OPTIMISATION OF PRODUCED WATER DEOILING IN HYDROCYCLONE-INDUCED GAS FLOTATION SYSTEM: A CASE OF MACHINE LEARNING

by

SANDRO DUARTE CÉSAR

Thesis submitted in fulfilment of the requirements for the degree

Doctor of Engineering: Chemical Engineering
in the Faculty of Engineering and the Built Environment

at the

CAPE PENINSULA UNIVERSITY OF TECHNOLOGY

Supervisor: Dr. Debbie de Jager
External Supervisor: Dr. Mahomet Njoya

Bellville, South Africa
December 2025

DECLARATION

I, **SANDRO CÉSAR**, declare that all the work in this thesis, save for the ones appropriately acknowledged, represents my own unaided work, apart from the normal guidance of my supervisors. This thesis represents my own opinions and not necessarily those of the Cape Peninsula University of Technology and its sponsors.

This thesis has not been submitted for any degree or examination in any other University.

____

(Signature)

Signed on this 12th day of July 2025

ABSTRACT

Produced water represents the largest liquid waste stream generated by offshore oil and gas operations, often exceeding hydrocarbon output in mature fields. Its complex composition, including dispersed and emulsified oils, creates significant environmental and operational challenges. To comply with stringent discharge standards, offshore platforms employ multistage treatment systems such as Hydrocyclones and Induced Gas Flotation (IGF) units. However, their performance is highly sensitive to fluctuating inlet conditions, equipment aging, and chemical dosing variability, making consistent regulatory compliance difficult.

This study developed an integrated, Machine Learning–aided digital framework to model, predict, optimise and intelligently control offshore produced water deoiling systems, focusing on Hydrocyclone–Induced Gas Flotation (HIGF) units. Using operational data from offshore facilities, ensemble learning models, particularly Extreme Gradient Boosting (XGBoost) and Gradient Boosting, effectively captured nonlinear process dynamics, achieving predictive accuracies of $R^2 = 0.78$ for the Hydrocyclone and $R^2 = 0.59$ for the IGF, with corresponding RMSE values of 5.7 ppm and 3.9 ppm. These models identified key operational drivers, including inlet oil-in-water concentration, clarifier dosage and differential pressures, providing actionable insights for optimisation.

By integrating the predictive models within a Monte Carlo simulation framework, the research quantified compliance risk and performance variability under uncertain conditions. The conceptualisation of a digital twin further demonstrated the feasibility of real-time monitoring, adaptive optimisation and control under dynamic offshore environments. Overall, this work provides a validated foundation for intelligent, data-driven produced water management, advancing environmental compliance, operational efficiency and sustainability in offshore oil and gas production.

DEDICATION

Blessed are those who hunger and thirst for
righteousness, for they shall be filled

Jesus Christ (4-6 BC – AD 30-33)
The Only Begotten Son of God

To my precious daughter, Alanis Uakaiela, whose laughter brightens my darkest days and whose love inspires me to reach higher every day.

May you always dream big, stay curious and know that you are deeply loved.

A man without a cause is a man already defeated. We chose the hardest path because it is the only honest path

Jonas Malheiro Sidónio Sakaita Savimbi (1934 – 2002)
Revolutionary Leader for Angolan Independence

Time is a mind construct. It’s not real. Sometimes it takes years for a person to become an overnight success

Prince Rogers Nelson O(+> (1958 – 2016)
Musician and Cultural Icon

LIST OF OUTPUTS

César, S.D., De Jager, D. and Njoya, M., 2024. Environmental trade-offs in energy production: A review of the produced water life cycle and environmental footprint. *Marine Pollution Bulletin*, 203, p.116480

César, S.D., De Jager, D. and Njoya, M., 2025. The Role of Hydrocyclone and Induced Gas Flotation Technologies in Offshore Produced Water Deoiling Advancements. *Petroleum Research*, 10(2), pp.342-351.

César, S.D., De Jager, D. and Njoya, M., 2025. Physicochemical Characterisation of Oilfield Produced Water and its Treatment Constraints. *Petroleum Research.* (Article in Press)

* César, S.D., De Jager, D. and Njoya, M., 2025. Ensemble Learning Models for Predicting Oil-In-Water Outlet Concentration in Offshore Produced Water Deoiling. *Engineering Applications of Artificial Intelligence*

* César, S.D., De Jager, D. and Njoya, M., 2025. Machine Learning-aided Monte Carlo Simulation for Compliance Risk and Performance Assessment in Offshore Produced Water Deoiling. *Journal of Cleaner Production*

* *Under Review*

ACKNOWLEDGEMENTS

First and foremost, I wholeheartedly thank God for granting me strength, wisdom and unwavering perseverance throughout this journey. Without His blessings and grace, none of this would have been possible.

I would like to express my deepest gratitude to my supervisor, Dr. Debbie de Jager, for the constant attention, dedicated guidance and invaluable time shared during this journey. Your expertise, patience and support were fundamental to the development and success of this work.

To my co-supervisor, Dr. Mahomet Njoya, without whom this research would not have been possible. My sincere gratitude for your exceptional guidance, steadfast support, brilliant ideas and all the technical advice that profoundly enriched this study. Your dedication has been truly inspiring.

To my beloved wife, Elielsa César, thank you for your unconditional love, patience and endless encouragement. Your faith in me gave me the strength to persevere through every challenge and celebrate every milestone.

Finally, to my precious daughter, Alanis Uakaiela, your joy, laughter and boundless light have been my greatest inspiration and motivation every single day. You remind me of the true meaning of dedication and love and this achievement is as much yours as it is mine.

PREFACE

This thesis examines the critical challenge of treating offshore produced water, with a focus on developing and integrating advanced digital, data-driven and control-oriented frameworks to enhance deoiling performance and ensure regulatory compliance. The study presents a comprehensive investigation into physicochemical characterisation, performance analysis, anomaly detection, Machine Learning modeling, probabilistic risk assessment and digital twin implementation for Hydrocyclone–Induced Gas Flotation (HIGF) systems.

Chapter 1 provides the background to the research problem, outlines the motivation and rationale behind the study, defines the hypothesis, articulates the research aims and objectives, presents the research questions and delineates the scope and significance of the work.

Chapter 2 describes the life cycle of hydrocarbons and the generation of produced water, highlighting the environmental and operational issues posed by its complex composition. This chapter also explains the importance of treatment to meet discharge regulations and discusses the characteristics and impacts of produced water in offshore oil and gas operations.

Chapter 3 offers a critical technical review of produced water deoiling systems, focusing on Hydrocyclones and Induced Gas Flotation units. It discusses their operational principles, design challenges, global performance criteria, and future research needs, setting the stage for developing advanced control and optimisation approaches.

Chapter 4 provides a detailed physicochemical characterisation of offshore produced water, evaluating variations in properties such as salinity, oil-in-water concentration and chemical additives, and discusses their impacts on treatment efficiency and system performance.

Chapter 5 introduces the application of anomaly detection algorithms to offshore produced water deoiling systems. It describes methods for identifying operational deviations and faults in real-time, supporting early warning systems and enhancing operational resilience.

