrosion current density and corrosion potential with good agreement between the actual and predicted values (Narimani et al., 2015). Abbas et al. (2018) established a neural network model as a tool for corrosion prediction with the main focus on the characterisation of transfer and training functions available in MATLAB software to check the suitability for the prediction of CO_2 corrosion rate. The dataset used for the neural network modelling was obtained from weight loss experiments using autoclaves (Hesjevik et al., 2003; Choi and Nešic, 2009; Zhang et al., 2013) with unevenly spread 22 sample points used for modelling.

Xu et al. (2023) stated that the low accuracy of the ANN models available in the literature was due to problems such as over-fitting and local minimum in the training process. They stated that this issue could be eliminated by adjusting the network parameters with the help of population-based optimisation algorithms such as the Beetle Antennae Search (BAS) Algorithm, PSO, Whale Optimisation Algorithm (WOA), Sparrow Search Algorithm (SSA), and Cuckoo Search Algorithm (CSA). The accuracy of the BPNN was improved by optimising the initial weights and deviations of BPNN using GA (Wen et al., 2019). Ossai (2020) tuned the hyper-parameters using PSO of a feed-forward sub-space clustering neural network to predict leakage failures and corrosion defects of pipelines long in service. Some innovative algorithms for the prediction of corrosion density include a combination of ANN with levy flight weighted quantum PSO algorithm, which provided twice as much an improvement in the accuracy of mean averaged error (MAE) compared to BPNN (Wang et al., 2022). Other examples include the use of the WOA (Ouladbrahim et al., 2022) and SSA (Xin et al., 2022) to optimise hidden layer neurons, deviation and the weights of the BPNN model.

After ANNs, SVMs were found to be the most popular technique, accounting for 27% of the available machine learning techniques, with RF at 9% and DNN at 7% for the prediction of CO_2 corrosion (Xu et al., 2023). SVM was useful for predicting the stress concentration coefficient of pipeline corrosion pits (Ji et al., 2015), monitoring pipeline conditions (Lee et al., 2013), and

identifying leakage apertures and leakage locations (Sun et al., 2016). A least-square SVM (LSSVM) model was developed by Hatami et al. (2016) to predict the corrosion rate based on input parameters such as pH, pCO_2, temperature and flow velocity obtained from the experimental dataset of Dugstad et al. (1994). Although an attempt was made to predict the CO_2 corrosion rate using LSSVM, it did not provide descriptive statistics for the data used, which could have helped to identify outliers and missing values in the dataset. Peng et al. (2021) developed a model that combined support vector regression (SVR), PCA and chaos PSO (CPSO) for the internal corrosion prediction of the multiphase pipeline. It was found that a combined model of PCA-CPSO-SVR provided a lower RMSE value (2.7%) compared to remaining models viz, SVR (5.9%), PCA-SVR (8.1%), PCA-GA-SVR (3.1%) and PCA-PSO-SVR (2.9%). SVR, ANN, RF and K-nearest neighbours are used to predict the time-dependent corrosion rates in the presence of corrosion inhibitors (Aghaaminiha et al., 2021). The SVR model provided a 13% RMSE value, which was 2nd best after the RF model.

RFs have been used for classification purposes in corrosion modelling for leakage detection and classification (Ning et al., 2021) and for locating and estimating pipeline damage (Bao et al., 2022); however, the model developed had lower accuracy than numerical and experimental studies. This was overcome by combining the RF with the random decrement technique (RDT) used in data processing to eliminate the influence of noise (Bao et al., 2022). Zahedi et al. (2018) used RF to predict erosion ratio in single-phase and multiphase flow situations and found that RF was the best-performing machine learning model. However, it over-predicted low erosion ratio values. Aghaaminiha et al. (2021) found that the RF model outperformed other machine learning models when trained on a combination of numerical and categorical datasets and provided MSE values ranging from 0.005 to 0.093. RF model was then used to study the sensitivity of corrosion rates to changes in the input variables such as type of corrosion inhibitor and concentration, pCO_2, temperature, wall shear stress, brine type.

Some other machine learning models used for corrosion modelling include Gaussian Process Regression (GPR) and Deep Neural Network (DNN). For example, Zhang et al. (2022) constructed a predictive model using GPR that included an adaptive sampling strategy for the CFD-based corrosion modelling of flow in an elbow. The advantage of using the adaptive sampling technique is that it controls the uncertainty in the model prediction Zhang et al. (2022). In addition, there has been an increase in the use of DNN classification problems to predict leakage from gas pipelines (Kopbayev et al., 2022), to predict the damage location (Bao et al., 2022) and corrosion detection for oil and gas pipelines (Forkan et al., 2022).

Research Gap

This section has shown that there has been significant interest in applying machine learning modelling to predict corrosion in oil and gas pipelines in the last two decades. However, some machine learning models still experience issues such as over-fitting and local minimum in the training process and difficulty adjusting hyper-parameters, resulting in lower accuracy and high calculation costs (Xu et al., 2023). In addition, very little attention is given to the sample points of the variables used in machine-learning models to predict corrosion rates. The most popular machine learning models, ANN, SVR, GPR, and RF, have been applied to predict corrosion rates, and their performance is assessed on the basis of evaluation metrics.

2.8 Surrogate-based Optimisation to find Optimal Parameters for Mathematical Modelling

Surrogate models replace high-fidelity CFD simulations to reduce computational time and power without compromising accuracy (Liu et al., 2023). Surrogate modelling-based optimisation (SBO) is widely used in various areas but is more prevalent in aerodynamic analyses due to the massive time required by full-scale models. One of the earliest studies that addressed the SBO method included the design of sampling space using orthogonal-array Latin hypercube sampling, Gaussian process regression (Kriging) to build a surrogate model and assessing the performance of optimisation methods on helicopter rotor blade design problem of 31 design variables (Booker et al., 1999). Queipo et al. (2005) carried out multi-objective optimisation of liquid-rocket injectors to find an optimal design. The two primary objectives in their studies were improving the performance and life of injector design, which is associated with an axial length of the thrust chamber and thermal field inside the thrust chamber, respectively. It included using orthogonal arrays to generate sample space given as OA (54,4,3,2) corresponding to 54 designs/runs, four factors, and three levels with strength 2 (Queipo et al., 2005). The parametric polynomial regression surrogate model was used for sensitivity analysis and multi-objective optimisation. They demonstrated that the SBO approach could effectively solve multi-objective optimisation problems but did not explore other surrogate modelling techniques. Thapa and Missoum (2022) constructed a SBO framework for optimising horizontal-axis wind turbine composite blades using Kriging and SVM approach with adaptive sampling viz., max-min sampling and generalized max-min for the refinement. Adaptive sampling can improve the accuracy of the surrogate model and has been used along with GPR to predict the corrosion rate in elbows (Zhang et al., 2022). The SBO framework has replaced CFD simulations to optimise the rotor airfoil's aerodynamic shape (Li et al., 2020; Du et al., 2021; Liu et al., 2023). Liu et al. (2023) used a deep neural network (DNN) surrogate model to optimise the aerodynamic shape of the SC1095 rotor airfoil with the primary aim of suppressing dynamic stall. The objective function was to minimise the drag and moment coefficients of the airfoil using a multi-island genetic algorithm (MIGA). This approach was significantly faster than CFD, with high prediction accuracy.

Surrogate modelling-based parameter optimisation has been used in several studies to find the best set of parameters, for example, in the proton exchange membrane fuel cell (PEMFC) model (Li et al., 2021; Fan et al., 2022), pressure swing adsorption (PSA) model (Subraveti et al., 2019) employed for adsorption and cyclically removal of contaminants, and in granulation modelling (Braumann et al., 2010). PEMFC cells, due to their low emissions and high energy efficiency, are considered one of the most prominent energy conversion devices (Li et al., 2021). Hence, maximizing their performance based on operating conditions and structural parameters has been a significant area of interest for many researchers (Mohamed and Jenkins, 2004; Peng et al., 2017). The SVM approach (Peng et al., 2017) and ensemble learning model with the base learner as ANN model (Li et al., 2021) were used to identify the significant variables of the PEMFC model. Peng et al. (2017) solved a single objective function of maximization of power density values using the standard simplex approach, whereas Li et al. (2021) solved a multi-objective optimisation problem which included power density, system efficiency and uniformity of O_2 distribution on the cathode catalyst layer using a non-dominated sorting genetic algorithm (NSGA-II). These studies showed that surrogate models efficiently identify significant parameters and their optimal values. Subraveti et al. (2019)

proposed Surrogate-assisted Optimisation (SOpt) and Dimensional Reduction-based Optimisation (DROpt) techniques to find the set of optimal parameters of the mathematical model of PSA. The SOpt approach consisted of ANN model to construct a surrogate and NSGA-II model for the multi-objective optimisation. They found that the SOpt approach was the most efficient when compared with predictions of the mathematical model of PSA, again highlighting that surrogate-based optimisation is a reliable methodology to replace physics-based expensive simulations. Braumann et al. (2010) carried out surrogate modelling-based parameter optimisation for the complex full-scale granulation model. This model consisted of unknown parameters often identified by solving inverse problems in which predictions obtained by full-scale models were compared with experimental results (Braumann et al., 2010; Myers et al., 2022). Their approach included using a quasi-random screening for sampling and a first and second-order polynomial to fit the sample points. They found that the second-order response surfaces showed more minor relative uncertainties for estimating four unknown rate constants. The above studies showed that surrogate modelling can replace complex, full-scale physics-based models. However, these studies did not explore other surrogate models as the applicability of a surrogate model depends on several factors such as input data quality, dimensionality of the input variables, relationship complexity between input and output variables, constraints and model assumption (Forrester, 2008).

Research Gap

The predictive tools in the literature for the prediction of CO_2 corrosion consists of empirical values for reaction rate kinetics. These values have been tuned in several studies to get the desired predictions of corrosion rate (Nordsveen et al., 2003; Al-Khateeb et al., 2018; Hu et al., 2018; Nešić et al., 2019). The lack of reliable predictive tools makes it difficult to make the decision required to mitigate corrosion in the oil and gas industries. Conducting a sensitivity analysis of these empirical values to obtain a reliable set of values requires running the mass transfer model several times, which is not feasible as it takes a tremendous amount of time and power. To circumvent this, the SBO approach can be implemented as mentioned above. A machine learning based surrogate modelling to find a reliable set of electrochemical reaction rate constants for the prediction of CO_2 corrosion is implemented for the first time. It generates a surrogate model using ANN, GPR, SVR, and RF, then solves the optimisation problem that is subject to constraints. Finally, a comparison is made with the experimental values, CFD-driven mass transfer prediction and values obtained using the SBO approach for corrosion rates.

2.9 Summary and Conclusions

This chapter provided theory, background information, and a literature review related to CO_2 corrosion, CFD, surrogate modelling and machine learning techniques. It covered the primary mechanism behind CO_2 corrosion and current prediction models are available in the literature. It highlighted the assumptions on which these models are based, stating that a reliable prediction model is needed. One of the significant issues related to these models is the use of empirical correlations of viscous sublayer thickness and turbulent diffusivities, which restricts their ability to predict corrosion rates in situations where the flow is not fully developed. CFD is a reliable tool that can be used to resolve this problem. This technique will involve accurate predictions of viscous sublayer thickness and turbulent diffusivity profiles and then subsequently coupling these predictions with the mass transfer model for the prediction of CO_2 corrosion.

A review of machine-learning-enabled surrogate modelling provided the background theory related to the different surrogate models, which can be applied to find the optimum parameters in the mass transfer models. Electrochemical reaction rate constants used in the current mass transfer models are often tuned to predict better results than the experimental data. This study focuses on identifying the most significant electrochemical reaction rate constants and then finding the optimum values of these constants. Machine learning techniques can be applied to find out the impact of significant parameters such as pH, velocity, temperature, and pCO_2. This methodology will assess all the supervised machine-learning models and will provide the most suitable model for the prediction of CO_2 corrosion can then be used to find the optimum combination of the parameters.

As mentioned above, the current CO_2 corrosion prediction models are not suitable for flow situations that are not fully developed. This study resolves this issue by accurately coupling the flow field obtained from CFD with the mass transfer model in an expansion/constriction pipe.

Chapter 3: Computational Fluid Dynamics Driven Mass Transfer Model for the Prediction of CO₂ Corrosion in Pipelines

3.1 Introduction

The flow in a pipeline is characterised as laminar, transitional, or turbulent based on the Reynolds number (Re). The flow field in most of the mechanistic models is assumed to be turbulent (Nesic et al., 1995; Nesic et al., 2001; Nordsveen et al., 2003; Nešić et al., 2009; Zheng, 2015; Nešić et al., 2019; Kahyarian and Nesic, 2020) due to high flow rate and length of the pipes used for transportation resulting in a fully developed turbulent flow (Nešić, 2007). This flow field is divided into three regions: the viscous sublayer, the buffer layer, and the turbulent layer, as shown in Figure 3.1.

Figure 3.1 Turbulent boundary layer profile over a flat plate (not to scale).

Very close to the wall, mass transfer is dominated by molecular diffusion. However, studies related to the observation of fluid elements adjacent to the wall have shown that vertical components of velocity fluctuations continue to exist until they reach the wall and thus occur within the viscous sublayer (VSL) of a turbulent flow (Popovich and Hummel, 1967). The effect of turbulent convection on an element carried within the fluid can be quantified using the notion of D_t. The dispersion of particles in each region is caused by eddies within the fluid flow. These eddies cascade in scale to a size smaller than the region's; larger eddies advect the fluid region but do not cause element separation. As the region grows in size due to advection, the range of eddies captured facilitates the dispersion of particles in the flow. This is turbulent diffusion, different from eddy diffusion, which is the process by which eddies are dispersed in a turbulent flow. The available equations for measuring turbulent behaviour are based on steady-state approximations since capturing the time-dependent behaviour of turbulence without an experiment for a particular flow is highly challenging. Hence, the time-dependent fluid flow and mass transport changes are described using eddy and turbulent diffusivity. As a result, researchers have no consensus on predicting turbulent diffusivity in horizontal pipelines. All empirical equations for the turbulent diffusivity are either assumed in a general form (Notter and Sleicher, 1971) or obtained by experimental data fitting (Notter and Sleicher, 1971; Davies, 1972; Nordsveen et al., 2003; Wang and Nesic, 2003). These

empirical correlations only apply to a fully developed pipe flow, limiting their use in non-idealised geometries such as elbows, rectangular channels, or sudden expansion/constriction. This has become one of the main drawbacks of using empirical correlations in the mechanistic modelling of CO_2 the present study aims to alleviate corrosion using CFD to determine the turbulent boundary layer profile and flow behaviour in the near wall region. The following set of objectives are studied in this chapter:

- Explore the limitations of the empirical correlations used in the mechanistic models available in the literature.
- Build the mass transfer model with the empirical correlations of viscous sublayer thickness and turbulent diffusivity.
- Modify the mass transfer model's chemistry to accommodate carbonic acid's dissociation.
- Build a CFD flow model in a straight pipeline using different turbulence models and choose the best model that resolves the near-wall fluid flow.
- Run CFD simulations for different velocities to obtain viscous sublayer thickness and turbulent diffusivity predictions.
- Use the predictions obtained from CFD simulations to drive the mass transfer model of CO_2 corrosion.
- Compare the CO_2 corrosion rate predictions with experimental data in the literature and values obtained from the empirical correlations-based mass transfer model.

This study's CFD-driven corrosion prediction model provides a robust approach for coupling the fluid flow and mass transfer models to predict corrosion rates. A vital feature of this approach is the accurate calculation of viscous sublayer thickness and turbulent diffusivity profiles, which sets a benchmark for the corrosion rate predictions in disturbed, developing and fully developed flow conditions. The mass transfer model is based on the Nordsveen et al. (2003) multi-node model, which integrates through the viscous sublayer to calculate the concentration of each species at a series of points by accounting for the transport of species to and from the bulk. It includes homogeneous chemical reactions, diffusion of species, electrochemical reactions at the steel surface and transport of species to and from the bulk, including diffusion and convection through the viscous sublayer (Nordsveen et al., 2003). Section 3.2 describes the mass transfer, velocity and eddy distribution in pipeline flows and CFD methodology. Section 3.3 explains the numerical approach used in the mass transfer model of axisymmetric pipe. Section 3.4 describes the CFD methodology. Section 3.5 presents a comprehensive series of results containing calculations of viscous sublayer thickness, turbulent diffusivities, corrosion rates, and verification using the experimental data in the relevant literature.

3.2 Materials and Methods

3.2.1 Mass Transfer Modelling

The transport of species is described using species conservation equations. The governing equation for the transport of diluted species j in a fluid medium is given by Eq. (3.1):

$$\frac{\partial c_j}{\partial t} = -\frac{\partial \boldsymbol{N_j}}{\partial x} + R_j \qquad (3.1)$$

where c_j the concentration of species j, N_j is the flux of species j due to diffusion, R_j is the rate of production of species j due to homogeneous chemical reactions, t is the time, and x is the spatial coordinate (distance from the metal surface). Figure 3.2 shows the transport of species in the mass transfer modelling.

Figure 3.2. Transport of species in mass transfer modelling in horizontal pipelines.

The flux of species N_j consists of diffusion, electro-migration, and convection. The model developed here is applied to cases where the electrolytes have high conductivity, including the major species Na^+ and Cl^-, with the result that electro-migration in the electric field is negligible compared to diffusion and convection. For a comprehensive justification of this assumption, see references (Stephens and Mauzeroll, 2019; Balsara and Newman, 2021). In addition, since the aim is to assess the importance of modelling the VSL accurately on corrosion modelling, the model is developed for cases where the precipitation of $FeCO_3$ is negligible, obviating the need to introduce additional empiricism into the corrosion modelling, as in previous models for corrosion product formation, e.g., Nesic et al. (2001). In fact, in the cases considered here, the temperatures are low (around 20°C), which means that the super-saturation S given by Eq. (3.2),

$$S = \frac{c_{Fe^{2+}} c_{CO_3^{2-}}}{K_{sp}}$$

(3.2)

is small, and the precipitation of $FeCO_3$ will, therefore, be negligible. This issue is discussed in greater detail by Nesic et al. (2001), who concluded that the precipitation rate of $FeCO_3$ is extremely low for temperatures less than 60°C. The ability to neglect corrosion product formation enables a steady-state modelling approach to be adopted.

The flux of species N_j is given by Eq. (3.3),

$$N_j = -(D_j + D_t)\frac{\partial c_j}{\partial x}$$

(3.3)

The turbulent diffusivity (D_t) is shown by Eq. (3.4),

$$D_t = \frac{\mu_t}{\rho Sc_t} \tag{3.4}$$

The sensitivity of the term Sc_t to corrosion rate predictions by varying it from 0.5 to 0.9, and 0.71 is chosen for the mass transfer calculations upon comparison with the experimental corrosion rates. The D_t of each species, j varies within the domain as μ_t varies through the viscous sublayer of the flow (Davies, 1972). Table 3.1 and Table 3.2 show the species properties as a function of temperature and reference values of molecular diffusion coefficients for chemical species involved in the corrosion process.

Table 3.1 Species properties as a function of temperature T(K) used in the mass transfer model.

Property	Equation
Density (Kg/m^3)	$\rho(T) = 1152.3 - 0.5116T$
Dynamic viscosity (Pa*s)	$\mu(T) = \mu_{ref} \times 10^{\frac{1.3272(20-T)-0.001053(20-T)^2}{T+105}}$
Diffusion coefficient (m^2/s)	$D = D_{ref}\left(\frac{T}{T_{ref}}\right)\left(\frac{\mu_{ref}}{\mu}\right)$

where T_{ref} is the reference temperature =20°C, μ_{ref}=1.002 *kg/(m.s)*.

Table 3.2 Reference values of molecular diffusion coefficients of chemical species involved in the corrosion process.

Species	Diffusion Coefficients (m^2/s)	Reference
CO_2	1.96×10^{-9}	(Perry and Green, 1987)
H_2CO_3	2×10^{-9}	(Kvarekval, 1997)
HCO_3^-	1.105×10^{-9}	(Newman, 1991)
CO_3^{2-}	0.92×10^{-9}	(Kvarekval, 1997)
H^+	9.312×10^{-9}	(Newman, 1991)
OH^-	5.26×10^{-9}	(Newman, 1991)
Fe^{2+}	0.72×10^{-9}	(Kvarekval, 1997)

3.2.2 Water Chemistry Modelling

CO_2 a stable and inert gas, when dissolved in water, gives a reactive chemical species called carbonic acid (H_2CO_3). The carbonic acid then dissociates to give bicarbonate ions (HCO_3^-). The bicarbonate ions are further dissociated, giving carbonate (CO_3^{2-}) and Hydrogen ions. Table 3.3 shows chemical reactions related to this equilibrium (Nesic et al., 2001; Nordsveen et al., 2003).

The reaction rates depend on the temperature, partial pressure of CO_2 (p_{CO_2}) and ionic strength (I) (Nordsveen et al., 2003). The ionic strength (I) of a solution provides the amount

of concentration present in the solution. Table 3.4 provides the equilibrium constants used in the corrosion prediction model. The forward reaction rate constants are denoted as K_f and backward reaction rate constants are represented as K_b. I is the ionic strength (mol/L), T_f is the temperature in degrees Fahrenheit, T_c is the temperature in degrees Celsius, T is the temperature in Kelvin, and p is pressure in psi.

Table 3.3 Chemistry of CO₂ dissociation in water

	Reaction	Equilibrium Constant
Carbon dioxide dissolution	$CO_{2(g)} \rightleftharpoons CO_{2(aq)}$	$K_{sol} = \dfrac{CO_{2(aq)}}{pCO_{2(g)}}$
Carbon dioxide hydration	$CO_{2(aq)} + H_2O_{(l)} \rightleftharpoons H_2CO_{3(aq)}$	$K_{hyd} = \dfrac{[H_2CO_3]}{\left[CO_{2(aq)}\right]}$
Dissolution of carbonic acid	$H_2CO_{3(aq)} \rightleftharpoons HCO_3^-{}_{aq} + H^+{}_{(aq)}$	$K_{ca} = \dfrac{[HCO_3^-][H^+]}{[H_2CO_3]}$
Dissolution of bicarbonate ion	$HCO_3^-{}_{aq} \rightleftharpoons CO_3^{2-}{}_{(aq)} + H^+{}_{(aq)}$	$K_{bi} = \dfrac{[CO_3^{2-}][H^+]}{[HCO_3^-]}$
Dissociation of water	$H_2O_{(l)} \rightleftharpoons OH^-{}_{(aq)} + H^+{}_{(aq)}$	$K_w = [OH^-][H^+]$

Table 3.4. Equilibrium constants used in the corrosion prediction model.