Chapter 6 presents a bibliometric analysis of machine learning applications in effluent quality prediction in wastewater treatment, outlining publication trends, emerging models and global research directions over the last decade. This sets a contextual basis for selecting and justifying ensemble-based machine learning models for offshore produced water applications.

Chapter 7 develops and evaluates ensemble-based Machine Learning models, including XGBoost, Gradient Boosting and Random Forest to predict oil-in-water outlet concentrations after Hydrocyclone and Induced Gas Flotation systems. The performance of these models is compared using statistical metrics and key operational features influencing treatment outcomes are identified.

Chapter 8 integrates the developed XGBoost Machine Learning model into a Monte Carlo simulation framework, enabling compliance risk assessment, exceedance probability forecasting and explicit quantification of operational uncertainties. This chapter emphasises the importance of probabilistic modeling in supporting proactive decision-making.

Chapter 9 proposes and validates a conceptual digital twin architecture for offshore produced water deoiling. The design integrates predictive analytics, probabilistic simulation, real-time optimisation and Model Predictive Control–inspired logic, demonstrating potential for dynamic recalibration and adaptive control in fluctuating offshore environments.

Chapter 10 concludes the thesis by summarizing key findings, directly addressing the research questions and offering actionable recommendations for future research and industrial applications. It also provides final reflections on the strategic value of integrating Machine Learning and digital twin technologies to transform offshore produced water deoiling into a more intelligent, predictive and environmentally responsible operation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BOD	Biochemical Oxygen Demand
BS&W	Basic Sediment and Water
BTEX	Benzene, Toluene, Ethylbenzene and Xylene
CFD	Computational Fluid Dynamics
COD	Chemical Oxygen Demand
DCS	Distributed Control System
DP	Differential Pressure
GB	Gradient Boosting
GBDT	Gradient Boosting Decision Trees
HIGF	Hydrocyclone–Induced Gas Flotation
IGF	Induced Gas Flotation
IQR	Interquartile Range
KDE	Kernel Density Estimate
L-BFGS-B	Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bound constraints
LOWESS	Locally Weighted Scatterplot Smoothing
MAE	Mean Absolute Error
MCS	Monte Carlo Simulation
ML	Machine Learning
MPC	Model Predictive Control
OIW	Oil-in-Water
OPC-UA	Open Platform Communications Unified Architecture
PAH	Polycyclic Aromatic Hydrocarbons
PCV	Pressure Control Valve
PDC	Pressure Differential Controller
PDCV	Pressure Differential Control Valve
PDP	Partial Dependence Plot
PDR	Pressure Drop Ratio
PDT	Pressure Differential Transmitter
PID	Proportional-Integral-Derivative

PW	Produced Water
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
SCADA	Supervisory Control and Data Acquisition
SHAP	SHapley Additive exPlanations
TDS	Total Disolved Solids
TOC	Total Organic Carbon
TPH	Total Petroleum hydrocarbons
TSS	Total Suspended Solids
XGBoost	Extreme Gradient Boosting

LIST OF SYMBOLS

$\boldsymbol{g_i}$	First-order gradient loss function
$\boldsymbol{h_i}$	Second-order gradient loss function
$\boldsymbol{f_t(x_i)}$	Prediction from tree t for instance i
$\boldsymbol{h_m(x_i)}$	Prediction from weak learner m for instance i
$\boldsymbol{\Omega(f_t)}$	Regularisation term for tree t
$\boldsymbol{y_i}$	Observed value for sample i
$\boldsymbol{\hat{y}_i}$	Predicted value for sample i
$\boldsymbol{L}$	Loss function
$\boldsymbol{M}$	Number of boosting iterations (tree)
$\boldsymbol{\gamma_m}$	Learning rate coefficient for tree m
$\boldsymbol{\sigma}$	Standard deviation
$\boldsymbol{\eta}$	Number of data points (sample)
$\boldsymbol{\mu}$	Mean value

GLOSSARY

Biodegradability Index: A metric that assesses the potential of a substance or wastewater to be broken down by biological processes, commonly expressed as the ratio of BOD_5 to COD (Lacalamita *et al.*, 2024; Choi *et al.*, 2017)

Biochemical Oxygen Demand (COD): The proportion of dissolved oxygen required for microorganisms to decompose organic matter in an aerobic environment at a certain temperature and time frame (Abbas *et al.*, 2021; Dawoud *et al.*, 2021).

Chemical Oxygen Demand (COD): The quantity of oxygen to be used during the chemical degradation of both organic and inorganic molecules (Amakiri, *et al.*, 2022; Abbas *et al.*, 2021).

Extreme Gradient Boosting (XGBoost): An efficient, scalable implementation of gradient boosting algorithms designed for high predictive accuracy and performance in machine learning tasks (Chen and Guestrin, 2016)

Floating Production Storage and Offloading (FPSO) Unit: A seabed-moored vessel designed to process, store and offload produced hydrocarbons (Leffler *et al.*, 2011).

Gradient Boosting: A machine learning technique that builds models sequentially, where each new model corrects the errors of the previous ones, resulting in strong predictive performance (Aziz *et al.*, 2020; Si and Du, 2020).

Hydrocarbon: An organic molecule consisting exclusively of hydrogen and carbon atoms (César, 2022)

Hydrocyclone: A high-speed separation device that removes dispersed oil droplets from produced water by applying centrifugal forces generated by a spinning flow (Durdevic *et al.*, 2017)

Induced Gas Flotation (IGF): A water treatment process that removes suspended oil droplets and solids by introducing fine gas bubbles, which attach to the particles and float them to the surface for removal (Amakiri *et al.*, 2022; Liu *et al.*, 2021).

Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bound constraints (L-BFGS-B): An optimization algorithm that extends the L-BFGS method to handle simple box constraints on variables, enabling efficient optimization of large-scale problems while limiting memory usage (Qi *et al.*, 2017).

Machine Learning (ML): A field of artificial intelligence that uses algorithms to learn patterns from data and make predictions or decisions without being explicitly programmed (Zhao *et al.*, 2020).

Monte Carlo Simulation: A computational technique that uses repeated random sampling to estimate the probability distribution of outcomes in a process or system (Balas, 2023).

Oil-in-Water (OIW): Also referred to as Total Oil and Grease (TOG), refers to a mixture of hydrocarbons including aliphatic and aromatic compounds, waxes, fats, and other organics found in produced water (Amakiri, *et al.*, 2022).

Produced Water (PW): Water co-extracted with oil and gas, mainly originating from natural formation fluids, often including seawater, present within deep underground reservoirs (Abbas *et al.*, 2021; Olajire, 2020; Zheng *et al.*, 2016).

Random Forest: An ensemble learning method that builds multiple decision trees and combines their outputs to improve prediction accuracy and control overfitting (Schonlau and Zou, 2020).

SHapley Additive exPlanations (SHAP): An interpretability framework based on cooperative game theory that uses Shapley values to fairly and consistently quantify each input feature's contribution to a machine learning model's prediction, providing a theoretically sound explanation of model outputs (Mosca *et al.*, 2022).