Equilibrium constants	Reference
$K_{sol} = \dfrac{14.5}{1.00258} \times 10^{-(2.27+0.00565\,\cdot T_f - 8.06\times10^{-6}\,\cdot\,T_f^2 + 0.075\,\cdot\, I)}\,\dfrac{\text{molar}}{\text{bar}}$	(Oddo and Tomson, 1982)
$K_{hy} = 0.00258$	(Palmer and Van Eldik, 1983)
$K_{f,hy} = 10^{329.85-110.541\times \log_{10}(T_k)-(\frac{17265.4}{T_k})}\,\text{s}^{-1}$	(Palmer and Van Eldik, 1983)
$K_{ca} = 387.6 \times 10^{-(6.41-1.594\times10^{-3}\,\cdot\,T_f + 8.52\times10^{-6}\,\cdot\,T_f^2 - 3.07\times10^{-5}\cdot p - 0.4772\cdot I^{0.5}+0.118\cdot I)}\,\text{molar}$	(Oddo and Tomson, 1982)
$K_{f,ca} = 10^{5.71+0.0526\cdot T_c - 2.94\times10^{-4}\cdot T_c^2 + 7.91\times10^{-7}\cdot T_c^3}\,\text{s}^{-1}$	(Bamford & Tiffer, 1972)
$K_{bi} = 10^{-(10.61-4.97\times10^{-3}\cdot T_f + 1.331\times10^{-5}\cdot T_f^2 - 2.624\times10^{-5}\cdot p - 1.166\cdot I^{0.5}+0.3466\cdot I)}\,\text{molar}$	(Oddo and Tomson, 1982)
$K_{f,bi} = 10^9\text{s}^{-1}$	(Nordsveen et al., 2003)
$K_{wa} = 10^{-(29.3868-0.0737549\cdot T_k + 7.47881\text{x}10^{-5}\cdot T_k{}^2)}(\text{molar})^2$	(Kharaka et al., 1988)
$K_{b,wa} = 7.85 * 10^{10}\ (\text{molar})^{-1}(\text{s})^{-1}$	(Delahay, 1952)

Table 3.5. Equations for the calculation of reaction rates.

Reaction	Reaction rate equation
Dissolution of carbon dioxide	$R_{CO_2,s} = \dfrac{\partial}{\partial t}\left(c_{CO_2,s}\right) = K_{b,hy}\, c_{H_2CO_3,s} - K_{f,hy} c_{CO_2,s}$
Formation of Carbonic acid (H_2CO_3)	$R_{H_2CO_3,s} = \dfrac{\partial}{\partial t}\left(c_{H_2CO_3,s}\right) = -\left(K_{b,hy}\, c_{H_2CO_3,s} - K_{f,hy} c_{CO_2,s}\right) - \left(K_{f,ca}\, c_{H_2CO_3,s} - K_{b,ca} c_{H_s^+} c_{HCO_3^-,s}\right)$
Formation of Bicarbonate ions(HCO_3^-)	$R_{HCO_3^-,s} = \dfrac{\partial}{\partial t}\left(c_{HCO_3^-,s}\right) = \left(K_{f,ca}\, c_{H_2CO_3,s} - K_{b,ca} c_{H_s^+} c_{HCO_3^-,s}\right) - \left(K_{f,bi}\, c_{HCO_3^-,s} - K_{b,bi} c_{H_s^+} c_{CO_3^{2-},s}\right)$
Formation of Carbonate ions (CO_3^{2-})	$R_{CO_3^{2-},s} = \dfrac{\partial}{\partial t}\left(c_{CO_3^{2-}}\right) = \left(K_{f,bi}\, c_{HCO_3^-,s} - K_{b,bi} c_{H_s^+} c_{CO_3^{2-},s}\right)$
Formation of Hydroxide ions (OH^-)	$R_{OH_s^-} = \dfrac{\partial}{\partial t}\left(c_{OH_s^-}\right) = K_{f,wa} - K_{b,wa} c_{H_s^+} c_{OH_s^-}$
Formation of H^+ ions	$R_{H^+.s} = \dfrac{\partial}{\partial t}\left(c_{H_s^+}\right) = \left(K_{f,ca}\, c_{H_2CO_3,s} - K_{b,ca} c_{H_s^+} c_{HCO_3^-,s}\right) + \left(K_{f,bi}\, c_{HCO_3^-,s} - K_{b,bi} c_{H_s^+} c_{CO_3^{2-},s}\right) + \left(K_{f,wa} - K_{b,wa} c_{H_s^+} c_{OH_s^-}\right)$

The equations presented in Table 3.5 are solved for the prediction of concentrations of species c_j in Python, j is used as an index for the species involved in the corrosion process. The representation of the concentration of species c_j is given below,

$$c = [c_{CO_2,s}, c_{H_2CO_3,s}, c_{HCO_3^-,s}, c_{CO_3^{2-}}, c_{OH_s^-}, c_{H_s^+}]$$

$$c_j = c(j)$$

$$j = 1,2, \dots N$$

where N is the total number of species involved in the corrosion process.

3.2.3 Electrochemical Reactions

Table 3.6 shows all the electrochemical reactions that take place in the current mass transfer model of CO_2 corrosion prediction, which includes the reduction of H^+ and H_2CO_3. The model is structured so that the additional cathodic reactions can be switched on and off.

Table 3.6. Electrochemical reactions take place on the metal surface.

Electrochemical Reaction	Reaction Type
$2H^+_{(aq)} + 2e^- \rightleftharpoons 2H_{2(g)}$	Cathodic
$2H_2CO_{3(aq)} + 2e^- \rightleftharpoons 2HCO_{3\,aq}^- + 2H_{2(g)}$	Cathodic
$Fe^{2+}_{(aq)} + 2e^- \rightleftharpoons Fe_s$	Anodic

The reaction rate of these electrochemical reactions is expressed in terms of an electrical current density given by Eq. (3.5),

$$i = \pm i_0 \times 10^{\pm \frac{E - E_{rev}}{b}} \tag{3.5}$$

where i_0 is the exchange current density with the positive sign for anodic reaction and negative sign for cathodic reactions, E is the electrical potential of the surface, E_{rev} is the reversible potential, and b is the Tafel slope constant. The exchange current density (i_0) is calculated using Eq. (3.6),

$$i_0 = i_{0ref} \left(\frac{C_{H^+}}{C_{H^+ ref}} \right)^{a_1} \left(\frac{C_{CO_2}}{C_{CO_2 ref}} \right)^{a_2} \left(\frac{C_{H_2CO_3}}{C_{H_2CO_3 ref}} \right)^{a_3} e^{-\frac{\Delta H}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}} \right)} \tag{3.6}$$

where i_{0ref} is the reference exchange current density, C_{H^+} is the concentration of H^+ ions at the surface, $C_{H^+ ref}$ is the reference concentration of H^+ ions, C_{CO_2} is the concentration of CO_2 species at the surface, $C_{CO_2 ref}$ is reference concentration of CO_2 species, $C_{H_2CO_3}$ is the concentration of H_2CO_3 at the surface, $C_{H_2CO_3 ref}$ is the reference concentration of H_2CO_3, a_{1-3}, are the power constants, ΔH is the activation energy, and T_{ref} is the reference temperature. The values for these constants are available for both cathodic reactions (Nesic et al., 1995) and anodic reactions (Nesic and Thevenot, 1996). Table 3.7 shows the values of electrochemical reaction rate constants for cathodic reactions used in different corrosion prediction models.

Table 3.7 Electrochemical reaction rate constants for electrochemical reactions (Nesic et al., 2001; Nordsveen et al., 2003; Al-Khateeb et al., 2018).

$i_{0,ref}$ (A/m^2)	a_1	$c_{H^+,ref}$ (mol/L)	a_2	$c_{CO_2,ref}$ (mol/L)	a_3	$c_{H_2CO_3,ref}$ (mol/L)	ΔH (KJ/mol)	T_{ref} (°C)	E_{rev} (V)	b (V)
0.05	0.5	10^{-4}	0	N/A	0	N/A	30	25	$-\dfrac{2.303\text{RT}}{F}pH$	$\dfrac{2.303RT}{0.5F}$
0.06	-0.5	10^{-5}	0	N/A	1	10^{-4}	50	20	$-\dfrac{2.303\text{RT}}{F}pH$	$\dfrac{2.303RT}{0.5F}$
1	2 if $pH < 4$ 1 if $4\leq pH < 5$ 0 if $pH\geq5$	10^{-4}	1 for $pCO_2 \leq 1$ bar 0 for $pCO_2 >1$ bar	0.0366	0	N/A	37.5	25	-0.488	$\dfrac{2.303RT}{1.5F}$

3.2.4 Velocity and Eddy Distribution in Straight Tubes

In a turbulent flow, momentum is usually transferred by viscosity and velocity fluctuations. The analogy of eddy diffusion of momentum is used to define the turbulent diffusion of mass and is given by Eq. (3.7)(Lin et al., 1953),

$$D_t = \left(\frac{y^+}{C}\right)^3 \frac{\mu}{\rho} \tag{3.7}$$

where constant C ranges from 8.9 to 14.5 (Lin et al., 1953; Davies, 1972). y^+ is the dimensionless distance from the wall calculated using Eq. (3.8),

$$y^+ = \frac{y u_\tau \rho}{\mu} \tag{3.8}$$

And,

$$u^+ = \frac{u}{u_\tau} \tag{3.9}$$

$$u_\tau = \sqrt{\frac{\tau_w}{\rho}} \tag{3.10}$$

where, u^+ is dimensionless velocity, y is the vertical distance from the wall (m), u_τ is the frictional velocity (m/s), and τ_w is the wall shear stress of the fluid (Pa).

Davies (1972) provided an empirical correlation shown using Eq. (3.11) to calculate the viscous sublayer thickness (δ) as a function of the Reynolds number and the diameter of the pipe, assuming that the edge of the viscous sublayer thickness is at $y^+ = 5$,

$$\delta = 25 \, Re^{-7/8} d \tag{3.11}$$

This correlation is based on the empirical friction factor equation for zero pressure gradient flow. It assumed $y^+ = 5$ at the edge of viscous sublayer thickness, which has been found to depend on the turbulence intensity. Some studies had found an average viscous sublayer thickness of $y^+ = 6.17$, questioning the validity of Eq. (3.11) (Popovich and Hummel, 1967).

A number of researchers (Nesic et al., 2001; Nordsveen et al., 2003) have used Eq. (3.11) for the viscous sublayer thickness and Eq. (3.12) for the turbulent diffusivity to predict the corrosion rate in pipe flow in the presence of a corrosion product film with a thickness δ_f.

$$D_t = \begin{cases} 0 & for \ y < \delta_f \\ 0.18 \left(\frac{y - \delta_f}{\delta - \delta_f}\right)^3 \frac{\mu}{\rho} & for \ y > \delta_f \end{cases} \tag{3.12}$$

There are several other empirical expressions which can be used (Notter and Sleicher, 1971; Rosen and Tragardh, 1995; Wang and Nesic, 2003b) and can be applied to a specific set of experimental conditions (Lin et al., 1953; Davies, 1972; Nesic et al., 2001; Wang and Nesic, 2003; Nordsveen et al., 2003; Kahyarian and Nesic, 2020). Therefore,

to make these predictions more robust, CFD will be used to provide accurate velocity distributions required for corrosion rate predictions.

3.2.5 Computational Fluid Dynamics Modelling

A 2-D axisymmetric pipe with a steady-state, isoviscous, isothermal and incompressible flow is used for the simulations. COMSOL Multiphysics® version 5.5 (COMSOL, 2016) is used to solve the governing equations given by Eq. (2.33) and Eq. (2.34).

The Shear Stress Transport (SST) $k - \omega$ turbulence model is used for its ability to resolve the flow in the viscous sublayer and buffer layers (Owen, 2018). The turbulent viscosity (μ_T) in the SST $k - \omega$ model is calculated using Eq. (2.57).

3.3 Axisymmetric Pipe Model

3.3.1 1D Model of Mass Transfer in a Pipeline

This section describes mass transfer modelling in an axisymmetric pipe with an overview of the computational domain used, initial and boundary conditions, numerical method and mesh convergence study. Figure 3.3 shows the computational domain used for the prediction of CO_2 corrosion in pipelines.

Figure 3.3. Computational grid and finite volumes for discretising the computational domain. The fluxes of species are computed on the finite volume boundaries, and concentrations and potential of species are calculated at the centre of finite volumes.

3.3.2 Initial and Boundary Conditions

It is considered that turbulence mixes all the species thoroughly, and uniform concentrations of species are considered for initial and boundary conditions. For the species not involved in the electrochemical reactions at the metal surface zero flux $(N_j = 0)$ is specified, and for the species j involved in the electrochemical reactions at the metal surface, the flux is specified by,

$$N_j = -\frac{i_j}{n_j F} \tag{3.13}$$

where i_j is the partial current of a specific electro-chemical reaction which contributes to the overall current density, n_j is the number of moles of electrons exchanged per mole of species j, and F is the Faraday constant (96485 C/mol).

3.3.3   Numerical Method

The viscous sub-layer thickness obtained from the CFD model is divided into a number of finite volumes (FV), as shown in Figure 3.3. There are a total of nFV finite volumes across the viscous sub-layer and a total of nFV+1 flux boundaries. Each FV has an upstream and downstream flux boundary. Eq. (3.1) can be rewritten as,

$$\frac{\partial c_j}{\partial t} = f(c_j) \tag{3.14}$$

Integrating over i^{th} FV leads to,

$$\left.\frac{\partial c_j}{\partial t}\right|_i = f_j(c)|_i \; \frac{-N_{i,j} + N_{i+1,j}}{\Delta x_i} + R_{j,i} \tag{3.15}$$

where $N_{i,j}$ is the flux of the jth species over the ith flux boundary, $N_{i+1,j}$ is the flux of the jth species over the i+1st flux boundary and $R_{j,i}$ is the rate of generation of species c_j in i^{th} FV.

The backward Euler method computes the approximations using Eq. (3.16),

$$c_j{}^{n+1} = c_j{}^n + \Delta t \, f(c_j{}^{n+1}, t^{n+1}) \tag{3.16}$$

where $f(c_j{}^{n+1})$ is the time derivative of the concentrations c_j at n+1st time step.

Here computing of $c_j{}^{n+1}$ from $c_j{}^n$ requires the solution of Eq. (3.17),

$$c_j{}^{n+1} - \Delta t \, f(c_j{}^{n+1}, t^{n+1}) = c_j{}^n \tag{3.17}$$

Eq. (3.17) is turned into a root-finding problem, and Newton's method is implemented to solve it. Let v^i be the ith Newton iterate approximation of c^{n+1} and $v_{old} = c^n$. Figure 3.4 shows the flow chart for the numerical method used for the corrosion rate predictions.

At each time step, an iterative matrix equation that has to be solved for $\Delta v = v^{i+1} - v^i$ subject to a specified error tolerance on the magnitude of Δv, starting from $v^0 = v_{old} = c^n$.

Numerical calculations are solved in the Python programming language. All the terms except flux of species are calculated at the centre of the FV. The flux of species is calculated at the finite volume boundaries (Nordsveen et al., 2003).

These equations need to be solved for the increments of Δx subject to charge balance equations for Finite Volumes 1 to nFV-1 (i.e. not adjacent to the corroding surface).

$$\Delta c^{H^+} + 2\Delta c^{Fe^{2+}} - \Delta c^{HCO3^-} - 2\Delta c^{CO3^{2-}} - \Delta c^{OH^-} + \Delta c^{Na^+} - \Delta c^{Cl^-} = 0 \tag{3.18}$$

This is done by replacing the equation for the Δc^{H^+} freedom in FVs 1 to nFV-1, with Eq. (3.18).

The conditions that the concentrations in the final Finite Volume, nFV-1, equal those calculated from the bulk steady state analysis are imposed by setting $\Delta c_j = 0$ for all concentration freedoms in the final Finite Volume, nFV-1.

Figure 3.4. Flow chart for the numerical method used for the corrosion rate predictions.

3.3.4 Mesh Convergence Study for Mass Transfer Model

A mesh convergence study was carried out to determine the sensitivity to the number of finite volumes used for the prediction of CO_2 corrosion at pH 4, velocity 1 m/s, pCO_2 1 bar and Temperature 20 ℃. The corrosion rate obtained at 25 nFV is 0.7563 mm/yr, which decreases to 0.7522 mm/yr for 50 nFV and stays constant for 55 and 60 nFV, as shown in Figure 3.5. Hence, the number of finite volumes chosen for all the corrosion predictions is 50 nFV.

Figure 3.5. Mesh convergence study for mass transfer model of corrosion prediction at pH = 4, velocity = 1 m/s, temperature = 20 ℃, pCO_2 = 1 bar and diameter of pipe = 0.015 m.

3.4 CFD Model of Axisymmetric Pipe

3.4.1 Geometry

Reynolds Averaged Navier Stokes (RANS) formulations can compute the flow field required for the computational domain shown in Figure 3.3. The axisymmetric pipe model is used to reduce computational time without compromising accuracy. The pipe diameter (D) is equal to 15 mm (Radius R =7.5 mm) with a length of 100 diameters (200*R) as the pipe length is chosen to obtain a fully developed flow at the outlet. The flow is simulated for different Reynolds numbers 18316 to 171795 by changing the inlet velocity. The axisymmetric pipe model is shown in Figure 3.6.

Figure 3.6. The axisymmetric computational domain used for the simulation (diameter = 15 mm).

3.4.2 Initial and Boundary Conditions

The velocity inlet and pressure outlet boundary conditions are used for the simulations. A flow velocity from 0.5 to 5 m/s was specified at the inlet with a fluid density (ρ) $998 \ \text{kg/m}^3$ and dynamic viscosity (μ) of $0.001 \ \text{Pa} \cdot \text{s}$. The values of k and ω are specified at the inlet as $0.00375 \ \text{m}^2/\text{s}^2$ and $11.18 \ \text{1/s}$ respectively. Table 3.8 describes the boundary conditions used for the simulation.

Table 3.8 Boundary conditions used for the CFD simulations of axisymmetric pipe.

Boundary	Description	p	u	k	ω
AB	Inlet	$\mathbf{n} \cdot \nabla p = 0$	$\mathbf{u} = \mathbf{u}_{in}$	$k = k_{in}$	$\omega = \omega_{in}$
CD	Outlet	$p = 0$	$\mathbf{n} \cdot \nabla u = 0$	$\mathbf{n} \cdot \nabla k = 0$	$\mathbf{n} \cdot \nabla \omega = 0$
AD	Symmetry axis	$\mathbf{n} \cdot \nabla p = 0$	$\mathbf{n} \cdot \nabla u = 0$	$\mathbf{n} \cdot \nabla k = 0$	$\mathbf{n} \cdot \nabla \omega = 0$
BC	Pipe wall	$\mathbf{n} \cdot \nabla p = 0$	$\mathbf{u} = 0$	$k = 0$	$\omega = 0$

3.4.3 Numerical Method

COMSOL Multiphysics ® (COMSOL, 2016) is based upon the finite element method (FEM) in which the field variables, in this case, $\boldsymbol{u}, p, k$ and ω are expressed using piecewise continuous basis functions on the discretised number of elements (de Boer et al., 2018). A form used for the approximation of a field variable φ is given using Eq. (3.19),

$$\varphi(x) = \sum_{i=1}^{n} \varphi_i N_i \tag{3.19}$$

where φ_i are the nodal values of the variable and N_i denotes the basis function (de Boer et al., 2018). A weighted integral (Galerkin) approach (Hysing, 2012) is then used to formulate nodal equations. The weak form of this formulation to obtain the discrete nodal equation for a field variable φ is shown using Eq. (3.20),

$$\int_{x_i}^{x_{i+1}} \mathcal{L}(\varphi) \, N_i(x) dx = \int_{x_i}^{x_{i+1}} s \, N_i(x) dx \tag{3.20}$$

where $\mathcal{L}(\varphi) = s$ is the strong form of the nodal equations. More details related to this can be found elsewhere (Hysing, 2012; COMSOL, 2016; de Boer et al., 2018).

3.4.4 Mesh Sensitivity

A structured quadrilateral mesh is used for the flow domain with rectangular boundary elements adjacent to the wall. The mesh adjacent to the wall should be fine enough to predict the fluid flow in the viscous sublayer accurately. The dimensionless distance (y^+) to the centre of the element adjacent to the wall was chosen to be 0.1. The mesh convergence study was conducted to assess the impact of the number of mesh elements on the flow velocity at the outlet. The number of elements varied along the pipe length for the mesh convergence study, as shown in Figure 3.7. The average velocity at the outlet did not change significantly for mesh elements more than 440000. Hence, to save computational time and power without compromising accuracy, the mesh with 440000 elements was chosen for the simulations.

Figure 3.7 Mesh convergence study for numerical simulations of axisymmetric pipe of diameter 0.015 m.

3.5 Results and Discussion

3.5.1 Turbulence Modelling – Justification of use of SST $k - \omega$ model

Figure 3.10 compares turbulent diffusivity profiles obtained using different turbulence models. The SST $k - \omega$ model is the most accurate turbulence model when compared against the turbulent diffusivity profile obtained using empirical correlation followed by the low Re $k - \omega$ model, Spalart Allmaras model and v2-f turbulence model. A similar method was used by Owen et al. (2019) to justify the use of the SST $k - \omega$ model by comparing the results with the Berger and Hau (1977) correlation for the calculation of the mass transfer coefficient in a straight pipeline.

Figure 3.8 Turbulent diffusivity profile comparison for different turbulence models at Reynolds number = 18316.

3.5.2 Prediction of Turbulent Boundary-layer Profiles

The values of VSL thickness from CFD and empirical correlation for VSL thickness given by Eq. (3.11) are compared for Reynolds numbers ranging from 18316 to 1717195. As shown in Table 3.9 and Figure 3.9, CFD predicts the values of viscous sublayer thickness reasonably well when compared with the empirical correlation, which is only applicable to specific cases. The percentage difference between the two predictions ranges from 11.51 to 17.50. Note that the empirical correlation in Eq. (3.11) is obtained by curve fitting using the friction factor obtained by Blasius for Re in the range between 3,000 and 100,000. Davies (1972) assumed that $y^+ = 5$ at the edge of VSL, which contradicts the findings by Popovich and Hummel (1967), as they had found that the average thickness of VSL was $y^+ = 6.2$ and proposed that it was most likely to be around $y^+ = 4.3$. These uncertainties in using the Davies (1972) empirical correlation can explain the discrepancy between the two sets of predictions shown in Table 3.9 and Figure 3.9.

Table 3.9 Viscous sublayer thickness predictions comparison between CFD and empirical correlation. The percentage differences between CFD and empirical VSL values are between 11.51 to 17.50.

Reynolds Number Re	Empirical VSL δ in mm	CFD VSL δ in mm	Percentage Difference %
18316	0.0698	0.0789	11.51
35686	0.0390	0.0456	14.52
52899	0.0276	0.0329	15.97
70019	0.0216	0.0255	15.44
87053	0.0179	0.0216	17.50
104138	0.0153	0.0179	14.80
121072	0.0134	0.0157	14.54
138079	0.0119	0.0139	14.39
155123	0.0108	0.0129	16.77
171795	0.0098	0.0118	16.53

Figure 3.9 Viscous sublayer thickness comparison between empirical and CFD values for Reynolds number 18316 to 171795.

Figure 3.10 compares the turbulent diffusivity profile obtained from CFD and the empirical Eq. (3.12) within the viscous sublayer thickness for Re = 18316. The difference between the shapes of the two profiles can be linked with the approximate constants obtained by fitting the experimental data for empirical correlation.

The values of viscous sublayer thickness and turbulent diffusivity profiles are then coupled with the mass transfer model to predict corrosion rates described in the following subsection.

Figure 3.10 Turbulent diffusivity comparison between CFD and empirical correlation for Re = 18316.

3.5.3 Verification of Corrosion Rate Predictions

Figure 3.11 shows the concentration profiles for dissolved species for pH = 4, velocity = 1 m/s, temperature = 20°C, pCO_2 = 1 bar and diameter of pipe = 0.015 m. At the metal surface, direct reduction and dissociation of carbonic acid results in depletion of H_2CO_3 species there, with a consequent increase in the concentration of HCO_3^-.

Figure 3.11 Concentration deviation of dissolved species from the values in bulk at for pH = 4, velocity = 1 m/s, temperature = 20 °C, pCO_2 = 1 bar and diameter of pipe = 0.015 m.

The corrosion rates shown in Figure 3.12, Figure 3.13, and Figure 3.14 are a function of significant parameters such as pH, Reynolds number, pCO_2 and temperature. The experimental data of straight pipe is obtained using potentiodynamic sweeps for a diameter of 15 mm, temperature of 20°C and pCO_2 of 1 bar (Nesic et al., 1995).