CHAPTER 1

INTRODUCTION

"Ask, and it will be given to you; seek, and you will find; knock, and it will be opened to you."
Matthew 7:7

CHAPTER 1: INTRODUCTION

1.1. Background to the Research Problem

The production of oil and gas inevitably generates substantial volumes of produced water, the largest liquid waste stream associated with hydrocarbon extraction. In mature offshore fields, the volume of produced water can surpass that of oil, creating significant operational and environmental challenges. Produced water typically contains dispersed and emulsified hydrocarbons, dissolved salts, and suspended solids, all of which must be effectively treated to prevent ecological harm (Amakiri *et al.*, 2022; Abbas *et al.*, 2021; Olajire, 2020).

To comply with increasingly stringent discharge regulations, such as the 30 ppm oil-in-water (OIW) limit commonly enforced in offshore regions including Angola, platforms employ multistage deoiling systems combining Hydrocyclones and Induced Gas Flotation (IGF) units. Hydrocyclones use centrifugal separation to remove bulk oil, while IGF units polish the effluent through gas-induced flotation. Despite their compactness and efficiency, these systems are sensitive to fluctuations in flow rate, pressure, temperature, and chemical dosing, which can degrade separation efficiency and jeopardise compliance (Vallabhan *et al.*, 2022; Piccioli *et al.*, 2020; Das and Jäschke, 2018; Durdevic *et al.*, 2017).

The growing emphasis on sustainable offshore operations and environmental stewardship has intensified the need for smarter, adaptive produced water management strategies. Machine Learning (ML) and digital technologies now offer promising pathways to model nonlinear system behaviour, detect anomalies, and enable predictive optimisation, moving beyond traditional reactive control strategies (Safeer *et al.*, 2022; Aghbashlo *et al.*, 2021; Icke *et al.*, 2020).

1.2. Statement of the Research Problem

HIGF performance depends on key parameters such as flow rate, inlet oil-in-water concentration and pressure differential. Optimisation is challenging due to the system's nonlinear and dynamic behaviour (Piccioli *et al.*, 2020; Vallabhan *et al.*, 2020; Durdevic *et al.*, 2017). CFD models replicate internal separation but are too computationally demanding for real-time use (Bram *et al.*, 2018). In practice, control relies on indirect indicators like pressure-drop ratio (PDR) instead of direct oil-in-water (OIW) readings, reducing responsiveness to changing conditions (Durdevic *et al.*, 2018). Conventional control schemes are also limited in capturing the complex, nonlinear interactions among variables and often rely on oversimplified assumptions that fail to accurately reflect real operational conditions (Aghbashlo *et al.*, 2021; Durdevic *et al.*, 2017).

Despite prior studies and deployment of advanced treatment technologies, a critical gap remains: few data-driven, control-oriented models can predict and optimise HIGF performance under full-scale offshore conditions. Existing work focuses on hydrodynamics or lab-scale trials with limited relevance for adaptive, real-time decision-making. This research fills that gap by introducing an ensemble Machine Learning framework integrating probabilistic risk analysis and digital-twin concepts for predictive optimisation and compliance assurance (Aghbashlo *et al.*, 2021; Bram *et al.*, 2018; Durdevic *et al.*, 2017).

1.3. Motivation for the Research Problem

Although conventional process controls remain essential, there is a pressing need for control-oriented ML models that incorporate reliable flow mechanics and can dynamically adjust to various operational scenarios. Developing a comprehensive, real-time, data-driven model for HIGF systems in full-scale industrial applications would enable improved operational efficiency and environmental compliance (Bram *et al.*, 2020; Das and Jäschke, 2018; Durdevic *et al.*, 2017).

The increasing availability of large-scale process data through advanced sensing and digitalisation makes ML approaches particularly timely and impactful in Process Systems Engineering (PSE). ML can leverage historical and real-time data to predict system behavior, model and optimise parameters, support early fault detection and enhance overall process efficieny and safety (Schweidtmann *et al.*, 2021; Lee *et al.*, 2018).

1.4. Research Rationale

ML algorithms have shown great promise in modeling complex systems, predicting performance and optimising operations. However, their application in modeling and optimising deoiling within HIGF systems remains limited. Addressing this gap offers significant potential for advancing process control, improving operational efficiency and ensuring regulatory compliance in offshore produced water treatment.

1.5. Hypothesis

By integrating real-time monitoring and feedback into a ML–based model, it is possible to dynamically adjust operational parameters in response to changing conditions, resulting in a more efficient deoiling process in HIGF systems. Such a model would provide deeper insights into the effects of each operational parameter, enabling process control designers to improve existing automatic and/or develop more effective control schemes.

1.6. Research Aim and Objectives

The main aim of this research is to develop an integrated Machine Learning–aided digital framework for modeling, predicting, optimising and controlling offshore produced water deoiling systems, with a focus on the Hydrocyclone–Induced Gas Flotation (HIGF) units, to enhance treatment performance and ensure regulatory compliance.

Objectives:

- To characterise the physicochemical properties of an offshore hydrocarbons produced water to establish a foundation for modeling and optimisation.
- To analyse the performance of an HIGF system in produced water deoiling by integrating operational data to better understand the system's behavior.
- To develop and compare the performance of three ensemble-based Machine Learning models (XGBoost, Gradient Boosting and Random Forest) for predicting oil-in-water outlet concentrations at various treatment stages using offshore operational data.
- To integrate a Machine Learning model into a Monte Carlo simulation framework to evaluate compliance risk, performance variability and probabilistic exceedance under uncertain operating conditions.
- To conceptualise and design a digital twin architecture for produced water deoiling, integrating predictive analytics, real-time optimisation and control recommendations.

1.7. Research Questions

To address the research problem and achieve the outlined objectives, this study sought to answer the following key questions:

- How do variations in physicochemical properties of produced water affect the efficiency of HIGF systems?
- Can detailed produced water characterisation be used to guide more adaptive and resilient treatment strategies?
- How can anomaly detection algorithms be applied to identify performance degradation in HIGF systems in real time?
- Can data-driven anomaly detection improve early warning capabilities and reduce downtime in offshore produced water deoiling?
- Which ensemble Machine Learning algorithm provides the highest accuracy and robustness for predicting oil-in-water concentrations in offshore produced water deoiling systems?
- What are the most influential operational features driving oil-in-water outlet concentration at each treatment stage, and how can these insights inform operational control strategies?
- How can predictive Machine Learning models be validated and trusted for deployment in real offshore facilities?
- How can Machine Learning–aided Monte Carlo simulations quantify compliance risks and forecast exceedance probabilities in offshore produced water deoiling?
- What are the benefits and limitations of integrating stochastic simulation with Machine Learning models to assess operational uncertainty and treatment performance?
- How can a digital twin architecture enhance predictive monitoring, control optimisation, and compliance assurance in offshore produced water deoiling?
- What operational advantages arise from integrating Model Predictive Control (MPC)-inspired logic within a digital twin for produced water deoiling?

1.8. Significance of the Research

This research addressed a critical environmental and operational challenge in offshore oil and gas production, namely, the efficient and compliant treatment of produced water to remove petroleum hydrocarbons. Additionally, it provided a detailed physicochemical characterisation of Angolan produced waters, for which limited data currently exists.

The integration of ML into produced water deoiling represents an innovative approach with significant potential to improve system efficiency, compliance and sustainability. By leveraging large, multi-source datasets to develop and train predictive models, this work introduced a unique methodology that advances current practice.

Furthermore, this study identified optimal operational strategies and demonstrated the value of ML in automating complex control tasks, potentially leading to significant environmental benefits through improved discharge water quality. The findings could also catalyse new developments in automatic control systems for offshore water deoiling, bridging a critical gap between environmental compliance and operational excellence.

1.9. Delineation of the Study

This study did not cover:

- Evaluating the HIGF process for other general water quality parameters beyond the effluent's oil-in-water concentrations.
- Modeling or assessing the degradation of mechanical or structural performance of system components over time.
- Extensive real-world field validation of the developed models beyond the historically collected operational data used in this research.
- Economic feasibility studies or cost-benefit analyses of implementing ML-based control in operational offshore facilities, including detailed economic savings projections or financial risk assessments.
- The feasibility or design optimisation of entirely new Hydrocyclone or IGF equipment hardware using Machine Learning.
- Real-time deployment trials onboard offshore platforms or pilot-scale experiments to validate live, dynamic performance.
- Assessment of the environmental fate, toxicity or ecological impacts of the discharged treated water beyond oil-in-water concentration metrics.
- Long-term reliability analysis or lifecycle performance prediction of the ML models under evolving operational or environmental conditions.

CHAPTER 2
HYDROCARBONS LIFE CYCLE AND PRODUCED WATER GENERATION

César, S.D., De Jager, D. and Njoya, M., 2024. Environmental trade-offs in energy production: A review of the produced water life cycle and environmental footprint. *Marine Pollution Bulletin*, 203, p.116480.

"For nothing is hidden that will not be made manifest, nor is anything secret that will not be known and come to light."

Luke 8:17

CHAPTER 2: HYDROCARBONS LIFE CYCLE AND PRODUCED WATER GENERATION

2.1. Introduction

Energy, deemed to be the pillar of modern civilisation's existence, plays an essential role in sustaining global economies and enabling countless human activities. A large portion of this energy is sourced from oil and gas, whose exploration and extraction inevitably produce significant volumes of water as a by-product, known as produced water. Produced water represents the most substantial waste stream by volume in the oil and gas industry, posing complex challenges at the intersection of environmental sustainability and energy production (Abbas *et al.*, 2021; Olajire, 2020; Zheng *et al.*, 2016).

This chapter offers a comprehensive review of the hydrocarbon lifecycle, covering the geological formation and maturation of hydrocarbons through to their extraction, processing, and separation phases. It thoroughly examined the origin, generation, and complex composition of produced water, emphasizing the significant global variations that influence its characteristics. Additionally, the chapter discusses regional regulatory frameworks designed to mitigate environmental impacts, highlighting the technological and economic challenges of standardizing treatment solutions. The chapter also critically evaluated current treatment methods, assessing their effectiveness in removing key contaminants and exploring their potential for reuse in sectors beyond oil and gas, including agriculture and various industrial applications.

2.2. Hydrocarbons Lifecycle

According to Schobert (2013), hydrocarbons are formed very slowly when buried old organic matter undergoes anaerobic pressure and temperature changes. Their origin has been traced to ancient plants that lived in the marine environment hundreds of million years ago, whose remains settled onto the seabed, accumulated on the ocean floor and were then buried under several kilometres of sand, silt and sediments (Kvenvolden, 2006). The layers of organic material were compressed beneath the weight of the sediments and together with an increase in temperature and pressure under anaerobic conditions, these sediments transformed the rock and organic matter into what is known as kerogen. After further burial and increase in temperature, kerogen is transformed into oil and gas by means of thermal cracking (Schobert, 2013; Kvenvolden, 2006; Tissot and Welte, 1984).

Tissot and Welte (1984), defined hydrocarbons as complex chemical mixtures containing different configurations of carbon and hydrogen atoms, water and gases such as carbon dioxide, nitrogen, hydrogen sulphide and helium. Hydrocarbons originate from organic matter and as stated by Schobert (2013), evidence supporting this idea includes: a) The presence of brine in the mixture's composition; b) Hydrocarbon fluids undergo a similar rotation like all organic oils when polarised light passes through; and c) The similarity between the carbon depletion of the oil and gas deposits and the organic plant matter through radioactive decay as seen in the process of photosynthesis.

2.2.1. Formation Mechanism of Hydrocarbons

The hydrocarbon's creation process is generally the same. Although there may be different types of plants dropping to the seabed that are buried under different conditions, based on the works of Overton *et al.* (2020) and Selley and Sonnenberg (2014), the following steps happen in a conventional deposit:

1. Photosynthesis produces a reduced form of carbon biomass and oxygen, which approximates to the ratio of carbon, hydrogen and oxygen in all living matter (Eq. 2.1). Dead plankton and organic tissue matter fall to the ocean floor, mix with inorganic sand and clay and create an organic rich mud.

$$\mathrm{CO_2 + H_2O + Energy \longrightarrow \{CH_2O\} + O_2} \qquad 2.1$$

2. As the mud is within an anoxic environment, the organic matter is not decomposed by bacteria as it is further buried by more sediment forming organic shales. Over thousands of years, some of this organic material is further stripped of its remaining oxygen and with an increase in temperature and pressure transforms into kerogen due to the shale's location in the interior of the earth. The kerogen transforms into complex chemical mixtures of carbon-hydrogen ratios referred to as hydrocarbons (Eq. 2.2).

$$\mathrm{\{CH_2O\} + Heat/Pressure \longrightarrow \{CH_2\,;\ CH\} + H_2O} \qquad 2.2$$

3. Since oil and gas are lighter than water, water is displaced from the oil shale source as oil and gas rise across the permeable rock pores. However, the hydrocarbon mixture remains trapped in the reservoir by an impermeable layer of rock, which acts as a sealant.

2.2.2. Maturation Process of Kerogen

Cheng *et al.* (2017), stated that the maturation of oil and gas depends on the temperature, depth and time of burial. After sedimentation, the maturation process takes place in three stages namely, diagenesis, catagenesis and metagenesis as graphically depicted in Fig. 2.1.