Figure 3.12, Figure 3.13, and Figure 3.14 show the corrosion rates for pH 4 to 6 at different Reynolds for conditions in the absence of protective corrosion product formation. As the Reynolds number increased, the corrosion rate increased for three pH values. The significant effect of turbulent flow in the absence of protective film is that it enhances the transport of species to and from the steel surface. At pH 4, there is a noticeable increase in the corrosion rate as the Reynolds number increases due to the presence of a much higher concentration of H^+ ions and enhanced transport of electrochemically active species towards and away from the metal surface. Conversely, at pH 5 and pH 6, the observed corrosion rate is much less than that at pH 4.

The CFD-driven mass transfer model predictions are compared with those based on the set of empirical correlations, the experimental dataset of Nesic et al. (1995), and the numerical model of Srinivasan (2015) for pH between 4 and 6, velocities between 1 m/s and 10 m/s, with a temperature of 20°C, pCO_2=1 bar and diameter of 0.015 m. The error bars on the experimental data are largest for pH 4, and it can be seen that each of the models are in reasonably good agreement with the experimental error bars for this case. Note that for pH 4 the corrosion rate predictions are particularly sensitive to the choice of electrochemical reaction rate constants, the precise values of which have been the subject of significant debate, see e.g. (Nordsveen et al., 2003; Al-Khateeb et al., 2018; Hu et al., 2018). Further sensitivity studies into the effects of the electrochemical reaction rate constants would be beneficial for pH values of 4 and below.

However, for the cases of pH 5 and 6, with much smaller error bars, it can be seen clearly that the new CFD-driven model is far superior to either the empirical or the Srinivasan (2015) model, and the limitations of the empirical approach are abundantly clear. For pH 5 and 6, typical discrepancies between the CFD-driven predictions and experimental data are only around 5%. This accuracy provides strong evidence that improving the modelling of flow in the viscous sublayer is very beneficial for mass-transfer based corrosion models and motivation to validate the approach for more complex, practically relevant corrosion scenarios.

Figure 3.12 CO_2 corrosion rate predictions for pH = 4, temperature = 20℃, pCO_2= 1 bar and diameter = 0.015 m and its comparison with the model based on empirical correlations of VSL and Dt, experimental data of Nesic et al. (1995), and numerical model of Srinivasan (2015).

Figure 3.13 CO_2 corrosion rate predictions for pH = 5, temperature = 20°C, pCO_2 = 1 bar and diameter = 0.015 m and its comparison with the model based on empirical correlations of VSL and Dt, experimental data of Nesic et al. (1995), and numerical model of Srinivasan (2015).

Figure 3.14 CO_2 corrosion rate predictions for pH = 6, temperature = 20°C, pCO_2 = 1 bar and diameter = 0.015 m and its comparison with the model based on empirical correlations of VSL and D_t, experimental data of Nesic et al. (1995), and numerical model of Srinivasan (2015).

3.6 Summary and Conclusions

A novel, CFD-driven modelling methodology for predicting CO_2 corrosion rates in pipelines is proposed. Computational fluid dynamics (CFD) based on the SST $k - \omega$ model, with its proven ability to resolve near-wall flow regions accurately, is used to provide accurate predictions of the viscous sublayer (VSL) thickness and turbulent diffusivities, which are then used within a mass transfer model of aqueous carbon dioxide (CO_2) corrosion. This new approach leads to significantly improved predictions of corrosion rate over a range of conditions and provides strong evidence that improving the modelling of flow in the VSL is very beneficial for mass-transfer based corrosion models. It also shows clearly that the current approach of using empirical models to predict VSL and associated turbulent diffusivities is both inaccurate and restrictive.

The key advantage of the new approach is its flexibility. It provides a firm scientific foundation for predicting corrosion rates by determining VSL and flow conditions in much more complex and practically relevant situations. The approach now needs further validation in such scenarios.

The present model is developed for application to cases with high electrolyte conductivity and low temperatures, for which electro-migration and $FeCO_3$ corrosion product formation can both be neglected. Since the goal of the study is to assess the benefits of determining the VSL conditions accurately, it is beneficial to avoid corrosion product formation conditions as this obviates the need for introducing additional empiricism into the corrosion model. However, if required, the model presented here can be extended to account for electro-migration by the addition of the term $z_j u_j F c_j \nabla \varphi$ in Eq. (3.3), whereas the inclusion of $FeCO_3$ The formation of corrosion products on the metal surface will necessitate a time-dependent solution scheme similar to those developed in the references (Nesic et al., 2001; Nordsveen et al., 2003). Finally, note that the electrochemical reaction rate constants can have a strong influence on corrosion rate predictions, particularly for lower pH values (Nordsveen et al., 2003; Al-Khateeb et al., 2018; Hu et al., 2018). It would, therefore, also be useful to carry out a systematic investigation into their influence on corrosion rate predictions, for example, using machine-learning methods. This approach is becoming increasingly popular in CFD-based simulations (Brunton et al., 2020).

Chapter 4: Application of Machine Learning Modelling for the Prediction of CO_2 Corrosion in Straight Pipelines

A CFD-driven mass transfer model developed for the prediction of CO_2 corrosion was elaborated on in the previous chapter. It has shown the limitations of using empirical correlations to predict fluid flow features and how CFD has the ability to construct a reliable tool in the areas of corrosion prediction.

4.1 Introduction

Measuring defect depth (depth of corrosion) and corrosion rate influences the decisions regarding pipeline corrosion control (Xu et al., 2023). These decisions include selecting economically cheap and sustainable pipe material, predicting pipe service, optimum wall thickness selection, and inhibitor performance analysis (Aghaaminiha et al., 2021). However, inaccurate prediction of corrosion rate, whether high or low, often results in economic losses for the industry. To avoid this, it is essential to have an accurate corrosion prediction model that will aid in corrosion control and reduce economic losses.

A CFD-driven mass transfer modelling technique developed in this work has been proven to be a robust tool for the prediction of viscous sublayer thickness, turbulent diffusivity profile, and CO_2 corrosion prediction in practical operating conditions. CO_2 corrosion is affected by several factors, which include temperature, flow rate, partial pressure of CO_2, pH and the presence of species such as H_2S. However, there are few studies into the factors affecting corrosion rates. Machine learning techniques offering efficient, non-linear data processing have shown great promise for corrosion prediction model development (Qi et al., 2019). The development of high-speed computing in the last couple of decades, in particular, has led to the application of machine learning methods such as Artificial Neural Networks (ANN) (Bassam et al., 2009; Weckman et al., 2010; De Masi et al., 2015; Chou et al., 2017; Pai et al., 2020), Random Forest (RF) (Ossai, 2019; Aghaaminiha et al., 2021; Ben Seghier et al., 2022), Support Vector Machines (SVM) (Lee et al., 2013; Ji et al., 2015; Hatami et al., 2016; Chou et al., 2017), Gaussian Process (GP) (Zhang et al., 2022) and Deep Learning (DL) (Ossai, 2019).

There has been significant interest in applying machine learning modelling to predict corrosion in oil and gas pipelines in the last two decades. However, some machine learning models still experience issues such as over-fitting and local minimum in the training process and difficulty adjusting hyper-parameters, resulting in lower accuracy and high calculation costs (Xu et al., 2023). In addition, very little attention is given to the sample points of the variables used in machine-learning models to predict corrosion rates. In the current study, the most popular techniques, such as ANN, GPR, SVR and RF are applied to predict CO_2 corrosion rate in straight pipelines.

Considering the capabilities of machine learning models, the following objectives were established:

- Find out the range for operating parameters in oil and gas industries and create the design of experiments using optimal Latin Hypercube sampling (OLHS).
- Obtain predictions at sampling points using the CFD-driven mass transfer model constructed in the previous chapter.
- Check the significance of each input variable.

- Construct machine learning models using ANN, GPR, SVR, and RF methods. Compare these machine learning models to choose an optimal machine learning model.

The section below covers the concise theoretical background of OLHS, PCA and machine learning methods. This is followed by training machine learning models and tuning their hyperparameters. The optimum machine learning model is then selected based on error metrics. The results section compares the metrics of machine learning models to provide the best-performing machine learning model.

4.2 Machine Learning Models

Machine learning techniques create predictive models for large datasets. It becomes useful when the output depends on several input variables to establish a relationship between them (Aghaaminiha et al., 2021). Machine learning is vital in data mining, difficult-to-program applications, and customised software applications (Michel, 1997). Machine learning is usually classified into supervised, unsupervised, and reinforcement learning, as discussed in Chapter 2. The current section outlines the machine learning models used to predict corrosion rates.

Artificial neural network (ANN) is one of the most popular machine learning techniques for creating data-driven models. In the architecture of ANN, the input layer consists of input variables (x_i) is connected to a hidden layer with weights(w_{ij}). A weighted sum ($F_j = \sum_i w_{ij}x_i$) of the inputs from the input layer is calculated at each node in the first hidden layer. An activation function is then used to transform F_j and is an output from each node. The output from each node in the first hidden layer becomes the input for the second hidden layer. This process is repeated until the output layer is reached. Weights that connect the nodes are adjusted to minimise the loss function between the predicted and target values.

SVR is a variation of SVM used for regression analysis. It is based on finding a hyperplane in higher dimensional space where the non-linear relationship between input and output variables may become linearly separable (Aghaaminiha et al., 2021). Different kernel functions, such as linear, polynomial, and radial basis functions (RBF), are used to map the input sample points to a higher dimensional space. Then, the hyperplane is selected in such a way that it maximises the margin between the hyperplane and the closest data points (Cherkassky and Ma, 2002).

GPR is a non-parametric regression analysis technique. It maps out the relationship between input and output variables, assuming that the target variable has a multivariate Gaussian distribution (Rasmussen and Williams, 2006). It consists of the mean function and covariance function; the former is an unknown function that needs to be modelled, whereas the latter captures the correlation between input and output variables. Bayesian inference is then used to obtain a probability distribution over the function's predicted values, allowing point estimates of the target variable (António and Rodrigues, 2021).

Random forest, an ensemble machine-learning technique, is used for both regression and classification problems (Quinlan, 1987). This technique consists of multitudes of decision trees in which smaller subsets (leaves) are created around input variables by splitting (branching) a dataset. As a result, these subsets will have minimal variances in the outcome values. This splitting is carried out until the terminal condition is met, which

could be either the number of splits or a cut-off value for the standard deviation for each subset. Following this, once the terminal condition is met, the average value in the leaf is the predicted outcome for that set of inputs. Random forest has two hyper-parameters: the number of trees and the maximum number of features that can be split.

4.3 Materials and Methods

Figure 4.1 shows the workflow for predicting corrosion rates using machine learning models. The modelling starts with defining a range of values for all the parameters, followed by creating sampling points using OLHS. The generated sampling points were then used to obtain corrosion rates from the CFD-driven mass transfer model described in the previous chapter. A dimensionality reduction technique, principal component analysis (PCA), is used to check if the large dataset can be transformed into a smaller one without losing significant information. Machine learning models such as ANN, SVR, RF and GPR are compared against each other to find out the optimal machine learning model for the prediction of CO_2 corrosion rates.

Figure 4.1 Flow chart for the CFD-enabled machine learning model.

4.3.1 Optimal Latin Hypercube Sampling

Efficient sampling of a design space obtains the maximum information from a limited sample (Yondo et al., 2018). Latin Hypercube sampling (LHS) (McKay et al., 1979) (Yondo et al., 2018) divides **N** design variables into equal **P** levels. For LHS DoE, there is only one sample point for each level, and this technique is classified into random Latin Hypercube sampling and optimal Latin Hypercube (OLH) sampling. This classification depends on how the sampling points are distributed in the design space (Bates et al., 2004). OLH samples used in this study are generated with the help of the permutation genetic algorithm (permGA) and based on the methodology developed by (Bates et al., 2004). This methodology solves a combinatorial optimisation problem of minimisation of the potential energy of repulsive forces for a set of sample points with a unit mass (Bates et al., 2004). The nature of the repulsive forces is such that the sample points are pushed away from each other to get an even distribution of sample points (Bates et al., 2004). The magnitude of these repulsive forces depends on the square of the distance between the points and is given by Eq. (4.1),

$$\sum_{p=1}^{P} \sum_{q=p+1}^{P} \frac{1}{r_{pq}^2} \rightarrow \min$$

(4.1)

where r is the distance between sample points p and q.

Table 4.1 summarises the initial variables used for the OLH sampling. These initial variables are chosen so that a protective layer of iron carbonate does not form on the surface. The diameter of the horizontal pipe (d) is 0.015 m. The total number of sample points generated from this technique is 100.

Table 4.1 Summary of the initial variables used for the OLH sampling.

Parameter	Base value	Minimum value	Maximum value
pH	4	4	6
pCO_2 bar	1	0.1	10
Temperature (T) °C	20	20	90
Velocity (u) m/s	1	0.5	10

The distribution of the initial variables is shown using the boxplots in Figure 4.2. The boxplots are helpful in showing the distributions of the variables around the measures of central tendency, i.e., mean, median and mode (Abbas, 2016) and measures of variation, i.e., range, standard deviation, and variance. The distribution of each variable is even around the measures of central tendency, which supports the fact that the data generated has symmetrical distributions.

Figure 4.2 Box plots to show the distribution of initial variables. The horizontal axis represents input variables, and the vertical axis represents the magnitude of data.

A quartile value breaks the distribution into four equal parts. The first quartile, the 25th percentile, divides the dataset between the first one-fourth and the last three-fourths of the total dataset. The third quartile, or 75th percentile, divides the dataset into the first three-fourths and the last one-fourth of the total dataset. The interquartile range is the range between the first quartile and the third quartile. This is useful for understanding the summary statistics of the corrosion data. Table 4.2 shows the descriptive statistics for the input and output variables of the corrosion modelling. Skewness measures the asymmetry of a distribution. It is evident from the values of skewness that the distribution of all the input variables is symmetrical. In contrast, the output corrosion rate has a positive distribution with the value of skewness equal to 0.32.

Table 4.2 Descriptive statistics of the corrosion sample data to build a surrogate model to find the optimal set of operating parameters.

Parameter	Range	Mean	First Quartile	Median	Third Quartile	Skewness
pH	4-6	5	4.50	5.00	5.49	0.00
pCO_2 (bar)	0.10-10	5.05	2.57	5.05	7.52	0.00
Temperature (T) (°C)	20-90	55	37.50	55	72.49	0.00
Velocity (u) (m/s)	1-10	5.50	3.25	5.50	7.74	0.00
Corrosion rate (mm/yr)	0.33-14.16	5.61	2.83	5.26	7.65	0.32

4.3.2 Data Normalisation

All the input variables and the output variable $(x_1, x_2, x_3, x_4, y)_j$ $(j = 1,2,3,\dots,n)$ are normalised to the range [0,1] for better representation using Eq.(4.2) and Eq.(4.3)

For $i = 1,2,3,4$

$$\frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}} \tag{4.2}$$

$$\frac{y - y^{min}}{y_i^{max} - y_i^{min}} \tag{4.3}$$

Normalised data is then fed into the PCA algorithm to assess the influence of all the design variables. The predictions obtained from the machine learning models can be easily transformed from normalised to physical space using the above equation.

4.3.3 Principal Component Analysis

PCA of the dataset obtained from OLH sampling is used to account for variations in the input variables used in this study: pH, pCO_2, T and u. This technique begins with standardisation of the continuous initial variables to ensure that each variable contributes equally to the analysis, an important step to decrease the dominance of initial variables over the remaining ones with smaller differences in value. This step is later followed by computing the covariance matrix and eigenvalues and eigenvectors to determine the principal components of the data (Jollife and Cadima, 2016). These principal components are linear combinations of initial variables, in this case, pH, pCO_2, T and u. Table 4.3 shows the values such as Eigenvalues, the proportion of those Eigenvalues, cumulative variance and cumulative variance percentage associated with each principal component.

Table 4.3 Principal component analysis of 4 design variables corrosion prediction model.

Principal component (PC)	PC 1	PC 2	PC 3	PC 4
Explained variance	0.26	0.25	0.24	0.25
Cumulative variance	0.26	0.51	0.75	1.00
Cumulative percentage (%)	26	51	75	100

Figure 4.3 Pareto chart showing variance among each of the principal components. All principal components exhibit equal influence, thereby contributing equally to the overall variance in the dataset. The solid line shows the cumulative variance of the principal components.

Figure 4.4 Scree plot showing eigenvalues of each principal component.

Figure 4.4 shows the Pareto chart of the data variance for each principal component. It is evident from Table 4.3 and Figure 4.4 that all principal components are significant for representing the data and, hence, cannot be neglected. Abbas et al. (2018) in their study assumed that it was possible to represent the data using the first two principal components as those two principal components accounted for 81 % of the data for pH, pCO_2, T and u. The remaining 19% was ignored in their study to proceed with the

calculation of the relative influence of initial variables on those two components (Abbas et al., 2018). On the other hand, the current study shows that the initial variables are uncorrelated or weakly correlated and, hence, cannot be ignored. Following PCA, it was found that all design variables are equally influential.

4.3.4   Modelling Process

CO_2 corrosion rate predictions obtained from CFD-driven mass transfer models are used as a dataset for machine learning modelling. The input variables for the machine learning models are pH, pCO_2, T and u. The output variable is CO_2 the corrosion rate in mm/yr. All the programs are written in Python programming language.

Determination of Model Parameters

For the ANN model, the set of hyperparameters consists of the number of hidden layers, nodes per hidden layer, batch size and Epochs. The ANN model was first studied for the variation in the number of hidden layers and nodes per hidden layer by keeping the epochs and batch size constant. Figure 4.5 shows the variation in RMSE for the number of hidden layers (nHL) and the number of nodes per hidden layer. It was found that the 15 neurons per hidden layer for 4 hidden layers showed the lowest RMSE. Hence, this configuration was considered for the final ANN model to determine the impact of batch size and Epochs. Figure 4.6 shows that a batch size of 5 and Epochs equal to 10000 provided the lowest RMSE value.

Figure 4.5 Selection of the number of hidden layers and nodes per hidden layer for the final ANN model.

Figure 4.6 Variation of RMSE value with the change in batch size and Epochs.

A grid search is a technique in which a grid of parameters is created to try all the possible combinations. This functionality is available in Scikit-learn named GridSearchCV, where CV stands for cross-validation. It takes a dictionary representing the parameters to be tuned and a model to train. Finally, a scoring criterion for evaluating the performance of the cross-validated model on the test dataset is set as a mean squared error (MSE). SVR, RF and GPR are tuned using this technique; the details are below.

There are four hyperparameters when it comes to SVR viz kernel function, epsilon(ε), cost (C) and gamma (γ). It is essential to study the influence of ε value as it determines the margin of tolerance where no penalty is given to errors. Cost (c) is a regularization parameter that decides the penalty for the misclassification of data points. Gamma (γ) the parameter determines the shape of the decision boundary, which controls the influence of the kernel function. A range of values were studied for these hyperparameters. The kernel functions studied for the performance are radial basis function (RBF), linear and polynomial. The value of cost C varied from 1 to 10000 for the γ values of 0.01, 0.1, and 1 and ε values of 0.1, 0.2, 0.3, and 0.5. The dataset was split into five groups during the tuning of the parameters. Table 4.4 shows the best parameters obtained using a cross-validated grid search.

Table 4.4 Tuned hyperparameters for the SVR model.

Parameter	Value
Kernel function	RBF
Epsilon (ε)	0.1
Cost (C)	10
Gamma (γ)	0.1

91

Different parameters need to be tuned for RF based on a negative RMSE value. The negative RMSE value is related to how the cross-validation function maximises the function; the higher, the better. These parameters include the number of trees, maximum depth of trees, and bootstrap. A Boolean input is provided for bootstrap with the 'default = True', resulting in training each tree on a subset of the dataset. If this input is changed to False, the whole dataset will be used to build each tree. For the current dataset, the number of trees varied from 10 to 100 with a maximum depth of trees from none to 10 for bootstrap = True. Table 4.5 and Figure 4.7 show the best parameters obtained using a cross-validated grid search.

Table 4.5 Tuned hyperparameters for the RF model.

Parameter	Value
Number of trees	10
Maximum depth of trees	None

Figure 4.7 Variation of negative root mean square with the number of trees and the maximum depth of trees.

Table 4.6 shows the parameters used for the GPR model. First, the prior of the GP needs to be specified for the GPR, which includes the prior mean and prior's covariance. Then, the kernel object is passed to specify the prior's covariance and its hyper-parameters are optimised while fitting the GPR model. A Matérn Kernel is used for the current model with a length scale value of 1. This kernel has an additional parameter, v, used to smooth the resulting function. The higher the value of v, the smoother the approximated function is.

Table 4.6 Tuned hyperparameters for the GPR model.

Parameter	Value
Alpha	0.001
Kernel	Matérn
Length scale	1.0
v	1.5

4.4 Results and Discussion

The machine learning regression algorithm aims to learn an optimal model for predicting an outcome based on input vectors (Zahedi et al., 2018). The tuned hyper-parameters were obtained for all the machine learning models and then applied to the corrosion rates data obtained from the CFD-driven mass transfer model. It is important to note that the hyper-parameters varied in the current study were assumed to be the most significant parameters of machine learning modelling. In contrast, all the other parameters were used as the default for the modelling. A similar methodology for choosing hyperparameters has been used recently by Aghaaminiha et al. (2021).

A dataset was split into training and test datasets, where 80% of the total dataset was used for training the model and 20% for the test data.

Table 4.7 Metrics of machine learning models.

Model	**RMSE**	$\mathbf{R^2}$
ANN	0.04646	0.9622
SVR	0.0741	0.9269
RF	0.0937	0.83
GPR	0.04898	0.9673

It can be seen from Table 4.7 that the ANN model has the lowest RMSE value of 0.04646 for the corrosion prediction model for design variables pH, pCO_2, T and u, followed by GPR, SVR and RF. ANN model performs well as it can map complex relationships between inputs and outputs (Wen et al., 2019). However, ANN models are considered highly sensitive regarding input data quality, which may result in lower performance due to noisy or missing data (Mrzygłód et al., 2020). In addition to this, ANN models result in overfitting for complex problems when training data is limited. On the other hand, the GPR model considers the uncertainty in the data and handles noisy data more effectively than ANNs (Zhang et al., 2022). The current study shows that the GPR model works well for limited training datasets. The lower performance of the SVR model for the current dataset can be attributed to the selection of hyper-parameters such as kernel function, epsilon parameter, cost and regularisation parameter shown in Table 4.4. RF model showed the worst performance compared to the other three machine learning models due to the limited training dataset.

Figures 4.8 to 4.11 compare corrosion rates predicted using the CFD-driven mass transfer and machine learning models of ANN, GPR, RF, and SVR models, respectively. The solid line on these figures shows x=y and the metrics corresponding to these figures is R^2 shown in Table 4.7. Based on this metric, the GPR model performed best, followed by ANN, SVR and RF. Again, this underlines that the GPR model works well for a limited training dataset.

The findings obtained in the current study are subjected to several limitations. The models' dependence on a limited training dataset may constrain the effective generalization of the findings. In addition, the selection of hyperparameters, as observed in the lower performance of the SVR model, emphasizes the difficulty of optimizing the model parameters to obtain the best performance of machine learning models.

Figure 4.8 Corrosion rate predictions using ANN model show actual corrosion rate (CFD driven mass transfer model) vs predicted corrosion rate (machine learning model). The solid line represents x = y.

Figure 4.9 Corrosion rate predictions using the GPR model show the actual corrosion rate (CFD-driven mass transfer model) vs predicted corrosion rate (machine learning model). The solid line represents x = y.