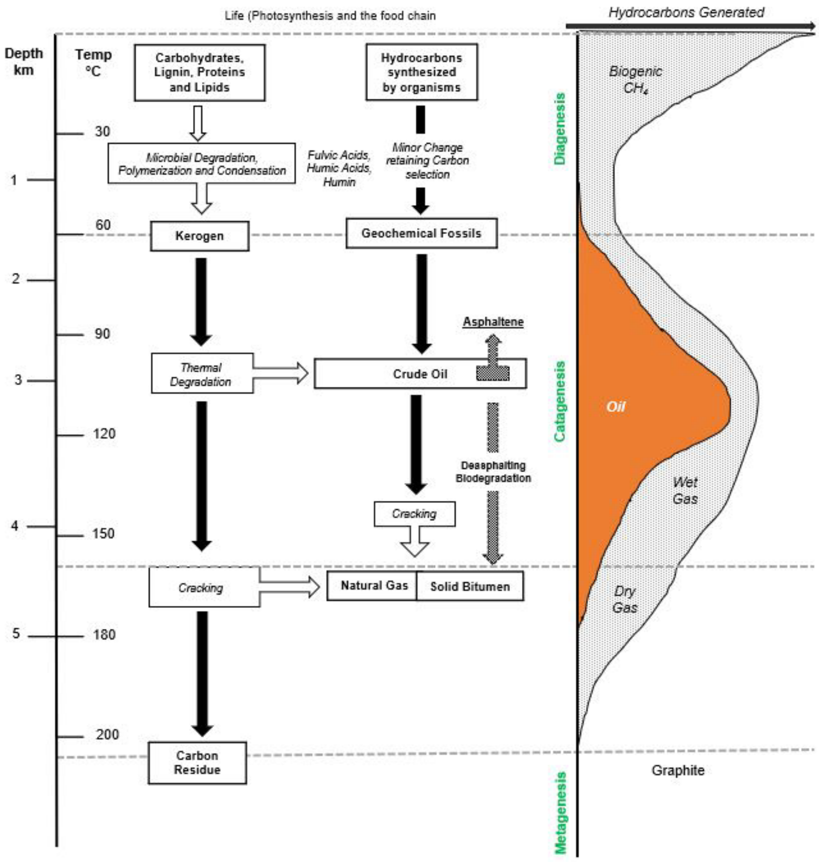


Figure 2.1*: Formation and evolution of macromolecules in geological conditions (adapted as a combination from Cheng et al., 2017; Tissot and Welte, 1984)*

Diagenesis marks the initial stage of sediment alteration where organic matter undergoes microbial degradation through bacterial oxidation and fermentation under mild temperature and pressure conditions. Water-saturated organic sediments with minerals undergo compaction, microbial activity, and reactions during burial, resulting in the expulsion of water and the formation of kerogen (Durand, 2019; Cheng *et al.*, 2017). This process begins at temperatures below 60°C in shallow surfaces, occurring up to a depth of 1 km. The type of hydrocarbon generated depends on various factors including temperature, pressure, organic matter type and time. Microbial activity under anaerobic conditions contributes to the breakdown of organic matter, yielding biogenic gas. Both biogenic decay and non-biogenic reactions facilitated by bacteria, produce methane, carbon dioxide, water and kerogen (Selley and Sonnenberg, 2014; Jahn *et al.*, 2008; Tissot and Welte, 1984). Jahn *et al.* (2008) affirm that there are four types of kerogen, characterised by the hydrogen to carbon ratio (H:C).

- Liptinite (Type I): Fine algae in freshwater lakes, rich in lipids, having the highest H:C ratio with the tendency of forming oil and low amounts of gas.
- Exinite (Type II): Single cell plankton, algae and bacteria in marine environments, having a ratio of 1H:4C, with the tendency of forming a proportional mixture of oil and gas.
- Vitrinite (Type 3): Organic material from land plants, spores and pollen, having a H:C ratio of less than 1.0, with the tendency of forming mostly dry gas.
- Inertinite (Type): Often termed as dead carbon, it is high in carbon, very low in hydrogen and has no effective potential to yield oil and gas.

Catagenesis, succeeding diagenesis, involves thermal degradation of kerogen as it is buried deeper under increased temperature and pressure without bacterial involvement. At around 60°C, kerogen molecules begin to crack, giving rise to oil formation. With increased burial depth, higher temperatures lead to the cracking of oil molecules, transforming heavy oil into lighter oil or gas within the temperature range of 60 to 150°C (Tissot and Welte, 1984). The extent of catagenesis determines the resulting hydrocarbon products, as higher temperature and pressure results in more comprehensive kerogen cracking and the formation of lighter and smaller hydrocarbons (Selley and Sonnenberg, 2014).

Metagenesis, the final stage of organic matter maturation, occurs at elevated pressures and temperatures ranging from 150 to 200°C. During this phase, oil molecules are further cracked into smaller gas molecules, which eventually break down into carbon, manifesting as graphite (Cheng *et al.*, 2017).

2.3. Hydrocarbon Extraction

Hydrocarbons are extracted through a process known as drilling, which involves creating a well in the earth's surface and utilising specialised equipment to retrieve oil and gas from underground reservoirs (Ali, 2017). As described by Tudorache *et al.*, 2020, this procedure employs a drilling rig to bore into the earth using a rotating drill bit powered by robust motors. As the drill bit penetrates the rock, a drilling fluid (also referred to as mud) is injected into the well, serving to both lubricate the bit and carry away rock cuttings. Once the well reaches the desired depth, casing is inserted to reinforce the well and prevent collapse (Epelle and Gerogiorgis, 2020; Tudorache *et al.*, 2020, Ali, 2017). Subsequently, cement is injected into the space between the casing and the rock formation to seal the wellbore, preventing any fluid migration between distinct rock layers. Hydrocarbons can now be extracted from the reservoir using various techniques, including hydraulic fracturing and artificial lift methods (Epelle and Gerogiorgis, 2020).

There are several types of drilling techniques used in the oil and gas industry, designed to meet specific needs and challenges as summarised in Fig. 2.2. The selection of the drilling technique depends on several factors, including the location of the reservoir, the type of rock formation and the desired production rate (Epelle and Gerogiorgis, 2020).

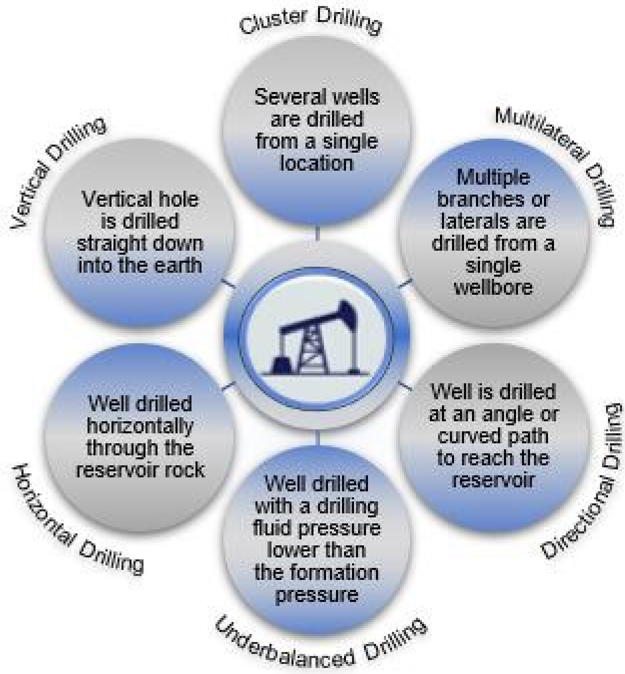


***Figure 2.2**: Common drilling techniques widely used in the oil and gas industry (adapted from Epelle and Gerogiorgis, 2020)*

2.3.1. Hydrocarbons Phase Separation

The processes and systems involved in separating hydrocarbons into phases, namely, oil, gas and produced water, are highly complex, capital-intensive and require state-of-the-art technology (César, 2022; Gou *et al.*, 2011). According to Lyons *et al.* (2015), a typical hydrocarbon fluid processing system comprises the following steps: phase separation, oil stabilisation, gas treatment and produced water treatment. Fig. 2.3 is a representation of a conventional offshore system used for processing hydrocarbons once brought to surface.