Figure 4.10 Corrosion rate predictions using the RF model show the actual corrosion rate (CFD-driven mass transfer model) vs predicted corrosion rate (machine learning model). The solid line represents x = y.

Figure 4.11 Corrosion predictions using the SVR model show the actual corrosion rate (CFD-driven mass transfer model) vs predicted corrosion rate (machine learning model). The solid line represents x = y.

4.4.1 Surrogate Modelling based Optimisation

Figure 4.12 shows different responses between pH, pCO_2, T and u with corrosion rate. The obtained surrogate models for the corrosion rate predictions show the relationship between operating parameters and corrosion rates. The optimal machine learning model, ANN for corrosion rate prediction, has helped in understanding how the operating parameters influence corrosion rates, which allows pipeline designers to make more informed decisions. The generation of response surfaces also helps in obtaining the best combination of these parameters to lower the corrosion rate. The expansion of design space for input variables provides corrosion rates for different operating conditions, which is useful in highlighting critical regions.

Upon training, surrogate models significantly reduce the dependency on expensive and time-consuming experiments (Yondo et al., 2018). This is one of the main advantages of the current methodology, as it provides an efficient exploration of a wide range of scenarios.

Figure 4.12 The response surfaces of corrosion rate generated using ANN model to find corrosion rate in single phase flow pipelines. Response surfaces of (a) pH and pCO_2 (b) T and u (c) pH and u (d) pH and T (e) pCO_2 and T (f) pCO_2 and u shows a relation with corrosion rate. The response surfaces show how changes in factors such as pH, pCO_2, T, and u affect the rate of corrosion in pipelines.

4.5   Conclusions

CFD-enabled machine learning modelling is used for the prediction of CO_2 corrosion in horizontal pipelines based on significant parameters such as pH, pCO_2, temperature and

velocity to find the best-performing machine learning model. A design of experiments (OLHS) was used to sample the dataset, which was then used to obtain corrosion rates from the CFD-driven mass transfer model. A dimensionality reduction technique – PCA- is used to determine the significance of each parameter in the prediction of corrosion rate. It is found that all four input variables are essential to model corrosion rates.

After this, machine learning models such as ANN, GPR, SVR and RF are applied to corrosion rate predictions. The ANN model performs well compared to GPR, SVR and RF based on RMSE value. To summarise,

- It is found that several variables impact the prediction of corrosion rates, out of which pH, pCO_2, T and u are the most important ones. A well-trained machine learning model is a model trained on sufficient data, with optimized hyperparameters and validated on test-data. A well-trained machine learning model could eliminate the need for a flow field obtained using CFD. This will save a massive amount of computational time and power.
- A well-tuned ANN model reasonably predicts the output without overfitting the training data. This study has also shown that ANN and GPR models are suited for small-medium-sized (a few hundred to a few thousand sample points) data.
- The response surfaces generated using the ANN model offer a comprehensive understanding of the correlation between operating conditions and corrosion rates.

This CFD-enabled machine learning modelling can be extended to more complex situations such as flows in elbows, expansion/constriction or even multiphase flow situations.

Chapter 5: Machine Learning Enabled Parameter Optimisation of a Mathematical Model of CO$_2$ Corrosion Predictions

5.1 Introduction

As discussed earlier, the nature of CO$_2$ corrosion is electrochemical. The evolution of hydrogen ions takes place, which provides the necessary electron sink for the dissolution of iron. For CO$_2$ corrosion, the hydrogen evolution reactions are cathodic, with hydrogen as a product for all the reactions. These reactions are part of mass transfer modelling for the prediction of CO$_2$ corrosion. Table 3.6 shows all the electrochemical reactions that take place in the current mass transfer model of CO$_2$ corrosion prediction, which includes the reduction of H$^+$ and H$_2$CO$_3$ along with the anodic dissolution of iron. The values for electrochemical reaction rate constants required in Eq. (3.5) and Eq. (3.6) are available for both cathodic reactions (Nesic et al., 1995) and anodic reactions (Nesic and Thevenot, 1996). There is no consensus among researchers when it comes to the use of electrochemical reaction rate parameters for the prediction of corrosion rates (Nesic et al., 2001; Nordsveen et al., 2003; Al-Khateeb et al., 2018; Nešić et al., 2019) with reference to their original source (Nesic et al., 1995; Nesic and Thevenot, 1996).

Tuning of these parameters in different studies underlines a need to find the set of reliable values that will be applicable to the broader range of data and assess their sensitivity with respect to corrosion rate prediction. This can be achieved through the use of surrogate modelling-based optimisation (SBO). Physics-based modelling poses constraints in terms of computational time and power for the analysis of high-fidelity simulations (Yondo et al., 2018). Although there is an advancement in computing resources, it is inadvisable to be completely dependent on these computationally demanding simulations. To circumvent this, 'cheap-to-evaluate' mathematical models, also known as surrogate models or response surface models, were introduced (Box and Draper, 1987). A surrogate model replicates high-fidelity simulation by providing an evaluation of an objective function at any sample point in the design space. A good understanding of design space, physics-based modelling and optimisation techniques is essential to extract the required information from SBO (Yondo et al., 2018). The steps carried out in the study are given as follows,

- Find out the range of electrochemical reaction rate constants from the literature.
- Use a design of experiment technique to create sampling point distribution.
- Obtain CFD-driven mass transfer model corrosion rate predictions for these sampling points.
- Construct machine learning models using ANN, GPR, SVR and RF methods and tune their hyperparameters.
- Choose the optimal machine learning model by comparing it with each other based on evaluation metrics.
- Solve a surrogate modelling-based optimisation problem to minimise an error between predicted values of corrosion rate and experimental values of corrosion rates.

In the current study, parameters related to Tafel's equation used to describe the kinetics of electrochemical reactions are optimized for mass transfer modelling of CO$_2$ corrosion. A design of experiment (DOE) technique, viz., random Latin hypercube sampling (RLHS), is used

to obtain the sampling points for which function evaluations (corrosion rate) were obtained. Machine learning modelling techniques such as ANN, SVR, GPR, and RF are then used to construct a surrogate model. A best-performing surrogate model is then integrated with different optimisation techniques such as Nelder-Mead method (Nelder and Mead, 1965), Powell's method (Powell, 1964) and Hooke-Jeeves method (Hooke and Jeeves, 1961) for optimisation.

5.2 Materials and Methods

An initial step consists of finding out the range of all the electrochemical reaction rate parameters from the literature. Random Latin hypercube sampling (RLHS) is then used to create sample points to obtain corrosion rates at each sample point. A total of 12 design variables, which are part of Tafel's equation for the kinetics of electrochemical reactions, are used in the current study. The PCA technique is then used to check if the dimensionality of the model can be reduced. Machine learning models such as ANN, SVR, RF, and GPR are then used for surrogate modelling, followed by an optimisation technique to find optimal electrochemical reaction rate constants. This process is shown in Figure 5.1.

Figure 5.1 Flow chart of methodology to find the optimal set of electrochemical reaction rate constants using machine learning enabled surrogate modelling.

Table 5.1 and Table 5.1 show the values of electrochemical reaction rate constants for cathodic and anodic reactions, respectively.

Table 5.1 Electrochemical reaction rate constants for cathodic reactions used in different studies.

Electrochemical reaction: $2H^+ + 2e^- \rightarrow H_2$											
$i_{0,ref}$ (A/m^2)	a_1	$c_{H^+,ref}$ (mol/L)	a_2	$c_{CO_2,ref}$ (mol/L)	a_3	$c_{H_2CO_3,ref}$ (mol/L)	ΔH (KJ/mol)	T_{ref} (°C)	E_{rev} (V)	b (V)	Source
0.05	0.5	10^{-4}	0	N/A	0	N/A	30	25	$-\dfrac{2.3RT}{F}pH$	$\dfrac{2.3RT}{2F}$	(Nesic et al., 2001)
0.05	0.5	10^{-4}	0	N/A	0	N/A	30	25	$-\dfrac{2.3RT}{F}pH$	$\dfrac{2.3RT}{2F}$	(Nordsveen et al., 2003)
0.05	0.5	10^{-4}	0	N/A	0	N/A	30	25	$-\dfrac{2.303RT}{F}pH$	$\dfrac{2.303RT}{0.5F}$	(Al-Khateeb et al., 2018)
Electrochemical reaction: $2H_2CO_3 + 2e^- \rightarrow H_2 + 2HCO_3^-$											
$i_{0,ref}$ (A/m^2)	a_1	$c_{H^+,ref}$ (mol/L)	a_2	$c_{CO_2,ref}$ (mol/L)	a_3	$c_{H_2CO_3,ref}$ (mol/L)	ΔH (KJ/mol)	T_{ref} (°C)	E_{rev} (V)	b (V)	Source
0.06	-0.5	10^{-5}	0	N/A	1	10^{-4}	50	20	$-\dfrac{2RT}{F}pH$	$\dfrac{2.3RT}{2F}$	(Nesic et al., 2001)
0.06	-0.5	10^{-5}	0	N/A	1	10^{-4}	50	20	$-\dfrac{2RT}{F}pH$	$\dfrac{2.3RT}{2F}$	(Nordsveen et al., 2003)
0.018	-0.5	10^{-5}	0	N/A	1	10^{-4}	50	20	$-\dfrac{2.303RT}{F}pH$	$\dfrac{2.303RT}{0.5F}$	(Al-Khateeb et al., 2018)

Table 5.2 Electrochemical reaction rate constants for anodic reactions used in different studies.

$i_{0,ref}$ (A/m^2)	a_1	$c_{H^+,ref}$ (mol /L)	a_2	$c_{CO_2,ref}$ (mol /L)	a_3	$c_{H_2CO_3,ref}$ (mol/L)	ΔH (KJ/mol)	T_{ref} (°C)	E_{rev} (V)	b (V)	Source
							Electrochemical reaction: $Fe \rightarrow Fe^{2+} + 2e^-$				
1	1 for pCO$_2$ < 1 0 for pCO$_2 \geq$1	10^{-4}	2 for pH < 4 1 if 4<pH < 5 0 if pH>5	0.0366	0	N/A	37.5	25	-0.488	0.03 for pH<4 0.08 for 4<pH<5 0.12 for pH>5	(Nesic et al., 2001)
1	1 for pCO$_2$ < 1 0 for pCO$_2 \geq$1	10^{-4}	2 for pH < 4 1 if 4 < pH < 5 0 if pH > 5	0.0366	0	N/A	37.5	25	-0.488	0.03 for pH<4 0.08 for 4<pH<5 0.12 for pH>5	(Nordsveen et al., 2003)
1	2 if pH < 4 1 if 4$\leq$ pH $\leq$ 5 0 if pH>5	10^{-4}	1 for pCO$_2 \leq$ 1 bar 0 for pCO$_2$ >1 bar	0.0366	0	N/A	37.5	25	-0.488	$\dfrac{2.303RT}{1.5F}$	(Al-Khateeb et al., 2018)

5.3   Random Latin Hypercube Sampling (RLHS)

RLHS is a type of LHS DoE in which points are obtained using random sampling and differs from optimal Latin hypercube sampling (OLHS) based on the distribution of sampling points in a design space (Bates et al., 2004; Yondo et al., 2018). A tremendous amount of computational time is required for the generation of OLHS. In addition, OLHS requires advanced computational resources for the higher number of design variables, which makes it impractical to use in the current study. Hence, to overcome this, RLHS is used to obtain a total sample point of 1000 for 12 design variables (Afzal et al., 2017).

Table 5.3 summarises the parameters of electrochemical reaction rates and their range used for RLHC sampling. The parameters related to the two cathodic reactions, hydrogen ions and carbonic acid reduction and one anodic reaction, iron dissolution reaction, are also shown in Table 5.3. The base value represents the value used in CFD driven mass transfer model of CO_2 corrosion discussed in Chapter 3. The minimum and maximum values are obtained by varying the base value by 20% for the parameters with a single value in the literature. This range is assumed to avoid invalid or impractical solutions while considering the physics constraints for each input variable. The corrosion rates were then predicted using CFD driven mass transfer model for a pipe diameter of 0.015 m, pH 4, velocity of 1 m/s, temperature of 20 °C, and pCO_2 of 1 bar.

The parameter range differs by several orders of magnitude, and hence, it is crucial to normalize the data. The normalized data is beneficial in terms of avoiding bias from the machine learning algorithms as these models assume that the input parameters are on the same scale (Izonin et al., 2022). In addition to this, normalized data become flexible for the interpretability and comparability of the input variables (Alasad & Bhaya, 2017). The following subsection shows the use of the data normalisation technique for the current dataset.

Table 5.3 Nomenclature of the 12-design variable corrosion modelling problem.

	Parameter	Notation used	Base value	Minimum value	Maximum value	Comments
Hydrogen ions reduction reaction	Reversible potential ($E_{rev,H}$) (V)	x_1	-0.232	-0.279	-0.186	20 % change
	Tafel slope(b_H) (V)	x_2	0.116	0.101	0.131	(Nordsveen et al., 2003)
	The activation energy (ΔH) (J/mol)	x_3	30000	27000	33000	20 % change
	Power constant ($a_{1,H}$)	x_4	0.5	0.4	0.6	20 % change
	Reference exchange current density (i_{0ref},H) (A/m^2)	x_5	0.05	0.04	0.06	20 % change
Carbonic acid reduction reaction	Reference exchange current density (i_{0ref},H_2CO_3) (A/m^2)	x_6	0.06	0.048	0.072	20 % change
	Power constant (a_{1,H_2CO_3})	x_7	-0.5	-0.6	-0.4	20 % change
	Activation energy (ΔH_2CO_3) (J/mol)	x_8	50000	40000	60000	(Kahyarian and Nesic, 2020)
Anodic dissolution reaction	Reference concentration of CO_2 species ($C_{CO_2 ref}$) (mol/L)	x_9	0.0366	0.02928	0.04392	20 % change
	Reversible potential ($E_{rev,Fe}$) (V)	x_{10}	-0.488	-0.51	-0.390	(Sainz-Rosales et al., 2022)
	Activation energy(ΔH) (J/mol)	x_{11}	37500	30000	45000	20 % change
	Tafel slope(b_{Fe}) (V)	x_{12}	0.0387	0.03	0.08	(Nordsveen et al., 2003)

5.4 Data Normalisation

All the input variables and the output variable $(x_1, x_2, x_3, \ldots, x_{12}, y)_j$ $(j = 1,2,3, \ldots, n)$ are normalised to the range [0,1] for better representation using Eq. (5.1) and Eq. (5.2). The parameters $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}$ are shown in Table 5.3, and y is the corrosion rate.

For $i = 1,2,3,4,..,12$

$$\frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}} \qquad (5.1)$$

$$\frac{y_i - y_i^{min}}{y_i^{max} - y_i^{min}} \qquad (5.2)$$

Normalised data is then fed into different machine-learning models to obtain corrosion rate predictions. These normalised values can be transformed back to physical space using the above equations.

5.5 Dimensionality Reduction Technique – Principal Component Analysis (PCA)

PCA, a dimensionality reduction technique, is used in the current study to determine the most significant variables in the dataset. In the current study of 12 design variables, the PCA technique transforms input features into uncorrelated principal components (Jollife and Cadima, 2016). Principal components are linear combinations of these 12 initial design variables and are ordered in such a way that the first component shows the maximum amount of variance of the dataset, followed by the second component, which tries to capture the remaining variance and so on (Abbas, 2016).

Table 5.3 shows the list of principal components, the amount of variance explained by each component, cumulative variance and cumulative percentage. The explained variance column indicates the proportion of total variance captured by each component from the original dataset. For example, PC1 has captured 0.098 or 9.8% of total variance, and PC2 has captured 0.092 or 9.2% of total variance. In contrast, cumulative variance shows the accumulated variance by each principal component at that particular point. The cumulative variance of PC2 is 0.191 or 19.2%, the sum of the explained variance of PC1 (0.098) and PC2 (0.092). In addition, cumulative variance identifies the number of components that retain a significant amount of information about the dataset. For the current dataset, the value of cumulative variance reaches 1.0 or 100% at the 12th principal component, exhibiting the importance of all the principal components to explain the variance in the original dataset. This suggests that all 12 design variables contribute towards the overall variance in the dataset. Hence, none of the original features can be considered redundant, and all 12 design variables are used for the machine learning modelling shown in the next section. Figure 5.2 shows the visual representation of the explained variance for each principal component.

Table 5.4 Prinicipal component analysis of 12 design variable problems to find the optimal set of electrochemical reaction rate constants.

Principal Component	Explained Variance	Cumulative Variance	Cumulative Percentage (%)
PC1	0.098	0.098	9.8
PC2	0.092	0.191	19.1
PC3	0.091	0.281	28.1
PC4	0.088	0.369	36.9
PC5	0.087	0.456	45.6
PC6	0.085	0.541	54.1
PC7	0.083	0.624	62.4
PC8	0.081	0.705	70.5
PC9	0.077	0.781	78.1
PC10	0.076	0.858	85.8
PC11	0.074	0.931	93.1
PC12	0.069	1.000	100

Figure 5.2 Principal component analysis (PCA) of electrochemical reaction rate constants for the prediction of corrosion rates in pipelines.

5.6 Machine Learning Modelling

Python programming language is used to write machine learning model programs. CO_2 corrosion rate predictions obtained from the CFD-driven mass transfer model based on different electrochemical reaction rate constants are used for machine learning modelling. An essential initial step following the data pre-processing for the machine learning modelling is to find the optimum set of hyperparameters for each machine learning model. The section below explains the tuning of hyperparameters for machine learning models.

Determination of Model Parameters

Various parameters must be tuned for each machine-learning model to obtain robust predictions. In the current study, for the ANN model, the set of hyperparameters consists of the number of hidden layers, nodes per hidden layer, batch size and Epochs. The ANN model was first studied for the variation in the number of hidden layers (nHL) and nodes per hidden layer without changing Epochs and batch size. After obtaining the optimal value of nHL and nodes per nHL, Epochs and batch size were varied to check for the best combination. This is shown in Figure 5.3 and Figure 5.4. The final set of hyperparameters based on the lowest RMSE value includes 50 nodes per hidden layer with 3 hidden layers, batch size of 10 and Epochs 50000.

Figure 5.3 Selection of the number of hidden layers and nodes per hidden layer for the ANN model of electrochemical reaction rate constants.

Figure 5.4 Selection of batch size and Epochs for the ANN model of electrochemical reaction rate constants.

Grid search cross-validation (GridSearchCV) technique is used to tune hyperparameters for SVR, RF and GPR. This technique involves the evaluation of the model's performance for different sets and combinations of hyperparameters to choose the optimal combination. An initial step in GridSearchCV involves defining a grid of hyperparameters for a machine-learning model. This grid of hyperparameters is provided using a dictionary in which values are ranges of parameters to be studied, and the names of the

hyperparameters are keys. Following this, an evaluation metric is selected to compare the performance of each combination of hyperparameters. The parameter grid and scoring metric are passed to the instantiated machine learning model and the GridSearchCV function, respectively. Finally, the 'best_params_' attribute returns the best hyperparameter combination, which is then used to obtain an improved machine-learning model.

For the SVR model, the hyperparameters studied are kernel function, epsilon(ε), cost (C) and gamma(γ), with the range of parameters shown in Table 5.4.

Table 5.5 Range of hyperparameters studied for the SVR model.

Parameter	Values
Cost (C)	[1,10,100,1000]
Epsilon (ε)	[0.1,0.2,0.3]
Gamma(γ)	[0.01,0.1,1]
Kernel function	[rbf, linear, poly]

A range of values was studied by fitting 5 folds for each of the 108 candidates to determine the optimum values. To elaborate, fitting 5 folds means the model is trained and tested five times. These combinations were compared with each other based on evaluation metrics. Figure 5.5 shows the plot of a combination of hyperparameters against the value of metrics. The best-performing combination based on the highest value of R^2 is then selected for the SVR machine learning model.

Figure 5.5 Mean test scores (R^2) for SVR with different parameter combinations.

A parameter combination with the best mean test score is then selected for the SVR model. Table 5.5 shows a tuned set of hyperparameters for the SVR model.
Table 5.6 Tuned hyperparameters for the SVR model of electrochemical reaction rate constants.

Parameter	Value
Kernel function	RBF
Epsilon (ε)	0.1
Cost (C)	1000
Gamma (γ)	0.01

Two hyperparameters were tuned for the RF model viz, number of trees and maximum depth of trees. A number of tree parameters determine the number of decision trees included in the RF model ensemble. The number of trees increases the model's performance to a specific value (Breiman, 2001). The maximum depth of trees splits each tree to a desired number, capturing the data's information (Osarogiagbon et al., 2021). Considering the significance of these hyperparameters, a range of combinations, as shown in, were studied. Bootstrap is a technique that creates multiple training datasets from a given dataset and then builds individual decision trees (Breiman, 2001). A model selects random observations for bootstrap samples, resulting in randomness and diversity in the model in the training process. A default input of 'True' is used for bootstrap, which results in training each tree on a subset of the dataset. Table 5.6 shows the range of values of hyperparameters studied for the RF model.

Table 5.7 Range of values of hyperparameters studied for RF model.

Parameter	Value
Number of trees	[10,20,30,40,50,60,70,80,90,100,200,300,500,750,1000]
Maximum depth of trees	[5,10, None]

Figure 5.6 shows the metric score variation for different parameter combinations for the RF model. A GridSearchCV expects a utility function instead of a cost function during cross-validation. A utility function is often maximised, whereas a cost function is minimised. Since RMSE is a cost function, it is inverted to a utility function by taking its negative value. Following the grid search, the highest negative value of RMSE is selected to choose the best combination of hyperparameters.

Figure 5.6 Metric score for different combinations of hyperparameters for the RF model.

The best set of hyperparameters for the RF model is shown in Table 5.8.
Table 5.8 Tuned hyperparameters for the RF model of electrochemical reaction rate constants.

Parameter	Value
Number of trees	50
Maximum depth of trees	None

As discussed in the previous chapter, the prior of the Gaussian process is specified, which consists of the prior's mean and covariance. The prior mean shows the expected value of the underlying function and is often viewed as the average value of the function. The prior mean value is set as zero in the current model. In contrast, prior covariance expresses the relationship between different sample points in the design space (Rasmussen and Williams, 2006). Then, the kernel object is passed to specify the prior's covariance and its hyperparameters are optimised while fitting the GPR model. The Matern Kernel, with a length scale of 1.0 and a smoothing parameter of 0.5, provided the lowest RMSE value. The Matern class kernel uses a positive smoothing parameter v, which controls the smoothness of the function. Table 5.8 shows the parameters used for the GPR model.

Table 5.9 Tuned hyperparameters for the GPR model of electrochemical reaction rate constants.

Parameter	Value
Alpha	0.01
Kernel	Matern
Length scale	1.0
v	0.5

5.7 Results and Discussion

A dataset was split into training and test datasets to evaluate the model's performance on each split. 5 different split strategies are used in the current study, 20% test data – 80% training data, 30% test data – 70% training data, 40% test data – 60 % training data, 50% test data – 50% training data. A similar technique has been used to assess the influence of various training-to-testing ratios in some research studies (Nguyen et al., 2021). RMSE value across multiple test-train splits is used to check the performance of each machine-learning model, as shown in Table 5.9. This strategy prevents any biases resulting from selecting an optimal machine learning model based on the lowest RMSE value for a single split. RMSE values are the performance metrics used in regression models which quantify the difference between predicted and actual values. The lower the value of RMSE, the better the performance of the regression model. Table 5.10 shows the R^2 values obtained for different test-train splits, representing the proportion of variance in the dependent variable.

GPR has consistently provided the lowest value of RMSE compared to other machine learning models. As it is evident, the GPR model can also consider uncertainty in the dataset and has also been found to be effective while handling non-linear relationships between electrochemical reaction rate constants and corrosion rate. Similarly, SVR and RF have shown that these models are able to provide predictions reasonably well, too. The performances of models are significantly improved through the tuning of hyperparameters. On the other hand, SVR and RF have shown that these models can provide predictions reasonably well, too. ANN model, however, has shown that it is ineffective for current dataset predictions. The lowest performance of the ANN model can be attributed to the type of architecture used and its high sensitivity to the input data quality (Mrzygłód et al., 2020). ANN requires a large dataset to identify patterns effectively, which could have been a limiting factor considering the size of the dataset used in the current study. It is important to note that for any machine learning model, the performance of the model noticeably depends on the type of dataset used, choice of hyperparameters and pre-processing steps carried out. This has been shown through the various sensitivity studies carried out in the current study. The overall high value of RMSE is due to the fewer sample points for 12 design variable problems, which has become one of the drawbacks of the current SBO approach. A 70% training and 30% test dataset is used for validation purposes.

The results obtained provide valuable insights into the performance of machine learning models to predict electrochemical reaction rate constants. However, several limitations must be acknowledged to ensure an understanding of the study's findings. The overall high RMSE value for the ANN model shows that the ANN model has failed to identify patterns effectively. The lower correlations in values of R^2 across various splits show that certain models find it difficult to capture the underlying relationship in the datasets. Although hyperparameter tuning has shown promising improvement in the performance of models in the current study, further refinement and investigation is needed.

Table 5.10 Performance metrics of machine learning models based on the input of electrochemical reaction rate constants. The percentage in RMSE columns is for the test dataset out of the total dataset.

Models	RMSE			
	20 %	30%	40%	50%
ANN	0.260	0.245	0.272	0.250
SVR	0.100	0.101	0.103	0.100
RF	0.107	0.110	0.112	0.111
GPR	0.100	0.100	0.102	0.100

Table 5.11 Performance metrics of machine learning models based on the input of electrochemical reaction rate constants. The percentage in R^2 columns is for the test dataset out of the total dataset.

Models	R^2			
	20 %	30%	40%	50%
ANN	0.290	0.250	0.260	0.470
SVR	0.721	0.727	0.691	0.724
RF	0.698	0.681	0.637	0.662
GPR	0.738	0.732	0.699	0.726

Figure 5.7 Corrosion rate predictions based on electrochemical reaction rate constants using the ANN model. The horizontal axis represents the actual corrosion rates (CFD-driven mass transfer model), and the vertical axis represents the corrosion rate predictions using machine learning models. The solid line represents x = y.

Figure 5.8 Corrosion rate predictions based on electrochemical reaction rate constants using the SVR model. The horizontal axis represents actual corrosion rates (CFD-driven mass transfer model), and the vertical axis represents the corrosion rate predictions using machine learning models. The solid line represents x = y.

Figure 5.9 Corrosion rate predictions based on electrochemical reaction rate constants using the GPR model. The horizontal axis represents the actual corrosion rates (CFD-driven mass transfer model), and the vertical axis represents the corrosion rate predictions using machine learning models. The solid line represents x = y.

Figure 5.10 Corrosion rate predictions based on electrochemical reaction rate constants using RF model. The horizontal axis represents the actual corrosion rates (CFD-driven mass transfer model), and the vertical axis represents the corrosion rate predictions using machine learning models. The solid line represents x = y.

5.7.1 Surrogate Modelling based Optimisation

Based on the performance metrics, the surrogate model developed using the GPR model determines the optimal electrochemical reaction rate constants. The electrochemical reaction rate constants in Table 5.3 are used as the design variables, and the formulation of an objective function (f) is shown below,

$$\text{minimise} \quad f = CR_{expt} - CR\,(E_{rev,H}, b_H, \Delta H, a_{1,H}, i_{0ref,H}, i_{0ref,H_2CO_3}, a_{1,H_2CO_3}, \Delta H_2CO_3, C_{CO_2,ref}, E_{rev,Fe}, \Delta H, b_{Fe})$$

$$\text{subject to:} \quad -0.279 < E_{rev,H}\,(\text{V}) < -0.186$$
$$0.101 < b_H(\text{V}) < 0.131$$
$$27000 < \Delta H(\text{J/mol}) < 33000$$
$$0.4 < a_{1,\text{H}} < 0.6$$
$$0.04 < i_{0ref,H}(\text{A/m}^2) < 0.06$$
$$0.048 < i_{0ref,H_2CO_3}(\text{A/m}^2) < 0.072$$
$$-0.6 < a_{1,\text{H}_2\text{CO}_3} < -0.4$$
$$40000 < \Delta H_2CO_3(\text{J/mol}) < 60000$$
$$0.0292 < C_{CO_2,ref}(\text{mol/L}) < 0.0439$$
$$-0.51 < E_{rev,Fe}(\text{V}) < -0.39$$
$$30000 < \Delta H(\text{J/mol}) < 45000$$
$$0.03 < b_{Fe}(\text{V}) < 0.08$$

where CR_{expt} is an experimental value of corrosion rate in mm/yr (Nesic et al., 1995).

The minimum value of the objective function is obtained using three different optimisation methods viz., Nelder-Mead simplex method, Powell's method and Hooke-Jeeves method for optimisation. Table 5.12 shows the optimal values obtained using different optimisation methods.

Table 5.12 Optimal values of electrochemical reaction rate constants obtained using the SBO model.

Parameter	Previous Value	Optimal values using Powell's method	Optimal values using the Nelder-Mead Simplex method	Optimal values using Hooke-Jeeves method
$E_{rev,H}$ (V)	-0.232	-0.186	-0.187	-0.189
b_H (V)	0.116	0.117	0.116	0.102
ΔH (J/mol)	30000	27005	27077	28623
$a_{1,H}$	0.5	0.4	0.453	0.406
$i_{0ref,H}$ (A/m^2)	0.05	0.060	0.060	0.059
i_{0ref,H_2CO_3} (A/m^2)	0.06	0.059	0.049	0.053
a_{1,H_2CO_3}	-0.5	-0.547	-0.521	-0.552
ΔH_2CO_3 (J/mol)	50000	46022	41740	42399
$C_{CO_2 ref}$ (mol/L)	0.0366	0.035	0.034	0.030
$E_{rev,Fe}$ (V)	-0.488	-0.509	-0.496	-0.505
ΔH_{Fe} (J/mol)	37500	41802	41979	33460
b_{Fe} (V)	0.0387	0.037	0.042	0.032

5.7.2 Corrosion Rate Predictions and Comparison

Figure 5.11 shows the comparison between the corrosion rate (CR) predictions for pH 4 obtained using electrochemical reaction rate constants available in the literature with CFD-driven mass transfer model, empirical VSL and turbulent diffusivity profile with the same electrochemical reaction rate constants, CFD-driven mass transfer model with surrogate based optimal parameters obtained using Nelder-Mead Simplex method, Powell's method and Hooke-Jeeves method, and experimental corrosion rate predictions obtained from Nesic et al. (1995). It can be seen that the CFD-driven mass transfer model with electrochemical reaction rate constants in the literature underpredicted the corrosion rate values, whereas the same electrochemical reaction rate constants with empirical VSL and Dt over predicted corrosion rates (Thorat et al.,

2024). The optimal values obtained using surrogate-based optimisation shown in Figure 5.11 predicted CR values quite well, as shown below.

Figure 5.11 Corrosion rate predictions and comparison between CFD-driven mass transfer model with literature constants, mass transfer model with empirical viscous sublayer and turbulent diffusivity (Empirical VSL and Dt), CFD-driven mass transfer model with surrogate-based optimal parameters obtained using Nelder-Mead Simplex method, Powell method, and Hooke-Jeeves method and experimental corrosion rate predictions shown using solid points and error bars (Nesic et al., 1995). Powell's method has provided the lowest RMSE value upon comparison with the experimental dataset.

CFD driven-mass transfer model with literature constants of electrochemical reaction rates underpredicted the corrosion rate values with the RSME value of 0.66. In contrast, the prediction model based on empirical VSL and Dt over-predicted with an error of 0.47 upon comparison with the experimental data of Nesic et al. (1995). On the other hand, CFD driven mass transfer model with optimal values of electrochemical reaction rate constants using the Nelder-Mead simplex method, Powell's method, and Hooke Jeeves method predicted corrosion rates extremely well with the RMSE of 0.30, 0.28, and 0.29 respectively. The optimisation method with the lowest RMSE value, Powell's method, is used for the corrosion predictions in the subsequent study. As described in Chapter 3, the error bars for pH4 in experimental datasets are the largest compared to other pH values, and the corrosion rates obtained using optimal electrochemical reaction rate constants fall within these error bars.

Machine learning-based surrogate models are significant in finding optimal electrochemical reaction rate constants in the current model. These surrogate models have primary advantages, such as the ability to approximate complex mathematical equations and computational efficiency (Yondo et al., 2018). The methodology developed in this study can be applied to several other mathematical models to obtain optimal parameters.

5.8   Summary and Conclusions

 In this research work, a comprehensive study of corrosion rate prediction and optimisation of electrochemical reaction rate constants is carried out using a combination of data-driven modelling approach and optimisation techniques. Initially, upper and lower limits were defined for the input variables from the literature and on the basis of experimental bias. Following this, the RLHS sampling technique is used to obtain random distribution samples. PCA on a dataset consisting of 12 electrochemical reaction rate constants and 1 output variable – corrosion rate obtained using methodology developed in Chapter 3. The PCA revealed that all the PCs contribute almost equally towards the explained variance of the dataset; hence, all the PCs were retained for the current model.

Different training-testing split ratios were studied for RMSE values for all machine learning models, out of which GPR showed the most consistent performance with the lowest RMSE value. A 70-30% training-testing split ratio was then selected to construct a surrogate model integrated with different optimisation techniques to find an optimal one based on RMSE values. Employing Powell's method for minimisation of the difference between CFD-driven mass transfer model predictions and experimental corrosion rate predictions, a set of optimal electrochemical reaction constants is obtained. Upon comparison of the corrosion rates with the newly derived electrochemical reaction constants with results from Chapter 3, it is found that the RMSE value has decreased from 0.66 to 0.28, showing the robustness of the technique proposed in this research work.

To conclude, this study has demonstrated the effectiveness of the SBO approach to obtain an optimal set of electrochemical reaction rate constants. The use of PCA for dimensionality reduction confirmed the importance of each principal component. The surrogate model constructed using GPR was integrated with an optimisation technique to find the optimal set of electrochemical reaction rate constants, resulting in the enhancement of the reliability of corrosion rate predictions.

Chapter 6: CFD-driven Mass Transfer Modelling of CO_2 Corrosion in Complex Flows

6.1 Introduction

A novel methodology is developed to predict the CO_2 corrosion using CFD driven mass transfer model in Chapter 3. This model was applied to the fully developed flow condition of the straight pipeline. However, when it comes to situations where complex flow occurs, there are not many CO_2 corrosion prediction tools that are available in the literature. Hence, with the advantage of CFD driven mass transfer model, CO_2 corrosion rate predictions are obtained in complex flow situations. The model developed in Chapter 3 with optimal electrochemical reaction rate constants obtained in Chapter 5 is applied to the complex flow situation of a gradual constriction.

The models that are available in the literature for CO_2 corrosion predictions are validated under certain operating conditions. This phenomenon is well established for corrosion prediction in fully developed single-phase pipeline flow (Kahyarian et al., 2017; Thorat et al., 2024). Available predictive tools in the literature are valid for fully developed flow and uniform of CO_2 corrosion (Nordsveen et al., 2003; Prasad et al., 2018; Hu et al., 2018; Nešić et al., 2019; Kahyarian and Nesic, 2020). The corrosion rate predictions obtained in Chapter 3 are for steady state, fully developed single-phase flow. These corrosion predictions lack reliability due to their limitations when considering real-life flow situations. In addition, much less is known about the corrosion predictions when complex flow situations occur, such as flows in expansions/constrictions domains. There is a clear research gap in the literature when it comes to the modelling of CO_2 corrosion in complex geometries.

In oil and gas industry pipelines, the connection between tubing or tapered sections is gradual constrictions (Zhong et al., 2020). These connections are not sudden and always have a transfer region that connects large-diameter pipe to small-diameter pipe. Zhong et al. (2020) used in situ electrochemical measurements to study flow accelerated corrosion (FAC) in gradual constriction pipe. They combined the array electrode technique with CFD to predict and mitigate corrosion rate across the surface. The combination of CFD and array electrode technique was helpful in correlating local corrosion rate with local hydrodynamics obtained from CFD. A similar approach was used in the study of Zeng et al. (2023), in which FAC of X65 gradual contraction pipe was studied for high pCO_2 conditions. They reported significantly high corrosion rate values; hence, verification with another dataset was needed. There are very few studies that predict corrosion rate in gradual contraction pipes (Zhong et al., 2020). The available studies use CFD to observe local hydrodynamics and correlate it with the local corrosion predictions (Zhong et al., 2020; Zeng et al., 2023).

Hence, a methodology is developed in the current study that predicts the VSL and turbulent diffusivity profile along the length of an expansion or constriction type domain. These predictions are subsequently coupled to the mass transfer model to obtain corrosion rates that will account for the variation in VSL due to changes in the flow field as the geometry changes. The following section discusses materials and methods used in the current study involving mass transfer modelling and CFD methodology. Following this, results and a discussion on predicting turbulent boundary

layer and corrosion rates are presented. The following steps are carried out in the current study,

- The velocity field calculation for expansions/constrictions pipe along the length of the domain uses a 2D CFD simulation.
- Extraction of VSL and D_t profile along the length of the domain and its subsequent coupling to the mass transfer model.
- CO_2 corrosion rate predictions using a novel mass transfer model in expansions/constrictions pipe in which flow field obtained from 2D CFD model is coupled with a series of 1D mass transfer models with newly derived electrochemical reaction rate constants.

6.2 Materials and methods

6.2.1 Velocity Field Calculation

As described in Chapter 3, a methodology must be developed that accurately calculates the flow field near the wall along the length of the domain. By considering this, a methodology is developed to calculate VSL thickness (δ) and D_t profile. This is done using MATLAB software. The process starts with evaluating coordinates and friction velocity at the bottom wall of the domain. Table 6.1 provides the variables obtained from the 'mpheval' function of MATLAB software, which evaluates expressions on mesh nodes to calculate the velocity field.

Table 6.1 List variables evaluated at the wall using the 'mpheval' function.

Variable	Description
x	x domain coordinate
y	y domain coordinate
nx	Normal x wall component
ny	Normal y wall component
tx	Tangential x wall component
ty	Tangential y wall component

Euclidean distance r_0 is calculated for each boundary point using evaluated 'x' and 'y' coordinates. These values are sorted and rearranged in ascending order for the following steps. An arbitrary projection length of $1.5\ e^{-4}$ m is then used to determine the reference length along which all variables are evaluated. These variables include velocity components, dimensionless velocity (u^+) and dimensionless wall distance (y^+) to find a value of δ. The value of δ is determined by evaluating the slope (m) of u^+ vs y^+ and then finding the value where m exceeds the tolerance value of $1e^{-2}$ given below in Eq. (6.1),

$$m_{(i,j)} = \frac{u^+_{(i,j+1)} - u^+_{(i,j)}}{y^+_{(i,j+1)} - y^+_{(i,j)}} = \frac{\partial u^+}{\partial y^+} \qquad (6.1)$$

where i and j are the boundary point and slope evaluation index.

The tangential velocity (ut) is calculated based on tx and ty as expression shown in Eq. (6.2),

$$ut(i,:) = abs\big(u(i,:) * tx(i) + v(i,:) * ty(i)\big) = |\boldsymbol{u} \cdot \boldsymbol{t}| \qquad (6.2)$$

where *u* and *v* are the tangential and normal components of velocity.

Figure 6.1 shows normal and tangential components evaluated along the length of the domain. The values of *tx* and *ty* change at the inclined region, indicating variations in the direction of the velocity vector along the boundary. These changes in the direction of *tx* and *ty* along the boundary of the domain will influence the velocity and, consequently, the point where the deviation in velocity occurs for the calculation of delta.

Figure 6.1 Graphical representation of the normal and tangential components evaluated along the length of the domain.

Once the value of δ is determined along the length of the pipe, the velocity field is calculated between the wall and δ. Figure 6.2 shows the flow chart for the calculation of δ and D_t.

Figure 6.2 Flow chart for calculating velocity field required in the mass transfer modelling.

6.3 Constriction – Expansion Model

The velocity field obtained from the above method is coupled with the mass transfer model for the constriction-expansion pipe. The coupling is done so that VSL conditions are obtained using the CFD model, and then, for each VSL condition, the mass transfer model is solved. This section describes mass transfer modelling in gradual constriction pipe with details of the computational domain, initial and boundary conditions,

124

numerical method, and mesh convergence studies for both mass transfer and CFD model.

6.3.1 1D Model of Mass Transfer of Expansion/Constriction Model

As discussed in Chapter 3, a 1D numerical model is developed to simulate CO_2 corrosion in pipelines. This 1D mass transfer model is based on the finite volume approach in which the fluxes of species are calculated at the boundaries. In contrast, the concentration of species is calculated at the centre of finite volumes. The computational length of the mass transfer model is based on the VSL thickness obtained from the CFD model, as shown in Figure 3.3. The governing equation for species concentration is solved using the Backward Euler method, as shown in Figure 3.4. The total number of finite volumes used is 50, and the program is implemented in Python. The optimal electrochemical reaction rate constants obtained in Chapter 5 are used in the current model.

Initial and Boundary Conditions

A uniform species concentration in chemical equilibrium is considered for initial and boundary conditions for the mass transfer model. It is assumed that all species are thoroughly mixed by turbulence. For the species not involved in the electrochemical reactions at the metal surface zero flux ($N_j = 0$) is specified. For the species j involved in the electrochemical reactions at the metal surface, the flux is specified by Eq. (6.3),

$$N_j = -\frac{i_j}{n_j F} \tag{6.3}$$

where i_j is the partial current of a specific electrochemical reaction which contributes to the overall current density, n_j is the number of moles of electrons exchanged per mole of species j, and F is the Faraday constant (96485 C/mol).

6.3.2 CFD Modelling of Gradual Constriction Pipe

This section gives an overview of steps carried out in CFD modelling of the gradual constriction model, including computational domain, initial and boundary conditions and mesh sensitivity study.

Geometry

Figure 6.3 shows the computational domain used for the current study. RANS formulation is used to compute the flow field required for the 2D model. The larger and smaller pipe diameters are 25.4 mm and 12.7 mm, respectively, with a cone angle of 18°. The lengths of expansion, gradual constriction, and straight constriction regions are 127 mm, 20 mm, and 127 mm, respectively.

Figure 6.3 Expansion/constriction pipe computational domain used for the simulations.

Initial and Boundary Conditions

The velocity inlet and pressure outlet boundary conditions are used for the simulations. A fully developed flow boundary condition is chosen at the inlet with a fluid density of $\rho = 998.8 \text{ kg/m}^3$ and dynamic viscosity μ = 0.001 Pa · s.

Table 6.2 Boundary conditions used for the simulation of gradual constriction geometry.

Boundary	Description	p	u	k	ω
AB	Inlet	$\mathbf{n} \cdot \nabla p = 0$	$\boldsymbol{u} = \boldsymbol{u}_{in}$	$k = k_{in}$	$\omega = \omega_{in}$
CD	Outlet	$p = 0$	$\mathbf{n} \cdot \nabla u = 0$	$\mathbf{n} \cdot \nabla k = 0$	$\mathbf{n} \cdot \nabla \omega = 0$
AD	Symmetry axis	$\mathbf{n} \cdot \nabla p = 0$	$\mathbf{n} \cdot \nabla u = 0$	$\mathbf{n} \cdot \nabla k = 0$	$\mathbf{n} \cdot \nabla \omega = 0$
BC	Pipe wall	$\mathbf{n} \cdot \nabla p = 0$	$u = 0$	$k = 0$	$\omega = 0$

Mesh Sensitivity Analysis of Gradual Constriction Pipe

A structured quadrilateral mesh with an extremely fine grid ratio at the wall is used for the simulation. The value of $y^+ \ll 1$ is chosen to ensure that the flow field in the viscous sublayer is calculated accurately. Following this, several mesh configurations were studied to evaluate mesh elements' effect on the outlet's flow velocity. This is shown in Figure 6.4, and the average velocity did not change significantly for the number of elements greater than 577500. Hence, a mesh configuration with 577500 elements is chosen for the simulations.

Figure 6.4 Mesh convergence study of expansion/constriction domain in the current study.

6.3.3 Expansion/Constriction Domain based on Experimental Setup in Literature

This section provides details related to the expansion/constriction domain used in Zhong et al. (2020) to study FAC in N80 steel gradual contraction pipe. Their study was focused on the prediction of flow-accelerated corrosion and its mitigation with the help of in situ electrochemical measurements. There are not enough studies in the literature related to the prediction of corrosion rate in the gradual constriction domain. Those that are available have provided very little information about the experimental setup, making it difficult to replicate the model (Zhong et al., 2020; Zeng et al., 2023). Hence, certain assumptions are made in the CFD-driven mass transfer model to carry out verification with the experimental dataset of Zhong et al. (2020).

Geometry

Figure 6.5 Expansion/constriction pipe computational domain used in the current model for verification purposes.

Figure 6.5 shows the expansion/constriction pipe for the verification of corrosion rates in the N80 steel gradual contraction pipe. It is important to mention that the length of

127

the domain was missing from the research study of Zhong et al. (2020). The experimental domain is shown in Figure 6.6. Hence, based on the available dimensions, the scaling method is used to assume the remaining dimensions. The diameter of expansion regions is 50 mm, the cone angle is 21.8°, and the constriction region diameter is 42 mm. The lengths of all the regions are assumed in the current model. The total length of this domain is 242 mm.

(a)

(b)

Figure 6.6 Schematic of expansion/constriction domain used in the experimental study of Zhong et al. (2020).

Initial and Boundary Conditions

The velocity inlet and pressure outlet boundary conditions are used for the simulations. A velocity of 2m/s is specified at the inlet with a fluid density of $\rho = 998.8 \text{ kg/m}^3$ and dynamic viscosity $\mu = 0.001 \text{ Pa} \cdot \text{s}$. Table 6.3 shows the boundary conditions used for the verification model.

The lack of significant details related to the setup used, such as length of expansion and constriction region, makes it difficult to replicate the results. Nevertheless, an attempt is made here to obtain corrosion rates from the details given in the study by Zhong et al. (2020).

Table 6.3 Boundary conditions used for the simulation of gradual constriction geometry.

Boundary	Description	p	$\boldsymbol{u}$	k	ω
AB	Inlet	$\mathbf{n}\cdot\nabla p = 0$	$\boldsymbol{u} = \boldsymbol{u}_{in}$	$k = k_{in}$	$\omega = \omega_{in}$
CD	Outlet	$p = 0$	$\mathbf{n}\cdot\nabla\boldsymbol{u} = 0$	$\mathbf{n}\cdot\nabla k = 0$	$\mathbf{n}\cdot\nabla\omega = 0$
AD	Symmetry axis	$\mathbf{n}\cdot\nabla p = 0$	$\mathbf{n}\cdot\nabla\boldsymbol{u} = 0$	$\mathbf{n}\cdot\nabla k = 0$	$\mathbf{n}\cdot\nabla\omega = 0$
BC	Pipe wall	$\mathbf{n}\cdot\nabla p = 0$	$\boldsymbol{u} = 0$	$k = 0$	$\omega = 0$

Mesh Sensitivity Analysis of Expansion/Constriction Pipe

The mesh sensitivity analysis of the model is shown in Figure 6.7. A total of 5 mesh configurations were studied to find out the variation in average velocity at the outlet. The mesh elements were varied along the length and height of the domain. The $y^+ \ll 1$ was assumed to completely evaluate VSL conditions. It can be seen that there is no significant variation in the average velocity after mesh configuration with 225000 elements compared to the previous configurations. Hence, the mesh configuration with 225000 elements is chosen for the simulations to ensure the reliability of the CFD model.