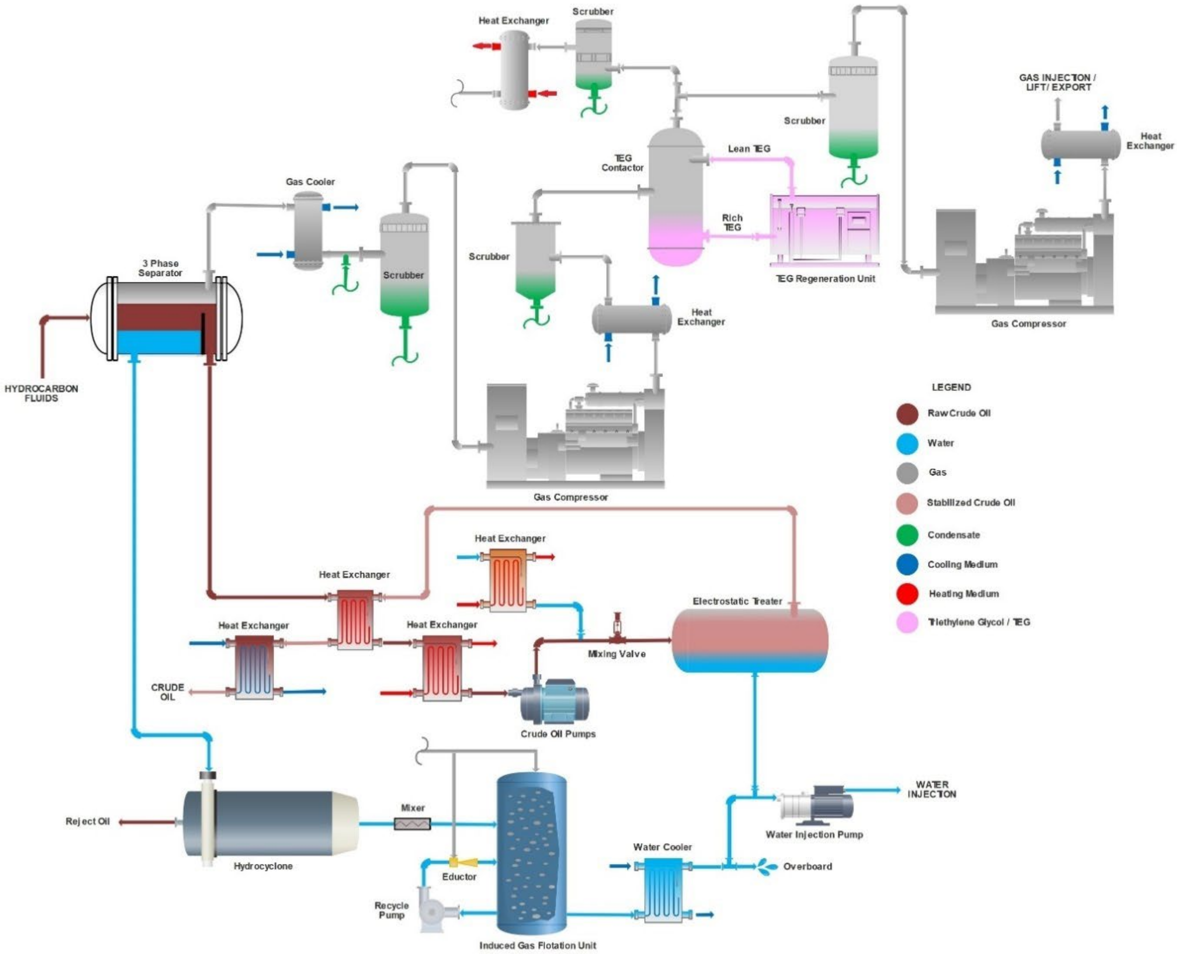


***Figure 2.3**: Conventional offshore oil and gas treatment process (adapted from César, 2022)*

The separation of well stream gas from free liquids is the first and most important stage of field-processing operations (César, 2022; Gou *et al.*, 2011). In offshore environments, phase separation is normally achieved by centrifugal separation and usually takes place in different pressure phases, giving a working volume for crude oil, water and gas separation (Guo *et al.*, 2011; Stewart and Arnold, 2008). Crude oil is stabilised by means of washing with heated fresh water injected directly into the oil stream for salt dilution before it is routed to an electrostatic treater for final dewatering. Water droplets are electrically charged for elongation and polarisation, acquiring a positive charge on one side and a negative charge on the other. The alternating current on the treater's electrical grid causes polarity inversion, resulting in water droplet movement and collisions with each other with enough force to destroy the thin layer that surrounds them, allowing them to congregate into larger droplets and settle to the bottom of the treating section for removal, while the oil rises to the top and is routed to storage (Ambrosio, 2014; Stewart and Arnold, 2008).

Gas extracted is treated by means of dehydration and compression for use as: 1) fuel gas for inhouse power generation; 2) injection gas to maintain reservoir pressure and to divert oil from injection to production wells; 3) gas lift, which is an enhanced-oil-recovery method, where oil is artificially lifted from wells when reservoir pressure is low; and 4) export gas for further treatment into domestic gas at dedicated facilities (Lyons *et al.,* 2015). In addition, unused hydrocarbon gas and condensate from process trains and utilities are disposed into the atmosphere by means of ignition, commonly referred flaring (César, 2022; Gou *et al.*, 2011).

Produced water recovered from the separators are treated for solid and oil particles to achieve the desired oil in water and total suspended solids content at an acceptable temperature for overboard discharge or reservoir injection to increase oil recovery (César, 2022).

2.4. Produced Water

According to the works of Nasiri and Jafari (2017) and Zheng *et al.* (2016), produced water is an integral part of the hydrocarbon recovery processes and consists mostly of naturally occurring water from hydrocarbon deposits that is often seawater bound with oil and gas in a deep-water reservoir (Abbas *et al*., 2021; Olajire, 2020; Zheng *et al.* 2016). According to Hedar (2018), produced water can originate from saline and formation water. Saline water has its sources from inflow within, above or below the hydrocarbon zone, as well as injected additives used to aid hydrocarbon extraction. Whereas formation water becomes produced water when it mixes with hydrocarbons and reaches the surface.

Over the course of a field's lifetime, the volume of produced water, the contaminants it contains and their concentrations change dramatically (Zheng *et al.*, 2016). In general, the proportion of produced water from gas fields is lower than that from oil fields because no water injection is utilised and acidity is higher given that it is solely a mixture of formation and condensed water (Fakhru'l-Razi *et al.*, 2009). Based on Abbas *et al.* (2021), the water generation rate can be a very small fraction of the oil production rate during the early years of field operations and usually increases as oil and gas production decreases, reaching up to 98% of the overall hydrocarbon fluids in nearly depleted fields.