Figure 6.7 Mesh convergence study of expansion/constriction domain used for the verification purpose.

6.4   Results and Discussion

6.4.1   Verification of Corrosion Rate Predictions with Experimental Data

The corrosion rate predictions obtained from CFD driven mass transfer model are compared with the experimental data of Zhong et al. (2020). The experimental setup includes in situ electrochemical measurements using the array electrode technique (Zhong et al., 2020). The pH value of 6.18 was obtained after de-aeration of the solution,

and the temperature of the solution is 65℃. The information related to pCO_2 is not mentioned in their study, and hence, its value is assumed to be 1.0132 bar.

Figure 6.8 shows the comparison between experimental (Zhong et al., 2020) and current model corrosion rate predictions. The corrosion rates are presented for different locations along the length of the domain at 2 m/s. The location of electrodes used in the experimental study is also shown in Figure 6.8. The difference between experimental corrosion rates and predicted corrosion rates with the current model is significant at locations 1, 2 and 3, which then reduces at locations 4 to 8 and again becomes significant at locations 9 and 10. Locations 1, 2 and 3 are near the entrance of the gradual constriction region, whereas locations 9 and 10 are just after the gradual constriction region. However, corrosion rate predictions show that the current model follows the trend of experimental data. For instance, the highest corrosion rate value is found at location 4. This can be attributed to increased velocity due to changes in the diameter of the domain. The corrosion rate decreases along the length following location 6 in both cases.

It is important to address the differences between the two datasets. The experimental data (Zhong et al., 2020) obtained from the literature posed several challenges due to missing pCO_2 values and incomplete information regarding the dimension, in particular, the lengths of both the expansion and constriction regions. These limitations, belonging to the reporting of the data and experimental setup, impact the reliability of the comparisons by introducing uncertainties. Furthermore, the experimental setup lacks details related to the location of the electrodes along the domain to extract corrosion rate predictions, adding more complexity to the verification of the model. These challenges have been acknowledged in the current study. Despite these limitations, the current model has shown more favourable agreement with the experimental predictions in the constriction region. This has highlighted its ability to capture certain aspects of the observed trends in the experimental data.

Figure 6.8 Comparison of corrosion rate predictions between experimental dataset of Zhong et al. (2020) and the current model at pH = 6.18, pCO2 = 1.0132 bar, Temperature = 65℃ and velocity = 2 m/s. The location of the electrodes is shown on the right side of the graph obtained from the experimental setup of Zhong et al. (2020).

6.4.2 Prediction of Turbulent Boundary-layer Profiles

The CFD model developed in this study calculated VSL and D_t profile near the wall along the length of the expansion/constriction pipe. A fully developed velocity profile of 1 m/s to 5 m/s is specified at the inlet of the computational domain. Figure 6.10 shows the velocity profile for expansion/constriction geometry. It can be seen that the flow velocity increases as the flow approaches the gradual constriction part, and the maximum velocity is seen in the straight constriction part. This maintains the principle of continuity.

Figure 6.10 shows the prediction of VSL for 1 m/s to 5 m/s at the gradual constriction region. The VSL values decrease as the inlet velocity increases, as found in Chapter 3. The VSL values are the highest for each velocity at the location just near the start of the gradual constriction region. The VSL values take a significant dip in the gradual constriction region and are then found to increase in the straight constriction region. It is also found that the VSL values for all velocities are almost equal in the gradual constriction region, highlighting the reduced laminar region near the wall.

Figure 6.9 Velocity profile in expansion/constriction geometry for velocity 1 m/s at inlet.

Figure 6.10 Viscous sublayer thickness prediction in expansion/constriction geometry for velocity 1 m/s to 5 m/s.

6.4.3   Corrosion Rate Predictions

As discussed earlier, the accurate prediction of corrosion rates in complex flow situations is of paramount importance to ensure the integrity of the pipeline. The CFD-driven mass transfer model developed in this work is used to predict corrosion rate at pH 4 to 6, pCO_2 1 bar, Temperature of 20°C, and velocity of 1 m/s to 5 m/s. Figure 6.11, Figure 6.12, and Figure 6.13 show the prediction of corrosion rate for all scenarios with a notable trend of increase in the corrosion rate in the direction of flow, with the highest values obtained in the region of gradual constriction. This ascent in corrosion rate is indicative of the dynamic nature of the corrosion process in which the flow field plays a significant role, as shown in Figure 6.10 (Zeng et al., 2023).

A particularly interesting observation of these predictions is a sudden increase in corrosion rate at the entrance of the inclination region. This phenomenon can be attributed to the increased turbulence due to abrupt changes in the pipe diameter in all cases. This change in turbulence enhances the transport of species to and from the bulk, which leads to an increased mass transfer rate (Li and Woollam, 2012). The enhanced mass transfer rate results in a higher corrosion rate value in the constriction region. Zhong et al. (2020) found that an increase in flow velocity and wall shear stress in the gradual constriction region increased the corrosion rate and observed a similar trend to that found in the current model predictions. As the pH value increased, the corrosion rate decreased, whereas as the velocity increased, the corrosion rate increased for constant pCO_2 and temperature similar to the observations made in Chapter 3.

Figure 6.11 CO_2 corrosion rate predictions for pH = 4, temperature = 20°C and pCO_2 = 1 bar in the expansion/constriction domain.

Figure 6.12 CO_2 corrosion rate predictions for pH = 5, temperature = 20°C and pCO_2 = 1 bar in expansion/constriction domain.

Figure 6.13 CO₂ corrosion rate predictions for pH = 6, temperature = 20°C and pCO₂ = 1 bar in the expansion/constriction domain.

Figure 6.14 shows the variation of corrosion rate with temperature at pH 4, pCO_2 = 1 bar and velocity 5 m/s. The operating temperature accelerates the physicochemical processes involved in corrosion (Nordsveen et al., 2003). Hence, as the temperature increases, the corrosion rate increases for a given pH value, as shown. Figure 6.15 shows the corrosion rate predictions for different pCO_2 values at temperature 20°C, and velocity 5 m/s. As the pCO_2 increases concentration of H_2CO_3 species increases, resulting in an increase in the rate of catholic reactions (Chen et al.,2023). Hence, as pCO_2 increases corrosion rate increases.

This shows the robustness of the current model in predicting the corrosion rate in complex flow situations. The current model can be applied to different situations, such as transient flows and flows in elbows/bends. The ability of the current model to accurately calculate VSL conditions and its subsequent coupling with the mass transfer model provides a pipeline designer with the ability to study a wide range of conditions.

Figure 6.14 CO_2 corrosion rate predictions for different temperatures at pH = 4, velocity = 5 m/s, and pCO_2 = 1 bar in the expansion/constriction domain

Figure 6.15 CO_2 corrosion rate predictions for different pCO_2 values at pH = 4, temperature = 20℃, and velocity = 5 m/s in the expansion/constriction domain.

6.5 Summary and Conclusions

A CFD-driven mass transfer model for the prediction of CO_2 corrosion in complex flow situations is presented in this chapter. The empirical correlations available in the literature for predicting VSL and turbulent diffusivity profiles are valid for fully developed flow conditions, restricting their applicability in developing and disturbed flow conditions. The CFD model developed in this study calculates VSL and turbulent diffusivity profile at any location in the domain. This model has also shown that it can predict viscous sublayer conditions in the domain with geometrical variations along the length.

The VSL conditions obtained at each location along the domain length are coupled with the mass transfer model to predict corrosion rates. The predictions obtained have shown that the highest corrosion rate values are experienced near the start of the straight constriction region after the inclination. The corrosion rate stayed at a moderate range following the inclination region.

While challenges in the experimental data, such as missing information related to the value of pCO_2 and incomplete dimensions introduced uncertainties in the verification of the current model. Our model has demonstrated favourable agreement in the gradual and straight constriction regions. Acknowledging these constraints, future research should focus on acquiring thorough experimental datasets as well as enhancing reporting standards.

To summarise, the current approach has shown robustness in predicting VSL, turbulent diffusivity profiles and corrosion rate predictions in complex flow situations.

Chapter 7: Discussion and conclusion of Computational Fluid Dynamics Driven Mass Transfer Modelling of CO_2 Corrosion in Pipelines and Machine Learning Modelling Based Optimisation

This chapter discusses key findings from the results and the methodology developed for the prediction of CO_2 corrosion in complex flow geometries. In addition, the significance of machine learning modelling in corrosion decisions is explored in this chapter. The accurate prediction of viscous sublayer conditions and their subsequent coupling to mass transfer modelling is also discussed in this chapter.

7.1 Introduction

A CFD-driven mass transfer modelling methodology is developed to understand the importance of accurately predicting the VSL conditions in complex flow situations. These predictions are then coupled with the mass transfer model to predict CO_2 corrosion in complex flow situations. This novel methodology is described in detail in Chapter 3, validated with the 1-D mass transfer model only applicable for fully developed flow conditions. With the aid of this methodology, optimal operating parameters to mitigate corrosion in pipelines are obtained using machine learning enabled surrogate modelling shown in Chapter 4. Machine learning models have been found helpful in predictive maintenance (Yang et al., 2020), corrosion rate estimation (Abbas, 2016), risk assessment (Senouci et al., 2014) and optimisation of corrosion inhibitors (Aghaaminiha et al., 2021). Furthermore, machine learning models are beneficial for tuning parameters for mathematical models (Li et al., 2021). A sensitivity analysis of electrochemical reaction rate constants is carried out with the help of machine learning enabled surrogate modelling. An optimisation methodology is then applied to find out the optimal set of electrochemical reaction constants provided in Chapter 5. A novel CFD-driven mass transfer modelling methodology developed in Chapter 3 is then expanded to complex flow situations with accurate mathematical modelling for the prediction of CO_2 corrosion.

7.2 Flow-induced CO₂ Corrosion Mechanism

7.2.1 Use of CFD Methodology to Build a Robust Mass Transfer Model for Prediction of CO₂ corrosion in Straight Pipe

Following the identified research gaps in the literature, a CFD methodology was initially built to calculate VSL conditions accurately in straight pipelines. RANS formulations, along with the SST $k - \omega$ model, compute the flow field required for the computational domain. The values of VSL thickness obtained from CFD are then compared with the empirical correlation obtained by (Davies, 1972). The error between the CFD obtained VSL thickness and that obtained by empirical correlation is between 11.51% and 17.50% for Reynolds number between 18316 and 171795. The discrepancy between these values can be explained by the way the empirical correlation of VSL was created. This equation is obtained by curve fitting using the friction factor by Blasius for a specific range of Reynolds numbers from 3000 to 100,000. In addition to this, the equation is simplified by considering the $y^+ = 5$ at the edge of VSL. However, this assumption of Davies (1972) disputes with Popovich and Hummel (1967) as they had found the average thickness of VSL equal to $y^+ = 6.2$, and the most probable value of VSL at $y^+ = 4.3$.

This argument of Popovich and Hummel (1967) related to varying VSL thickness based on y^+ the findings of the current study support value instead of relying on simplified empirical correlation. Table 7.1 shows the values of at y^+ at the edge of VSL obtained for straight pipelines.

Table 7.1 Values of at y^+ at the edge of VSL obtained for straight pipelines using CFD.

Reynolds number	y^+ at the edge of VSL
18316	4.92
35686	5.03
52899	5.09
70019	5.04
87053	5.00
104138	4.96
121072	4.94
138079	4.90
155123	5.02
171795	5.05

As discussed above, to account for the turbulent flow field, a notion of turbulent diffusivity is used. The comparison between the predicted turbulent diffusivity profile and empirical turbulent diffusivity has shown that the overall error is less than 10%. Following the validation of the viscous sublayer conditions obtained from CFD, corrosion rates are predicted for the operating conditions in which film formation did not take place and effects of electro-migration can be ignored. Corrosion rates are obtained for diameter of 0.015 m, pH 4 to 6, velocity 1 to 10 m/s, temperature 20 °C and pCO_2 1 bar. These predictions are compared with the numerical model of Srinivasan (2015) and the experimental data of Nesic et al. (1995). Table 7.2 shows the RMSE for corrosion rates obtained using the current model, the numerical model of Srinivasan (2015) and the experimental data of Nesic et al. (1995) for pH 4, 5 and 6.

Table 7.2 Comparison between the current model, numerical model of Srinivasan (2015) and experimental data (Nesic et al., 1995) of CO_2 corrosion.

pH	RMSE	
	Current model	(Srinivasan, 2015)
4	0.66	0.52
5	0.08	0.16
6	0.15	0.38

For pH 4, the experimental data has the largest error bars compared to other pH values, as described in Chapter 3. As can be seen from Table 7.2, pH 4 has the highest deviation from the experimental corrosion rates as compared to pH 5 and pH 6. This can be related to the choice of electrochemical reaction constants used in the current model, which needed a sensitivity analysis to find optimal values of these constants, as described in Chapter 5. For pH 5 and pH 6, the current model is in much better agreement than the numerical model of Srinivasan (2015). Following this, machine learning based surrogate modelling is used to find the optimal set of operating parameters in pipelines discussed in the following sections.

7.3 Importance of Machine Learning Modelling in Corrosion Modelling

7.3.1 Machine Learning Modelling Enabled Surrogate Modelling to Check the Influence of Operating Parameters

In the current research work, machine learning models such as ANN, SVR, GPR, and RF are used to predict CO_2 corrosion in pipelines based on input parameters such as pH, temperature, velocity, and pCO_2. The methodology started with defining the range of input parameters, followed by the use of OLHS for sampling of input parameters. Corrosion rate is then predicted for these sampling points using a robust model developed in Chapter 3. PCA technique is then applied to check if the dimensionality of the model can be reduced. Hyperparameters of all the machine learning models are then tuned to predict corrosion rates. The discussion covers the model's performance, the importance of input features used, implications of the optimisation techniques used, and potential areas for future research.

The results obtained from PCA are discussed here to give an overall idea of the dataset's structure and dimensionality reduction. The explained variance obtained for each principal component (PC) showed the fraction of the total variance accounted for by each component. PC1, PC2, PC3, and PC4 each contributed 24% to 26% of the variance, suggesting that the input features are evenly spread across the PCs. The cumulative variance shows how successive PCs add to the cumulative variance explained in the PCA analysis (Jollife & Cadima, 2016). In the current study, the cumulative approach enabled us to make informed decisions related to retaining all the PCs for the machine learning model. This approach is crucial as some other research work (Abbas, 2016) found that

temperature and pCO_2 input features had the most significant contribution to the variance in the datasets. To summarise, PCA results provided valuable information for the selection of features and dimensionality reduction of the dataset.

The performance evaluation metrics used for machine learning models are MSE, RMSE, and R^2. These metrics provided valuable insights into the accuracy and determined the goodness of the fit of machine learning models. ANN and GPR models outperformed the SVR and RF models for all the metrics evaluated. The lower values of MSE and RMSE for ANN and GPR models showed that these models are capable of minimising prediction errors, showing their high accuracy in corrosion rate predictions. In addition, high R^2 values obtained for ANN and GPR around 0.97 indicated their capability of capturing a significant portion of the variance in the data. In contrast to ANN and GPR, RF and SVR models performed reasonably well but showed higher values of MSE and RMSE. The R^2 values for both RF and SVR fall below those of ANN and GPR, raising their suitability for this corrosion prediction modelling. These results highlight the use of performance evaluation metrics for the selection of the best machine-learning model for a given corrosion prediction task. ANN, with its performance, is then chosen to find the optimal operating conditions.

To summarise, this work has shown how machine learning models perform when applied to corrosion scenarios. Machine learning models integrated with optimisation algorithms have the potential to save costs and improve the safety and operational efficiency of pipeline networks. Furthermore, these findings will aid in the construction of operational guidelines to be applied in real-world scenarios with some modifications. For example, the inclusion of time series analysis will look into the corrosion inhibition performance of some other research work (Aghaaminiha et al., 2021). This study has shown that it is essential to find out the best machine learning model as the performance of each model will vary depending on the type of datasets used.

7.4 Tuning of the Mathematical Model for the Prediction of CO₂ Corrosion

Machine learning models play a significant role in the optimisation of parameters in mathematical models for various industrial applications such as proton exchange membrane fuel cells (PEMFC) (Li et al., 2021; Fan et al., 2022) and PSA (Subraveti et al., 2019) and granulation model (Braumann et al., 2010). Considering this, different machine learning models such as ANN, SVR, RF and GPR are used to construct surrogate models. The main advantage of surrogate models is the approximate mapping of the complex non-linear relationship between input parameters and the objective functions, as discussed in Chapter 4 and above section. A 12-design variable problem with 1 objective function of a difference between experimental corrosion rate and CFD-driven mass transfer model described in Chapter 3 is modelled. RLHS technique is used for sampling, and PCA is applied to check if the dimensionality of the model can be reduced. Surrogate models obtained using the best-performing machine learning model obviate the necessity of running the computationally expensive corrosion model for objective function evaluation. Integrated optimisation algorithms allow surrogate models to search through the design space to find the optimal set of parameters, in this case, a set of electrochemical reaction rate constants.

A PCA technique was used to simplify the complexity of the dataset; however, upon analysing the explained variance of each PC, it was found that all the PCs contributed

equally to the total variance in the dataset. The Explained variance values of all PCs ranged between 6.9% and 9.8%. This provided essential information that suggested retaining all the PCs during machine modelling.

The various train-test split ratios were studied for RMSE value for ANN, SVR, RF and GPR models. It was found that the RMSE value for GPR was the lowest and consistent for all train-test split ratios. Finally, a 70-30% split ratio was selected to construct a surrogate model using the GPR model. This enabled the optimisation of critical electrochemical reaction rate constants with the help of Powell's method. The optimal values obtained from this method have provided robustness to the corrosion prediction model described in Chapter 3. Table 7.3 shows the comparison of corrosion predictions developed in Chapter 3, a numerical model of Srinivasan (2015) and CFD driven mass transfer model using optimal electrochemical reaction rate constants with experimental data of Nesic et al. (1995) on the basis of RMSE value.

Table 7.3 Comparison of corrosion prediction model in Chapter 3, numerical model of (Srinivasan, 2015), corrosion prediction model in Chapter 3 using optimal electrochemical reaction rate constants and experimental data of (Nesic et al., 1995).

Corrosion Prediction models	RMSE
CFD-driven mass transfer model – Chapter 3	0.66
Srinivasan (2015)	0.52
CFD-driven mass transfer model - Chapter 5	0.28

In evaluating the performance of corrosion prediction models in the current study, the CFD-driven mass transfer model developed in Chapter 3 under-predicted corrosion rates, whereas the numerical model of Srinivasan (2015) over-predicted corrosion rates. Following the implementation of optimal electrochemical reaction constants obtained using the SBO approach, the RMSE value decreased from 0.66 to 0.28. The newly derived values accurately predicted corrosion rates for higher velocities and stayed within the error bar for low to moderate velocities.

To conclude, the SBO approach has proven to be significant. These findings have shown how significant the sensitivity studies can be in corrosion prediction modelling. This work also highlights the effectiveness of the SBO approach constructed with the help of machine learning modelling.

7.5 Coupling of CFD Model and Mass Transfer Model for the Prediction of CO_2 Corrosion in Complex Geometries

The development of a novel methodology of coupling CFD model and mass transfer model for the prediction of CO_2 corrosion in pipelines has been utilized in cases where complex flow situations occur. This methodology consists of evaluating the velocity profile at a series of mesh node points to obtain VSL conditions. These calculations are then coupled with the mass transfer model to predict corrosion rate in gradual constriction geometry. The gradual constriction geometry has a variation in diameter

along the length and consists of three regions: expansion, gradual constriction, and straight constriction. It is well known that as the flow velocity increases, the velocity gradient near the wall reduces significantly. Consequently, the laminar region decreases.

In the model developed in Chapter 6, the VSL predictions have shown that in the expansion region, the VSL value is the highest overall as the velocity is the lowest compared to the other regions in the domain. This is also due to the large diameter, which is responsible for having a thicker hydrodynamic boundary layer. The gradual constriction region has the lowest VSL values due to increased turbulence. The CFD model developed in this study has shown that it is capable of predicting VSL conditions for which empirical correlations are not applicable.

The VSL conditions obtained for 1 m/s to 5 m/s velocities were coupled to predict corrosion rates in gradual constriction geometry. The model developed found the highest corrosion rate value at the entrance of the constriction region after the inclination. This is due to an increase in the rate of species transport as a result of the increased rate of turbulent diffusion and reduced VSL value. The lowest corrosion rate in the domain for all velocities is in the expansion region, as the turbulence in this region is the lowest. It also found that the corrosion rate increased in the direction of flow in the gradual constriction region due to increased turbulence. In the verification of the current model against the experimental dataset, challenges were posed due to missing pCO_2 values and incomplete dimension information in Zhong et al. (2020). Despite these limitations, the current model has shown that it has the capability to predict corrosion rates in gradual and straight constriction regions quite well. In addition, the current model has demonstrated a trend similar to the experimental dataset of Zhong et al. (2020).

7.6 Future Work

As the aim of this study was to assess the effectiveness of predictions of VSL conditions using CFD, two areas are recommended here. The model developed in the current study did not take into account the effect of electromigration and is valid for low temperatures. The electromigration term can be added to the model by adding $z_j u_j F c_j \nabla \varphi$ in Eq. (3.3). This allows corrosion prediction for low-conductivity solutions. A time-dependent solution scheme is needed to make the model applicable for situations where the formation of $FeCO_3$ corrosion products are considered (Nesic et al., 2001; Nordsveen et al., 2003).

A CFD-driven mass transfer model developed in Chapter 6 should be verified with the reliable experimental dataset available in the literature. In addition to this, the model developed in the current research work did not take into account the formation of galvanic/electrochemical cells resulting from the disturbed flow. The difference in mass transfer levels in complex flow situations is due to the level of turbulence in the regions. These levels create potential variation across the surface, known as galvanic/electrochemical cells. The potential distribution across the surface is simplified by modelling the Laplace equation by ignoring concentration gradients (Li and Woollam, 2012). CFD-driven mass transfer model could potentially aid in understanding flow-induced galvanic corrosion more. A CFD model would then provide a term mentioned in Eq.(2.27), which can be obtained by interpolation of the averaged velocity term across the domain. This methodology is an extension of the model developed in Chapter 6 with the addition of velocity components in the x and y directions. A 2D mass transfer model

for CO_2 corrosion in complex geometries with the help of CFD would be able to determine the anodic and cathodic sites, resulting in potential variation across the surface. This model could experience issues such as a high geometric aspect ratio, which can be addressed by splitting the domain and linking the boundary conditions.

References

Abbas, M.H. 2016. Modelling of CO_2 corrosion of pipeline steels. Ph.D. Thesis, Newcastle University.

Abbas, M.H., Norman, R. and Charles, A. 2018. Neural network modelling of high pressure CO_2 corrosion in pipeline steels. Process Safety and Environmental Protection. 119, pp.36–45.

Abe, K., Kondoh, T. and Nagano, Y. 1994. A new turbulence model for predicting fluid flow and heat transfer in separating and reattaching flow. 1: Flow field calculations. International Journal of Heat and Mass Transfer. 37(1), pp.139–151.

Aburebaiya, H. H. 2020. Aerodynamic shape optimisation of conceptual over-body designs for heavy goods vehicles. Ph.D. Thesis, University of Leeds.