According to McCabe (2013), the ratio of water-to-oil can be above 3:1 on average. In essence, the age of mature fields drives produced water generation, whereas better management practices and the development of new oil wells decreases produced water generation (Amakiri *et al.*, 2022).

2.4.1. Produced Water Generation

Fig. 2.4 represents the annual produced water production from various global regions over a six-year period, spanning from 2017 to 2022. Notably, North America emerges as the leading producer of produced water, with a substantial output that increased gradually from 43.8 billion barrels (bbl.) in 2017 to 55.5 billion bbl. in 2022, underscoring the region's significant reliance on oil and gas extraction. Meanwhile, East Asia/Pacific, Middle East/Africa and Latin America/Caribbean have exhibited relatively stable produced water generation levels over the years, with only slight fluctuations. In contrast, Europe/Central Asia shows a consistent rise in production, reaching 40.5 billion bbl. in 2022, supporting the lack of new developments unlike Latin America for example. Furthermore, Southern Asia maintains a relatively low production level of produced water, suggesting a less intensive presence of the hydrocarbon industry in this area. These trends offer valuable insights into regional variations in produced water production, which are directly linked to oil and gas extraction activities (Amakiri *et al.*, 2022).

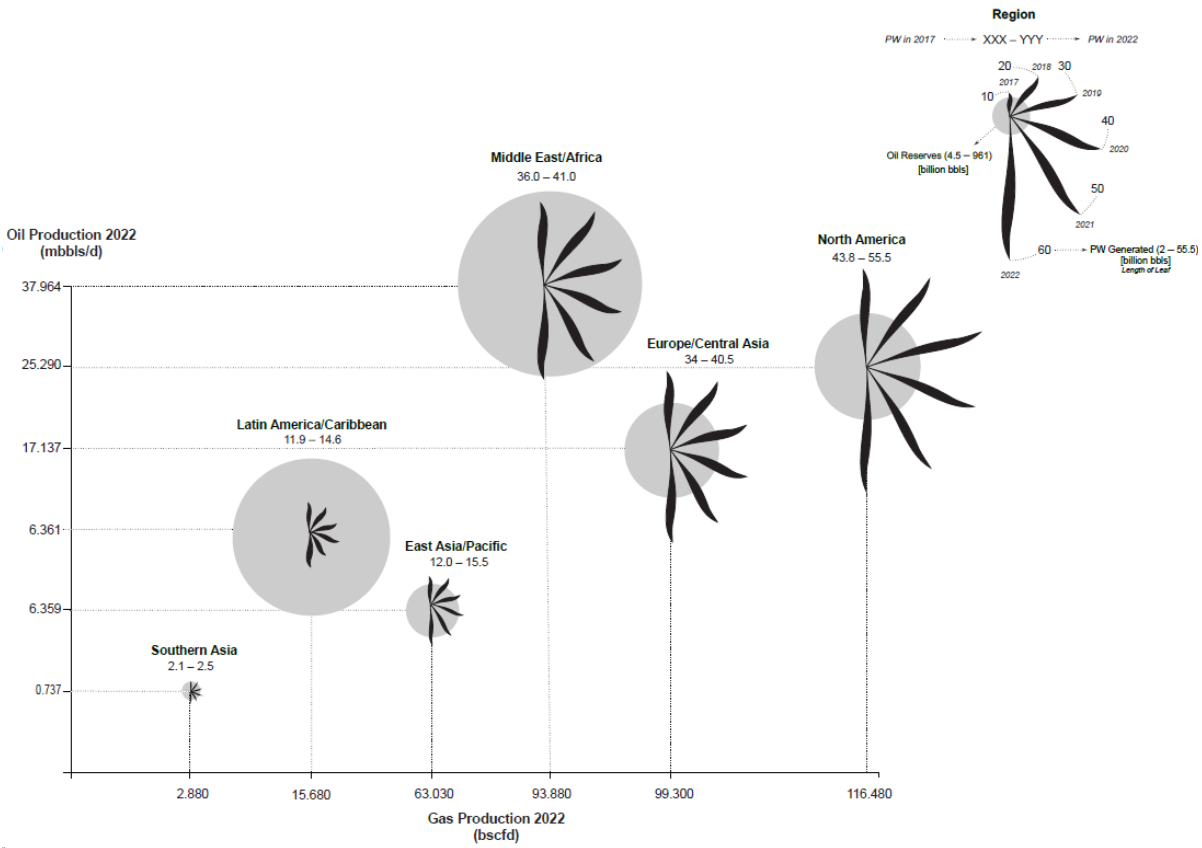


***Figure 2.4**: Produced water volume by area (adapted from Energy Institute., 2023; OPEC, 2023; Amakiri et al., 2022)*

Referring to the global outlook of produced water generation from 2012 to 2022 as shown in Fig. 2.5, beginning at 110 billion bbl. in 2012, the volume is steadily ascending, illustrating a consistent upward trajectory over the subsequent years, reaching a maximum volume of 170 billion bbl. in 2022 (Amakiri *et al.*, 2023). These trends are consistent with the maturation of oil fields and a peak in oil production worldwide, reflective of heightened industry activity and improved extraction and separation technologies (Energy Institute., 2023; OPEC, 2023; Fakhru'l-Razi *et al.*, 2009).

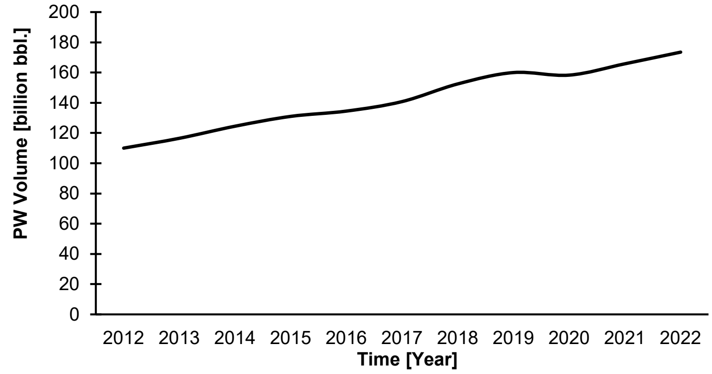


Figure 2.5: Oil field produced water volumes from 2012 to 2022 (Amakiri *et al.*, 2023)

According to the latest projections from the Produced Water Society, until 2030 a surge is anticipated in the produced water volume emanating from oil fields worldwide (Amakiri *et al.*, 2023). The projected dataset in Fig. 2.6, indicates an estimate of annual changes and the respective percentage changes from 2023 to 2030. Notably, there is a consistent positive trend in the absolute changes, ranging from 2 billion bbl. (2028 to 2029) to a substantial 20 billion bbl. (2029 to 2030). The percentage changes, which reflect the proportional increase relative to the previous year, exhibits some variability. While most years demonstrate modest percentage changes in the range of 3.6% to 4.3%. The year 2029 stands out with a minimal 0.9% increase and 2030 with a maximum of 9% (Amakiri *et al.*, 2023; Amakiri *et al.*, 2022).