Aghaaminiha, M., Mehrani, R., Colahan, M., Brown, B., Singer, M., Nesic, S., Vargas, S.M. and Sharma, S. 2021. Machine learning modeling of time-dependent corrosion rates of carbon steel in presence of corrosion inhibitors. Corrosion Science. 193(August), p.109904.

Ahmed, M. and Qin, N. 2009. Surrogate-Based Aerodynamic Design Optimization: Use of Surrogates in Aerodynamic Design Optimization. International Conference on Aerospace Sciences and Aviation Technology. 13(AEROSPACE SCIENCES), pp.1–26.

Alasad, S. A., & Bhaya, W. S. (2017). Review of Data Preprocessing Techniques in Data Mining. Journal of Engineering and Applied Sciences, 16(September), 4102–4107. https://doi.org/10.3923/jeasci.2017.4102.4107

Alawadhi, K. 2009. Inhibition of weld corrosion in flowing brines containing carbon dioxide. Ph.D. Thesis, Cranfield University.

Al-Khateeb, M., Barker, R., Neville, A. and Thompson, H. 2018. An experimental and theoretical investigation of the influence of surface roughness on corrosion in CO_2 environments. Journal of Corrosion Science and Engineering. (June).

Al-Moubaraki, A.H. and Obot, I.B. 2021. Corrosion challenges in petroleum refinery operations: Sources, mechanisms, mitigation, and future outlook. Journal of Saudi Chemical Society. 25(12), p.101370.

António, N. and Rodrigues, S. 2021. Gaussian Process Regression for Data-Driven Model Predictive Control. Delft University of Technology.

Antony, J. 2014. 6 - Full Factorial Designs In: J. B. T.-D. of E. for E. and S. Antony, ed. Oxford: Elsevier, pp.63–85. Available from: https://www.sciencedirect.com/science/article/pii/B9780080994178000067.

Aravinth, S. 2000. Prediction of heat and mass transfer for fully developed turbulent fluid flow through tubes. International Journal of Heat and Mass Transfer. 43(8), pp.1399–1408.

Bagherian, M.A., Mehranzamir, K., Pour, A.B., Rezania, S., Taghavi, E., Nabipour-afrouzi, H., Dalvi-esfahani, M. and Alizadeh, S.M. 2021. Classification and analysis of optimization techniques for integrated energy systems utilizing renewable energy sources: A review for chp and cchp systems. Processes. 9(2), pp.1–36.

144

Balint, T. S. (2001). Dynamics of the upper airway [University of Warwick]. http://go.warwick.ac.uk/wraphttp://go.warwick.ac.uk/wrap/66409

Balsara, N.P. and Newman, J. 2021. Electrochemical Systems. Newark, United States: John Wiley & Sons, Incorporated.

Bamford, C. H., & Tiffer, C. F. H. (1972). Comprehensive chemical kinetics (Volume 6). Elsevier.

Bao, X., Wang, Z., Fu, D., Shi, C., Iglesias, G., Cui, H. and Sun, Z. 2022. Machine learning methods for damage detection of thermoplastic composite pipes under noise conditions. Ocean Engineering. 248(September 2021), p.110817.

Bassam, A., Ortega-Toledo, D., Hernandez, J.A., Gonzalez-Rodriguez, J.G. and Uruchurtu, J. 2009. Artificial neural network for the evaluation of CO 2 corrosion in a pipeline steel. Journal of Solid State Electrochemistry. 13(5), pp.773–780.

Bates, S.J., Sienz, J. and Toropov, V. V. 2004. Formulation of the optimal latin hypercube design of experiments using a permutation genetic algorithm. Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. 7(April 2004), pp.5217–5223.

Berger, F.P. and Hau, K.F.F.L. 1977. Mass transfer in turbulent pipe flow measured by the electrochemical method. International Journal of Heat and Mass Transfer. 20(11), pp.1185–1194.

Bertsimas, D. and Tsitsiklis, J. 1993. Simulated annealing. Statistical science. 8(1), pp.10–15.

Bertsimas, D. and Tsitsiklis, J.N. 1997. Introduction to linear optimization. Athena scientific Belmont, MA.

de Boer, G.N., Johns, A., Delbosc, N., Burdett, D., Tatchell-Evans, M., Summers, J. and Baudot, R. 2018. Three computational methods for analysing thermal airflow distributions in the cooling of data centers. International Journal of Numerical Methods for Heat and Fluid Flow. (January).

Booker, A.J., Dennis, J.E., Frank, P.D., Serafini, D.B., Torczon, V. and Trosset, M.W. 1999. A rigorous framework for optimization of expensive functions by surrogates. Structural Optimization. 17(1), pp.1–13.

Box, G.E. and Draper, N.R. 1987. Empirical model-building and response surfaces. John Wiley \& Sons.

Box, G.E.P. and Wilson, . B. 1951. On the Experimental Attainment of Optimum Conditions Published by: Wiley for the Royal Statistical Society Stable URL : http://www.jstor.org/stable/2983966. Journal of the Royal Statistical Society - Series B. 13(1), pp.1–45.

Bozzini, B., Ricotti, M.E., Boniardi, M. and Mele, C. 2003. Evaluation of erosion-corrosion in multiphase flow via CFD and experimental analysis. Wear. 255(1–6), pp.237–245.

Braumann, A., Kraft, M. and Mort, P.R. 2010. Parameter estimation in a multidimensional granulation model. Powder Technology. 197(3), pp.196–210.

Braumann, A., Man, P.L.W. and Kraft, M. 2010. Statistical approximation of the inverse problem in multivariate population balance modeling. Industrial and Engineering Chemistry Research. 49(1), pp.428–438.

Breiman, L. 2001. Random forests. Machine learning. 45, pp.5–32.

Brereton, R.G. and Lloyd, G.R. 2010. Support Vector Machines for classification and regression. Analyst. 135(2), pp.230–267.

Brunton, S.L., Noack, B.R. and Koumoutsakos, P. 2020. Machine Learning for Fluid Mechanics. Annual Review of Fluid Mechanics. 52, pp.477–508.

Brusco, M.J. and Stahl, S. 2005. Branch-and-bound applications in combinatorial data analysis. Springer.

Bursztyn, D. and Steinberg, D.M. 2006. Comparison of designs for computer experiments. Journal of Statistical Planning and Inference. 136(3), pp.1103–1119.

Castellanos, V., Albiter, A., Hernández, P. and Barrera, G. 2011. Failure analysis expert system for onshore pipelines. Part - I: Structured database and knowledge acquisition. Expert Systems with Applications. 38(9), pp.11085–11090.

Chang, Y.-C., Woollam, R. and Orazem, M.E. 2014. Mathematical Models for Under-Deposit Corrosion. Journal of The Electrochemical Society. 161(6), pp.C321–C329.

Chen, C. and Mansfeld, F. 1997. Potential distribution in the Evan's drop experiment. Corrosion Science. 39(2), pp.409–413.

Chen, X., Yang, X., Zeng, M., & Wang, H. (2023). Influence of CO_2 partial pressure and flow rate on the corrosion behavior of N80 steel in 3.5% NaCl. International Journal of Electrochemical Science, 18(8). https://doi.org/10.1016/J.IJOES.2023.100218

Cherkassky, V. and Ma, Y. 2002. Selection of meta-parameters for support vector regression. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2415 LNCS, pp.687–693.

Choi, K.K., Youn, B.D. and Yang, R. 2001. Moving Least Square Method for Reliability-Based Design Optimization In: The Fourth World Congress of Structural and Multidisciplinary Optimization.

Choi, Y.-S. and Nešic, S. 2009. Corrosion Behavior of Carbon Steel In Supercritical CO_2-Water Environments In: NACE CORROSION.

Choi, Y.S., Nesic, S. and Young, D. 2010. Effect of impurities on the corrosion behavior of carbon steel in supercritical CO_2 - Water environments. NACE - International Corrosion Conference Series., pp.9233–9238.

Chou, J.S., Ngo, N.T. and Chong, W.K. 2017. The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate. Engineering Applications of Artificial Intelligence. 65(August 2016), pp.471–483.

COMSOL 2016. CFD Module User's Guide.

Crary, S.B. 2002. Design of computer experiments for metamodel generation. Analog Integrated Circuits and Signal Processing. 32(1), pp.7–16.

Crolet, J.L. and Bonis, M.R. 1991. Prediction of the risks of CO sub 2 corrosion in oil and gas wells. SPE production engineering. 6:4.

Davies, J.T. 1972. Turbulence Phenomenon. Academic Press New York and London.

Delahay, P. 1952. Implications of the Kinetics of Ionic Dissociation with Regard to Some Electrochemical Processes-Application to Polarography. Journal of the American Chemical Society. 74(14), pp.3497–3500.

Dhaubhadel, M.N. 1996. Review: Cfd applications in the automotive industry. Journal of Fluids Engineering, Transactions of the ASME. 118(4), pp.647–653.

Djukic, T., Topalovic, M. and Filipovic, N. 2022. Validation of lattice Boltzmann based software for blood flow simulations in complex patient-specific arteries against traditional CFD methods. Mathematics and Computers in Simulation. 203, pp.957–976.

Dorigo, M., Birattari, M. and Stutzle, T. 2006. Ant colony optimization. IEEE computational intelligence magazine. 1(4), pp.28–39.

Du, X., He, P. and Martins, J.R.R.A. 2021. Rapid airfoil design optimization via neural networks-based parameterization and surrogate modeling. Aerospace Science and Technology. 113, p.106701.

Dugstad, A. 1998. Mechanism of Protective Film Formation during CO_2 Corrosion of Carbon Steel. Corrosion. Houston, TX: NACE International.

Dugstad, A., Lunde, L. and Videm, K. 1994. Parametric study of CO_2 corrosion of carbon steel In: Corrosion. Houston, TX: NACE.

Efird, K. D. (2000). Jet impingement testing for flow accelerated corrosion. *NACE CORROSION*, NACE-00052.

Eggels, J.G.M., Unger, F., Weiss, M.H., Westerweel, J., Adrian, R.J., Friedrich, R. and Nieuwstadt, F.T.M. 1994. Fully developed turbulent pipe flow: a comparison between direct numerical simulation and experiment. Journal of Fluid Mechanics. 268, pp.175–210.

El-Batsh, H.M., Doheim, M.A. and Hassan, A.F. 2012. On the application of mixture model for two-phase flow induced corrosion in a complex pipeline configuration. Applied Mathematical Modelling. 36(11), pp.5686–5699.

El-Gammal, M., Mazhar, H., Cotton, J.S., Shefski, C., Pietralik, J. and Ching, C.Y. 2010. The hydrodynamic effects of single-phase flow on flow accelerated corrosion in a 90-degree elbow. Nuclear Engineering and Design. 240(6), pp.1589–1598.

Eslahpazir, M., Krull, R. and Krühne, U. 2019. Computational fluid dynamics In: Comprehensive biotechnology. Elsevier, pp.95–107.

Evans, U.R. 1960. The Corrosion and Oxidation of Metals: Scientific Principles and Practical Applications. London SE -: Edward Arnold.

Fan, W., Xu, B., Li, H., Lu, G. and Liu, Z. 2022. A novel surrogate model for channel geometry optimization of PEM fuel cell based on Bagging-SVM Ensemble Regression. International Journal of Hydrogen Energy. 47(33), pp.14971–14982.

Fang, K.-T. and Li, R. 2006. Uniform design for computer experiments and its optimal properties. International Journal of Materials and Product Technology. 25(1–3), pp.198–210.

Fisher, R. 1937. The Design of Experiments. London: Oliver and Boyd

Forkan, A.R.M., Kang, Y. Bin, Jayaraman, P.P., Liao, K., Kaul, R., Morgan, G., Ranjan, R. and Sinha, S. 2022. CorrDetector: A framework for structural corrosion detection from drone images using ensemble deep learning. Expert Systems with Applications. 193(December 2021), p.116461.

Forrester, A.I.J. 2008. Engineering design via surrogate modelling: a practical guide. Chichester: J. Wiley.

Gandomi, A.H., Yang, X.-S., Talatahari, S. and Alavi, A.H. 2013. Metaheuristic algorithms in modeling and optimization. Metaheuristic applications in structures and infrastructures, 1, pp. 1-24.

Gardner, T., Paolinelli, L.D. and Nesic, S. 2019. Study of water wetting in oil-water flow in a small-scale annular flume. Experimental Thermal and Fluid Science. 102(July 2018), pp.506–516.

Gentle, J.E. (ed.). 2003. Simulating Random Numbers from a Uniform Distribution BT - Random Number Generation and Monte Carlo Methods In: New York, NY, NY: Springer New York, pp.1–60.

Goldberg, D. and Richardson, J. 1987. Genetic algorithms with sharing for multi-modal functions optimization. in Genetic Algorithms and Their Applications. ICGA-87 pp., pp.44–49.

Gosmann, L., Geitner, C. and Wieler, N. 2022. Data-driven forward osmosis model development using multiple linear regression and artificial neural networks. Computers and Chemical Engineering. 165(June), p.107933.

Gray, L.G.S., Anderson, B.G., Danysh, M.J. and Tremaine, P.R. 1989. Mechanisms of carbon steel corrosion in brines containing dissolved carbon dioxide at pH 4. Corrosion. pp.1–19.

Gunaltun, Y.M. 1996. Combining Research and Field Data for Corrosion Rate Prediction In: The NACE international Annual Conference and Exposition.

Gunst, R.F. 1996. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Technometrics. 38(3), pp.284–286.

Halvorsen, A.M.K. and Sontvedt, T. 1999. CO_2 corrosion model for carbon steel including a wall shear stress model for multiphase flow and limits for production rate to avoid mesa attack. Corrosion. (42).

Hatami, S., Ghaderi-Ardakani, A., Niknejad-Khomami, M., Karimi-Malekabadi, F., Rasaei, M.R. and Mohammadi, A.H. 2016. On the prediction of CO_2 corrosion in petroleum industry. Journal of Supercritical Fluids. 117, pp.108–112.

Hebert, J.C., Thibodeaux, T.B. and Hernandez, J.G. 2019. A study of preferential weld corrosion in produced water with elevated organic acids and low partial pressure

CO_2 on an aged gulf of Mexico offshore facility In: NACE CORROSION., NACE-2019-13110.

Heitz, E. 1996. Mechanistically based prevention strategies of flow-induced corrosion. Electrochimica Acta. 41(4 SPEC. ISS.), pp.503–509.

Hesjevik, S.M., Olsen, S. and Seiersten, M. 2003. Corrosion at high CO_2 pressure In: NACE CORROSION. NACE.

Hooke, R. and Jeeves, T.A. 1961. ``Direct Search'' Solution of Numerical and Statistical Problems. Journal of the ACM (JACM). 8(2), pp.212–229.

Hu, H. and Cheng, Y.F. 2016. Modeling by computational fluid dynamics simulation of pipeline corrosion in CO_2 -containing oil-water two phase flow. Journal of Petroleum Science and Engineering. 146, pp.134–141.

Hu, H., Xu, H., Huang, C., Chen, X., Li, X. and Li, Y. 2018. Numerical predictions of uniform CO_2 corrosion in complex fluid domains using low Reynolds number models. Metals. 8(12).

Van Hunnik, E., Pots, B.F.M. and Hendriksen, E.L.J.A. 1996. The Formation of Protective $FeCO_3$ Corrosion Product Layers in CO_2 Corrosion In: NACE - International Corrosion Conference Series. NACE.

Hysing, S. 2012. Mixed element FEM level set method for numerical simulation of immiscible fluids. Journal of Computational Physics. 231(6), pp.2449–2465.

Isaac, J. 2013. CFD in the Aerospace and Aeronautics Industries. Journal of Aeronautics & Aerospace Engineering. 02(03), pp.2–5.

Ibrahimi, B. El, Nardeli, J.V. and Guo, L. 2021. An overview of corrosion In: Sustainable Corrosion Inhibitors I: Fundamentals, Methodologies, and Industrial Applications, pp. 1-19.

Izonin, I., Member, I. and Strauss, C. 2022. Towards data normalization task for the efficient mining of medical data (2021), pp.480–484.

Jepson, W.P., Bhongale, S. and Gopal, M. 1996. Predictive model for sweet corrosion in horizontal multiphase slug flow. NACE - International Corrosion Conference Series. 1996-March.

Ji, J., Zhang, C., Kodikara, J. and Yang, S.Q. 2015. Prediction of stress concentration factor of corrosion pits on buried pipes by least squares support vector machine. Engineering Failure Analysis. 55, pp.131–138.

Jollife, I.T. and Cadima, J. 2016. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 374(2065).

Shen-Haw Ju, Cheng-Han Hsieh, Optimal wind turbine jacket structural design under ultimate loads using Powell's method, Ocean Engineering, Volume 262, 2022, 112271, ISSN 0029-8018, https://doi.org/10.1016/j.oceaneng.2022.112271. (https://www.sciencedirect.com/science/article/pii/S0029801822015773)

Kacker, R.N., Lagergren, E.S. and Filliben, J.J. 1991. Taguchi's orthogonal arrays are classical designs of experiments. Journal of Research of the National Institute of Standards and Technology. 96(5), pp.577–591.

Kahyarian, A., Achour, M. and Nesic, S. 2017. Mathematical modeling of uniform CO_2 corrosion. Trends in oil and gas corrosion research and technologies, pp. 805-849.

Kahyarian, A. and Nesic, S. 2020. On the mechanism of carbon dioxide corrosion of mild steel: Experimental investigation and mathematical modeling at elevated pressures and non-ideal solutions. Corrosion Science. 173(May), p.108719.

Kahyarian, A., Singer, M. and Nesic, S. 2016. Modeling of uniform CO_2 corrosion of mild steel in gas transportation systems: A review. Journal of Natural Gas Science and Engineering. 29, pp.530–549.

Kanwar, S. 1994. Study and modelling of sweet corrosion of multiphase mixtures in horizontal pipelines. Master's thesis, Ohio University.

Keating, A. and Nesic, S. 2000. Numerical prediction of erosion-corrosion in bends In: CORROSION 2000. Houston, TX: NACE International.

Kennedy, J. and Eberhart, R. 1995. Particle swarm optimization In: Proceedings of ICNN'95- International conference on neural networks. IEEE, pp.1942–1948.

Kermani, M.B. and Harrop, D. 1996. The impact of corrosion on oil and gas industry. SPE production & facilities. 11(3), pp.186–190.

Kermani, M.B. and Morshed, A. 2003. Carbon dioxide corrosion in oil and gas production - A compendium. Corrosion. 59(8), pp.659–683.

Khalajestani, M.K. and Bahaari, M.R. 2014. Investigation of pressurized elbows containing interacting corrosion defects. International Journal of Pressure Vessels and Piping. 123, pp.77–85.

Kharaka, Y.K., Gunter, W.D. and Aggarwal, P.K. 1988. A computer program for geochemical modeling of water-rock interactions. California.

Kiefer, J. and Wolfowitz, J. 1959. Optimum designs in regression problems. The Annals of Mathematical Statistics. 30, pp.271–294.

Kim, H., Robert, C.P. and Casella, G. 2000. Monte Carlo Statistical Methods. Technometrics. 42(4), p.430.

Koch, G.H., Brongers, P.H.M. and Thompson, and N.G. 2002. Corrosion costs and preventive strategies in the United States.

Koehler, J.R. and Owen, A.B.B.T.-H. of S. 1996. 9 Computer experiments In: Design and Analysis of Experiments. Elsevier, pp.261–308.

Kopbayev, A., Khan, F., Yang, M. and Halim, S.Z. 2022. Gas leakage detection using spatial and temporal neural network model. Process Safety and Environmental Protection. 160, pp.968–975.

Kopliku, A. and Gunaltun, Y. 2006. Field data collection, evaluation and use for corrosivity prediction and validation of models - Part I: Collection of reliable field

data for validation of prediction models In: NACE CORROSION (pp. NACE-06117). NACE.

Kvarekval, J. 1997. A kinetic model for calculating concentration profiles and fluxes of CO_2 -related species across the Nernst diffusion layer. Corrosion. (97005).

Lam, C.K.G. and Bremhorst, K. 1981. A modified form of the k-ε model for predicting wall turbulence.

Launder, B.E. and Spalding, D.B. 1983. The numerical computation of turbulent flows In: Numerical Prediction of Flow, Heat Transfer, Turbulence and Combustion. Elsevier Ltd, pp.96–116.

Lee, L.H., Rajkumar, R., Lo, L.H., Wan, C.H. and Isa, D. 2013. Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach. Expert Systems with Applications. 40(6), pp.1925–1934.

Lemieux, C. 2009. Quasi–Monte Carlo Constructions BT - Monte Carlo and Quasi-Monte Carlo Sampling In: New York, NY, NY: Springer New York, pp.1–61.

Li, D. and Sun, X. 2006. Nonlinear integer programming. Springer.

Li, H. and Woollam, R. 2012. Mechanistic modelling of flow enhanced corrosion in complex geometry In: NACE - International Corrosion Conference Series.

Li, H., Xu, B., Lu, G., Du, C. and Huang, N. 2021. Multi-objective optimization of PEM fuel cell by coupled significant variables recognition, surrogate models and a multi-objective genetic algorithm. Energy Conversion and Management. 236(December 2020), p.114063.

Li, J., Zhang, M., Martins, J.R.R.A. and Shu, C. 2020. Efficient aerodynamic shape optimization with deep-learning-based geometric filtering. AIAA Journal. 58(10), pp.4243–4259.

Li, Q., Hu, H. and Cheng, Y.F. 2016. Corrosion of pipelines in CO_2-saturated oil-water emulsion flow studied by electrochemical measurements and computational fluid dynamics modeling. Journal of Petroleum Science and Engineering. 147, pp.408–415.

Li, Y. and Nielsen, P. V. 2011. Commemorating 20 years of Indoor Air: CFD and ventilation research. Indoor Air. 21(6), pp.442–453.

Lin, C.S., Moulton, R.W. and Putnam, G.L. 1953. Mass transfer between solid wall and fluid streams - Mechanism and eddy distribution relationships in turbulent flow. Industrial & Engineering Chemistry. 45(3), pp.636–640.

Liu, G., Tree, D.A. and High, M.S. 1994. Relationships between rotating disk corrosion measurements and corrosion in pipe flow. Corrosion. 50(8), pp.584–593.

Liu, J., Chen, R., Lou, J., Hu, Y. and You, Y. 2023. Deep-learning-based aerodynamic shape optimization of rotor airfoils to suppress dynamic stall. Aerospace Science and Technology. 133, p.108089.

Loweth, E.L., De Boer, G.N. and Toropov, V. V. 2011. Practical recommendations on the use of moving least squares metamodel building. Proceedings of the 13th International Conference on Civil, Structural and Environmental Engineering Computing. (September), pp.1–13.

Lv, W., Ou, G., Liu, X. and Gong, C. 2020. Experimental and numerical studies on the corrosion properties of AISI 316L stainless steel in two-phase upward slug flows. Engineering Applications of Computational Fluid Mechanics. 14(1), pp.897–909.

Ma, Y., Dong, H., Li, P., Wu, B., Huang, L., Zhang, L., Li, C. and Li, J. 2023. Galvanic corrosion of AA5052/304SS welded joint with Zn-based filler metal in marine engineering. Corrosion Science. 211(November 2022), p.110912.

Manninen, M., Taivassalo, V. and Kallio, S. 1996. On the mixture model for multiphase flow, Technical Research Centre of Finland, Finland.

De Masi, G., Gentile, M., Vichi, R., Bruschi, R. and Gabetta, G. 2015. Machine learning approach to corrosion assessment in subsea pipelines. MTS/IEEE OCEANS 2015 - Genova: Discovering Sustainable Ocean Energy for a New World., pp.8–13.

Martínez-Palou, R., de Lourdes Mosqueira, M., Zapata-Rendón, B., Mar-Juárez, E., Bernal-Huicochea, C., de la Cruz Clavel-López, J. and Aburto, J. 2011. Transportation of heavy and extra-heavy crude oil by pipeline: A review. Journal of Petroleum Science and Engineering. 75(3), pp.274–282.

McKay, M.D., Beckman, R.J. and Conover, W.J. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 21(2), pp.239–245.

Menter, F.R. 1994. Two-equation eddy-viscosity turbulence models for engineering applications. AIAA Journal. 32(8), pp.1598–1605.

Michel, T.M. 1997. Does machine learning really work? AI Magazine. 18(3), pp.71–83.

Mohamed, I. and Jenkins, N. 2004. Proton exchange membrane (PEM) fuel cell stack configuration using genetic algorithms. Journal of Power Sources. 131(1–2), pp.142–146.

Montgomery, D.C. 2013. Design and Analysis of Experiments Eighth Edition. Arizona State University.

Mrzygłód, B., Hawryluk, M., Janik, M. and Olejarczyk-Wożeńska, I. 2020. Sensitivity analysis of the artificial neural networks in a system for durability prediction of forging tools to forgings made of C45 steel. International Journal of Advanced Manufacturing Technology. 109(5–6), pp.1385–1395.

Mukerjee, R. and Wu, C.F.J. (eds.). 2006. Fundamentals of Factorial Designs BT - A Modern Theory of Factorial Designs In: New York, NY, NY: Springer New York, pp.9–48.

Myers, R.H. 1990. Classical and modern regression with applications Second edi. Boston: Duxbury.

Myers, T.G., Valverde, A. and Cabrera-Codony, A. 2022. On the Development of a Consistent Mathematical Model for Adsorption in a Packed Column (and Why Standard Models Fail). SSRN Electronic Journal., pp.1–31.

Narimani, N., Zarei, B., Pouraliakbar, H. and Khalaj, G. 2015. Predictions of corrosion current density and potential by using chemical composition and corrosion cell characteristics in microalloyed pipeline steels. Measurement: Journal of the International Measurement Confederation. 62, pp.97–107.

Nelder, J.A. and Mead, R. 1965. A simplex method for function minimization. The Computer Journal. 7(4), pp.308–313.

Nešić, S. 2007. Key issues related to modelling of internal corrosion of oil and gas pipelines - A review. Corrosion Science. 49(12), pp.4308–4338.

Nesic, S., Cai, J. and Lee, K.L.J. 2005. A multiphase flow and internal corrosion prediction model for mild steel pipelines In: Corrosion. pp.1–35.

Nešić, S., Hui, L., Huang, J. and Sormaz, D. 2009. An open source mechanistic model for CO_2/H_2S corrosion of carbon steel. NACE CORROSION (pp. NACE-09572). NACE.

Nešić, S., Kahyarian, A. and Choi, Y.S. 2019. Implementation of a comprehensive mechanistic prediction model of mild steel corrosion in multiphase oil and gas pipelines. Corrosion. 75(3), pp.274–291.

Nesic, S. and Lunde, L. 1994. Carbon Dioxide Corrosion of Carbon Steel in Two-Phase Flow. Corrosion (Houston, Tex.). 50(9), pp.717–727.

Nesic, S., Nordsveen, M., Maxwell, N. and Vrhovac, M. 2001. Probabilistic modelling of CO_2 corrosion laboratory data using neural networks. Corrosion Science. 43(7), pp.1373–1392.

Nesic, S., Nordsveen, M., Nyborg, R. and Stangeland, A. 2001. A mechanistic model for CO_2 corrosion with protective iron carbonate films In: CORROSION. Houston, TX: NACE International.

Nesic, Sr, Postlethwaite, J. and Olsen, S. 1995. An electrochemical model for prediction of corrosion of mild steel in aqueous carbon dioxide solutions. Corrosion. 52(4), pp.280–294.

Nesic, S., Solvi, G.T. and Enerhaug, J. 1995. Comparison of the rotating cylinder and pipe flow tests for flow-sensitive carbon dioxide corrosion. Corrosion. 51(10), pp.773–787.

Nesic, S. and Thevenot, N. 1996. Electrochemical properties of iron dissolution in the presence of CO_2- basics revisited In: The NACE international Annual Conference and Exposition.

Nesic, S., Wang, S., Cai, J. and Xiao, Y. 2004. Integrated CO_2 corrosion - Multiphase flow model. Proceedings - First International Oilfield Corrosion Symposium; Managing Corrosion in Oil and Gas Production: Fluids, Chemicals, Materials, and More. (04626), pp.43–54.

Newman, J.S. 1991. Electrochemical systems. Englewood Cliffs, N.J.: Prentice Hall.

Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Van Le, H., Tran, V.Q., Prakash, I. and Pham, B.T. 2021. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Mathematical Problems in Engineering. 2021.

Nordsveen, M, Nešić, S., Nyborg, R. and Stangeland, A. 2003. A mechanistic model for carbon dioxide corrosion of mild steel in the presence of protective iron carbonate films - Part 1: Theory and verification. Corrosion. 59(5), pp.443–456.

Nordsveen, M., Nešić, S., Nyborg, R., Stangeland, A. and Lee, K.-L.J.L. 2003. A mechanistic model for carbon dioxide corrosion of mild steel in the presence of protective iron carbonate films - Part 1: Theory and verification. Corrosion. 59(5), pp.443–456.

Notter, R.H. and Sleicher, C.A. 1971. The eddy diffusivity in the turbulent boundary layer near a wall. Chemical Engineering Science. 26(1), pp.161–171.

Nyborg, R. 2010. CO_2 corrosion models for oil and gas production systems In: NACE - International Corrosion Conference Series., pp.527–536.

Nyborg, R. 2002. Overview of CO_2 corrosion models for wells and pipelines. NACE - International Corrosion Conference Series. 2002-April (02233), pp.1–16.

Oddo, J.E. and Tomson, M.B. 1982. Simplified Calculation of $CaCO_3$ Saturation at High Temperatures and Pressures in Brine Solutions. JPT, Journal of Petroleum Technology. 34(7), pp.1583–1590.

Olsen, S. 2003. CO_2 Corrosion prediction by use of the Norsok M-506 model - guidelines and limitations In: Corrosion. NACE CORROSION., pp.1–12.

Osarogiagbon, A.U., Khan, F., Venkatesan, R. and Gillard, P. 2021. Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations. Process Safety and Environmental Protection. 147, pp.367–384.

Ossai, C.I. 2019. A data-driven machine learning approach for corrosion risk assessment—a comparative study. Big Data and Cognitive Computing. 3(2), pp.1–22.

Ossai, C.I. 2020. Corrosion defect modelling of aged pipelines with a feed-forward multi-layer neural network for leak and burst failure estimation. Engineering Failure Analysis. 110(March 2019), p.104397.

Ouladbrahim, A., Belaidi, I., Khatir, S., Magagnini, E., Capozucca, R. and Abdel Wahab, M. 2022. Experimental crack identification of API X70 steel pipeline using improved Artificial Neural Networks based on Whale Optimization Algorithm. Mechanics of Materials. 166(December 2021), p.104200.

Owen, J., Godfrey, J., Ma, W., de Boer, G., Al-Khateeb, M., Thompson, H., Neville, A., Ramsey, C. and Barker, R. 2019. An experimental and numerical investigation of CO_2 corrosion in a rapid expansion pipe geometry. Corrosion Science. p.108362.

Owen, J.J. 2018. Erosion-corrosion of carbon steel in complex flow geometries in oil & gas CO_2 environments. The University of Leeds.

Pai, K.N., Prasad, V. and Rajendran, A. 2020. Experimentally validated machine learning frameworks for accelerated prediction of cyclic steady state and optimization of

pressure swing adsorption processes. Separation and Purification Technology. 241(January), p.116651.

Palmer, D.A. and Van Eldik, R. 1983. The Chemistry of Metal Carbonato and Carbon Dioxide Complexes. Chemical Reviews. 83(6), pp.651–731.

Peng, S., Zhang, Z., Liu, E., Liu, W. and Qiao, W. 2021. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline. Journal of Natural Gas Science and Engineering. 85(May 2020), p.103716.

Peng, X., Wu, W., Zhang, Y. and Yang, W. 2017. Determination of operating parameters for PEM fuel cell using support vector machines approach. Journal of Energy Storage. 13, pp.409–417.

Perez, T.E. 2013. Corrosion in the oil and gas industry: An increasing challenge for materials. Jom. 65(8), pp.1033–1042.

Perry, R.H. and Green, D.W. 1987. Perry's Chemical Engineers' Handbook 50th ed. New York: McGraw-Hill.

Pietralik, J.M. 2012. The Role of Flow in Flow-Accelerated Corrosion under Nuclear Power Plant Conditions. Journal of Advanced Maintenance. 4(2), pp.63–78.

Popovich, A.T. and Hummel, R.L. 1967. Experimental study of the viscous sublayer in the pipe flow. AIChE journal. 13(5), pp.854–860.

Powell, M.J.D. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. The Computer Journal. 7(2), pp.155–162.

Prasad, M., Gopika, V., Sridharan, A. and Parida, S. 2018. Pipe wall thickness prediction with CFD based mass transfer coefficient and degradation feedback for flow accelerated corrosion. Progress in Nuclear Energy. 107(July 2017), pp.205–214.

Qi, X., Chen, G., Li, Y., Cheng, X. and Li, C. 2019. Applying Neural-Network-Based Machine Learning to Additive Manufacturing: Current Applications, Challenges, and Future Perspectives. Engineering. 5(4), pp.721–729.

Queipo, N. V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R. and Kevin Tucker, P. 2005. Surrogate-based analysis and optimization. Progress in Aerospace Sciences. 41(1), pp.1–28.

Quinlan, J.R. 1987. Simplifying decision trees. Int. J. Man Mach. Stud. 27, pp.221–234.

Rao, B.C.R. 1947. Factorial Experiments Derivable from Combinatorial Arrangements of Arrays Author (s): C . Radhakrishna Rao Source: Supplement to the Journal of the Royal Statistical Society, 1947, Vol. 9, No. 1 Published by: Wiley for the Royal Statistical Society. 9(1), pp.128–139.

Rasmussen, C.E. and Williams, K.I.C. 2006. Gaussian Process for Machine Learning. The MIT Press.

Roache, P.J. 1998. Verification of codes and calculations. AIAA Journal. 36(5).

Rosen, C. and Tragardh, C. 1995. Prediction of turbulent high Schmidt number mass transfer a low Reynolds number k--E turbulence model using. The chemical Engineering Journal. 59, pp.153–159.

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. 1989. Design and analysis of computer experiments. Statistical Science. 4(4), pp.409–423.

Sainz-Rosales, A., Ocampo-Lazcarro, X., Hernández-Pérez, A., González-Gutiérrez, A.G., Larios-Durán, E.R., Ponce de León, C., Walsh, F.C., Bárcena-Soto, M. and Casillas, N. 2022. Classic Evans's Drop Corrosion Experiment Investigated in Terms of a Tertiary Current and Potential Distribution. Corrosion and Materials Degradation. 3(2), pp.270–280.

Sakoe, H. and Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing. 26(1), pp.43–49.

Sanchez-Caldera, L.E., Griffith, P. and Rabinowicz, E. 1988. The mechanism of corrosion-erosion in steam extraction lines of power stations. Journal of Engineering for Gas Turbines and Power. 110(2), pp.180–184.

Ben Seghier, M.E.A., Höche, D. and Zheludkevich, M. 2022. Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques. Journal of Natural Gas Science and Engineering. 99(September 2021).

Senouci, A., Elabbasy, M., Elwakil, E., Abdrabou, B. and Zayed, T. 2014. A model for predicting failure of oil pipelines. Structure and Infrastructure Engineering. 10(3), pp.375–387.

Snelson, E.L. 2007. Flexible and efficient Gaussian process models for machine learning. ACM SIGKDD Explorations Newsletter. 7(2001), pp.1–135.

Sobol', I.M., Asotsky, D., Kreinin, A. and Kucherenko, S. 2011. Construction and Comparison of High-Dimensional Sobol' Generators. Wilmott. 2011(56), pp.64–79.

Spalart, P.R. and Allmaras, S.R. 1994. One-equation turbulence model for aerodynamic flows. Recherche aerospatiale. (1), pp.5–21.

Srinivasan, V. 2015. Evaluation of Flow Coupled CO_2 Corrosion Using CFD: Kinetics & Hydrodynamics In: NACE - International Corrosion Conference Series., pp.1–15.

Stanković, B., Stojanović, A., Sijerčić, M., Belošević, S. and Čantrak, S. 2014. Evaluation and limitations of standard wall functions in channel and step flow configurations. Journal of the Serbian Society for Computational Mechanics. 8(1), pp.1–22.

Stephens, L.I. and Mauzeroll, J. 2019. Demystifying mathematical modeling of electrochemical systems. Journal of Chemical Education. 96(10), pp.2217–2224.

Subraveti, S.G., Li, Z., Prasad, V. and Rajendran, A. 2019. Machine Learning-Based Multiobjective Optimization of Pressure Swing Adsorption. Industrial and Engineering Chemistry Research. 58(44), pp.20412–20422.

Sun, J., Xiao, Q., Wen, J. and Zhang, Y. 2016. Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis.

Measurement: Journal of the International Measurement Confederation. 79, pp.147–157.

Sun, Y. and Nesic, S. 2004. A parametric study and modeling on localized CO_2 corrosion in horizontal wet gas flow, CORROSION/2004, paper no. 380 (Houston, TX: NACE International, 2004).

Taguchi, G. 1959. Linear graphs for orthogonal arrays and their application to experimental design, with the aid of various techniques. Reports of Statistical Applications Research, UJSU. 6(4), pp.1–43.

Taguchi, G. 1989. System of experimental design. Technometrics. 31(2), pp.256–257.

Tapeh, A. T. G., & Naser, M. Z. (2023). Artificial Intelligence, Machine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices. Archives of Computational Methods in Engineering, *30*(1), 115–159. https://doi.org/10.1007/s11831-022-09793-w

Thapa, M. and Missoum, S. 2022. Surrogate-based stochastic optimization of horizontal-axis wind turbine composite blades. Structural and Multidisciplinary Optimization. 65(2), pp.1–18.

Thorat, U., Jones, M., Woollam, R., Owen, J., Barker, R., Thompson, H. and de Boer, G. 2024. Computational fluid dynamics driven mass transfer model for the prediction of CO_2 corrosion in pipelines. Journal of Pipeline Science and Engineering. 4(1), p.100148.

Tominaga, Y. and Stathopoulos, T. 2007. Turbulent Schmidt numbers for CFD analysis with various types of flowfield. Atmospheric Environment. 41(37), pp.8091–8099.

Toropov, V. V, Schramm, U., Sahai, A., Jones, R.D. and Zeguer, T. 2005. Design optimization and stochastic analysis based on the moving least sqaures method In: 6th World Congresses of Structural and Multidisciplinary Optimization.

Tu, J., Yeoh, G.H. and Liu, C. 2018. Computational Fluid Dynamics: A Practical Approach. Saint Louis: Elsevier Science & Technology.

Vapnik, V. 1999. The nature of statistical learning theory. Springer science & business media.

Veersteg, H.K. and Malalasekara, W. 2016. An introduction to computational fluid dynamics: the finite volume method.

Veldhuis, C., Gornicz, T. and Scholcz, T. 2016. Ship optimization using viscous flow computations in combination with generic shape variations and design of experiments. PRADS 2016 - Proceedings of the 13th International Symposium on PRActical Design of Ships and Other Floating Structures. (September).

Venkateswarlu, Ch. and Jujjavarapu, S.E. 2020. Chapter 2 - Classical analytical methods of optimization In: Ch. Venkateswarlu and S. E. Jujjavarapu, eds. Stochastic Global Optimization Methods and Applications to Chemical, Biochemical, Pharmaceutical and Environmental Processes. Elsevier, pp.27–52.

Versteeg, H.K. 2007. An introduction to computational fluid dynamics : the finite volume method Second edi. Harlow: Pearson Prentice Hall.

Versteeg, H.K. and Malalasekara, W. n.d. An Introduction to Computational Fluid Dynamics.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. Pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D. V, Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y. and Contributors, S. 1. 0 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 17(3), pp.261–272.

de Waard, C., Craig, B.D. and Smith, L.M. 2003. The influence of crude oils on well tubing corrosion rates. NACE - International Corrosion Conference Series. 2003-April(January 2003).

de Waard, C., Lotz, U. and Dugstad, A. 1995. Influence of liquid flow velocity on CO_2 corrosion: A semi-empirical model. NACE International. Paper No. 128.

de Waard, C., Lotz, U. and Milliams, D.E. 1991. Predictive model for CO_2 corrosion engineering in wet natural gas pipelines. Corrosion. 47(12), pp.976–985.

de Waard, C. and Milliams, D.E. 1975. Carbonic acid corrosion of steel. Corrosion. 31(5), pp.177–181.

Wang, C., Li, W., Xin, G., Wang, Y., Xu, S. and Fan, M. 2022. Novel method for prediction of corrosion current density of gas pipeline steel under stray current interference based on hybrid LWQPSO-NN model. Measurement: Journal of the International Measurement Confederation. 200(April), p.111592.

Wang, F. 1999. Modelling of aqueous CO_2 corrosion of iron in turbulent pipe flow. University of Saskatchewan.

Wang, K., Li, C., Li, Y., Lu, J., Wang, Y. and Luo, X. 2020. A fully coupled model of hydrodynamic-chemical-electrochemical processes for CO_2 uniform corrosion in multi-physics environment. Journal of Petroleum Science and Engineering. 193(May).

Wang, S. and Nesic, S. 2003. On Coupling CO_2 Corrosion and Multiphase Flow Models, NACE Corrosion 2003, NACE, Houston TX, Paper 3631

Wasserman, P.D. 1993. Advanced methods in neural computing. New York: Van Nostrand Reinhold.

Weckman, G., Young, W., Hernández, S., Rangwala, M. and Ghai, V. 2010. Extracting knowledge from carbon dioxide corrosion inhibition with artificial neural networks. International Journal of Industrial Engineering : Theory Applications and Practice. 17(1), pp.69–79.

Wen, K., He, L., Liu, J. and Gong, J. 2019. An optimization of artificial neural network modeling methodology for the reliability assessment of corroding natural gas pipelines. Journal of Loss Prevention in the Process Industries. 60(April), pp.1–8.

White, H. 1992. Artificial neural networks: approximation and learning theory. Blackwell Publishers, Inc.

Wilcox, D.C. 1998. Turbulence modeling for CFD. DCW industries La Canada, CA.

William L. Oberkampf, Timothy G. Trucano, Verification and validation in computational fluid dynamics, Progress in Aerospace Sciences, Volume 38, Issue 3, 2002, Pages 209-272, ISSN 0376-0421, https://doi.org/10.1016/S0376-0421(02)00005-2.

Wójtowicz-Wróbel, A., Kania, O., Kocewiak, K., Wójtowicz, R., Dzierwa, P. and Trojan, M. 2023. Thermal-flow calculations for a thermal waste treatment plant and CFD modelling of the spread of gases in the context of urban structures. Energy. 263(January 2022), p.125952.

Xin, J., Chen, J., Li, C., Lu, R. kun, Li, X., Wang, C., Zhu, H. and He, R. 2022. Deformation characterization of oil and gas pipeline by ACM technique based on SSA-BP neural network model. Measurement: Journal of the International Measurement Confederation. 189(September 2021), p.110654.

Xu, L., Wang, Y., Mo, L., Tang, Y., Wang, F. and Li, C. 2023. The research progress and prospect of data mining methods on corrosion prediction of oil and gas pipelines. Engineering Failure Analysis. 144(June 2022), p.106951.

Yang, H., Lu, L. and Tsai, K. 2020. Machine learning based predictive models for CO_2 corrosion in pipelines with various bending angles. Proceedings - SPE Annual Technical Conference and Exhibition. 2020-Octob.

Yeoh, G.H. and Tu, J. 2010a. Governing Equations and Boundary Conditions. Computational Techniques for Multiphase Flows., pp.21–94.

Yeoh, G.H. and Tu, J. 2010b. Introduction. Computational Techniques for Multiphase Flows., pp.1–20.

Yeoh, G.H. and Tu, J. 2010c. Solution Methods for Multi-Phase Flows. Computational Techniques for Multiphase Flows., pp.95–242.

Yondo, R., Andrés, E. and Valero, E. 2018. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. Progress in Aerospace Sciences. 96(March 2017), pp.23–61.

Zahedi, P., Parvandeh, S., Asgharpour, A., McLaury, B.S., Shirazi, S.A. and McKinney, B.A. 2018. Random forest regression prediction of solid particle Erosion in elbows. Powder Technology. 338, pp.983–992.

Zahura, F.T. and Goodall, J.L. 2022. Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model. Journal of Hydrology: Regional Studies. 41(December 2021), p.101087.

Zeng, L., Lv, T., Chen, H., Ma, T., Fang, Z., & Shi, J. (2023). Flow accelerated corrosion of X65 steel gradual contraction pipe in high CO_2 partial pressure environments. Arabian Journal of Chemistry, 16(8). https://doi.org/10.1016/j.arabjc.2023.104935

Zhang, G.A. and Cheng, Y.F. 2010. Electrochemical characterization and computational fluid dynamics simulation of flow-accelerated corrosion of X65 steel in a CO_2 - saturated oilfield formation water. Corrosion Science. 52(8), pp.2716–2724.

Zhang, Q., Yu, H., Barbiero, M., Wang, B. and Gu, M. 2019. Artificial neural networks enabled by nanophotonics. Light: Science and Applications. 8(1).

Zhang, S., Lu, L., Yang, H., Tsai, K. and Sidahmed, M. 2022. Accelerating Pipeline Corrosion Modeling via Bayesian Active Learning. Proceedings - SPE Annual Technical Conference and Exhibition (2009).

Zhang, Y., Gao, K., Schmitt, G. and Hausler, R.H. 2013. Modeling steel corrosion under supercritical CO_2 conditions. Materials and Corrosion. 64(6), pp.478–485.

Zheng, D., He, X. and Che, D. 2007. CFD simulations of hydrodynamic characteristics in a gas-liquid vertical upward slug flow. International Journal of Heat and Mass Transfer. 50(21–22), pp.4151–4165.

Zheng, Y. 2015. Electrochemical mechanism and model of H_2S corrosion of carbon steel. Ohio University.

Zhong, X., Shang, T., Zhang, C., Hu, J., Zhang, Z., Zhang, Q., Yuan, X., Hou, D., Zeng, D., & Shi, T. (2020). In situ study of flow accelerated corrosion and its mitigation at different locations of a gradual contraction of N80 steel. Journal of Alloys and Compounds, 824, 153947. https://doi.org/https://doi.org/10.1016/j.jallcom.2020.153947

Appendix

The numerical model developed in Chapter 3 to predict CO_2 corrosion rates using CFD driven mass transfer model is uploaded to GitHub repository. This repository can be accessed using the following link:

https://github.com/udayrajthorat/CFD-driven-mass-transfer-modelling-of-CO2-corrosion.git

The machine learning models developed in Chapter 4 to predict CO_2 corrosion rates for input parameter such as pH, pCO_2, velocity and temperature can be accessed using the following link:

https://github.com/udayrajthorat/machine_learning_modelling_CO2_corrosion.git

The codes used for the surrogate modelling based optimization of electrochemical reaction rate constants in Chapter 5 can be accessed using the following link:

https://github.com/udayrajthorat/machine_learning_enabled_parameter_optimization_CO2_corrosion_model.git

The codes used for CO_2 corrosion rate predictions in complex flow geometries in Chapter 6 can be accessed using the following link:

https://github.com/udayrajthorat/CFD_driven_mass_transfer_model_in_complex_flow_geometries.git