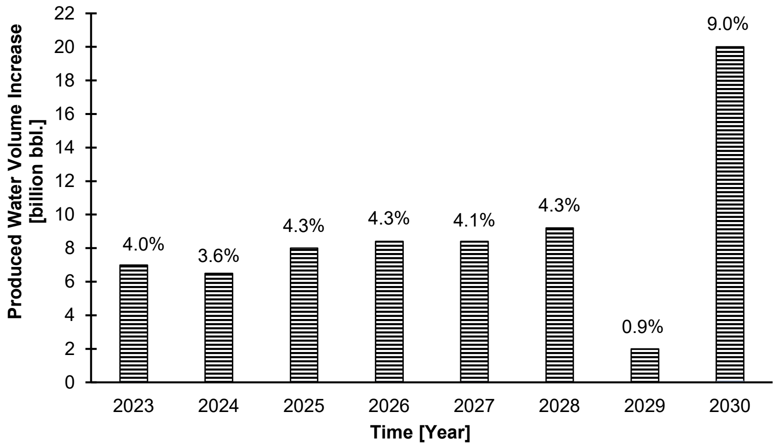


Figure 2.6: *Oil field produced water volume growth estimation from 2023 to 2030 (Amakiri et al., 2023)*

2.4.2. Factors Affecting the Volume of Produced Water

According to Dawoud *et al.* (2021), several factors have the capability to affect the volume of produced water throughout the lifespan of an oil or gas producing well, as listed, and described in Table 2.1

Table 2.1: *Factors affecting the amount of produced water generated (adapted* from Dawoud *et al.*, 2021*)*

Factor	Overview
Method of drilling	A vertical well can produce at a lower or higher rate than a horizontal well at similar drawdown.
Location of well within the reservoir	An incorrectly drilled or poorly oriented well inside the reservoir can result in earlier than anticipated water production. Furthermore, the use of horizontal or vertical wells can reduce water production depending on whether the reservoir is homogeneous or heterogeneous.
Type of well completion	The open hole technique enables for the testing of drilling zones while the perforated completion technique delivers a higher degree of control because the interval may be perforated and tested.
Single and commingled zones	Most wells are completed initially in a single zone. As the oil rate drops owing to well maturation, other zones may also be opened to sustain the oil yield, resulting in increased volumes of produced water.
Method of water separation technique	For wells that produce huge amounts of saline water, various strategies are employed to lower the costs of lifting and/or water handling.
Volume of water injection or water flooding	Water flooding produces an increased amount of water and as the flood continues, the volume of water required for injection grows. Therefore, makeup water with the appropriate chemical properties is required.
Poor mechanical integrity	Mechanical issues of the casting holes due to corrosion or wear, splits created by flows and high pressure can allow undesirable reservoir fluids to enter the casing and boost water production.

2.4.3. Composition of Hydrocarbon Produced Water

The physical and chemical properties of hydrocarbon produced water are complex, site-specific and vary with geological formation, geological age, the history of the reservoir, oil and water ratio, pH, the chemistry of the oil and water (both in-situ and injected), the classification and composition of the crude oil dominant hydrocarbon, the interactions of the rock and fluid, the production process and the additives used for oil related production activities (Amakiri, *et al.*, 2022; Siagian *et al.*, 2018; Nasiri *et al.*, 2017). Fakhru'l-Razi et al. (2009) emphasises that although the composition of the water may vary from one source to another, it is similar in quality to the oil and/or gas being produced.

Produced water may contain fluids from within, below or above the hydrocarbon zone in a well. Therefore, in addition to surface water and production chemicals injected into the reservoirs for production enhancement, petroleum hydrocarbons and associated organic contaminants are always present in substantial amounts in produced water, which penetrates the oil and gas zone of production (Neff *et al.*, 2011; Fakhru'l-Razi *et al.*, 2009). The main contaminants present in produced water are organic compounds, inorganic compounds, oil and grease, treatment chemicals, solids and dissolved gases.

2.4.3.1. Organic Compounds

Neff *et al.* (2011) classifies petroleum hydrocarbons into two groups, namely saturated and aromatic, which appear in both dispersed and dissolved forms. Saturated hydrocarbons are more soluble in water than aromatic hydrocarbons of the same molecular weight and in general, their solubility in water is inversely proportional to their size, i.e., the higher the molecular weight, the lower the solubility in water (Neff *et al.*, 2011).

Organic compounds generally contain a dissolved mixture of phenol and BTEX (benzene, toluene, ethylbenzene and xylene), which are categorised as high priority chemicals. In this group of contaminants, polyaromatic hydrocarbons (PAH) are also included, which are characterised as hydrocarbons containing two or more fused aromatic rings that are toxic, persistent pollutants, unsafe for the environment and have been categorised as high-risk chemicals for both public health and environmental safety (Amakiri, *et al.*, 2022; Abbas *et al., 2021*). BTEX compounds are highly soluble in water and volatile, allowing them to remain in the aqueous phase post-oil removal. Phenol similarly has higher water solubility. PAHs, though more associated with oil, vary in molecular size and solubility, affecting their removal efficiency (Amakiri, *et al.*, 2022; Abbas *et al.*, 2021; Neff *et al.*, 2011). Table 2.2 contextualises concentration ranges of phenol, BTEX and PAH from selected locations around the world.

***Table 2.2**: BTEX, phenol and PAH concentrations in produced water from selected locations*

Compound	Concentration [mg/L]					
	Nigeria[a]	Malaysia[b]	Gulf of Mexico[c]	Brazil[d]	Indonesia[c]	Norway[a]
Benzene	0.8 – 2.98	0.2 – 21	0.44 – 2.8	1.397	0.084 – 2.30	4
Xylene	0.017 – 2.1	0.13 – 24	0.16 – 0.72	0.096	0.013 – 0.48	–
Ethylbenzene	0.01 – 1.5	0.2 – 22	0.026 – 0.11	0.148	0.026 – 0.056	2.3
Toluene	0.8 – 2.60	0.7 – 24	0.34 – 1.7	1.263	0.089 – 0.80	2.77
Phenol	0.1 – 7.9	0.07	0.723	2.1	0.33 – 3.64	5.1

Compound	Concentration [mg/L]				
	Nigeria[a]	North Sea[c]	Gulf of Mexico[c]	Canada[c]	Brazil[d]
Total PAH	0.005	0.4 – 1.5	0.040 – 0.600	2.148	0.014

[a] Amakiri *et al.*, 2022; [b] Johnsen *et al.*, 2004; [c] Neff *et al.*, 2011; [d] Dórea *et al.*, 2007

2.4.3.2. Inorganic Compounds

Inorganic compounds include dissolved formation minerals in the form of natural radioactive materials, heavy metals, cations and anions. Cations like K^+, Na^+, Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and Fe^{2+}, as well as anions such as HCO_3^-, Cl^-, CO_3^{2-} and SO_4^{2-} usually affect the chemistry of produced water in terms of scale formation, buffering capacity and salinity (Jimenez *et al.*, 2018; Faksness *et al.*, 2004). The salt concentration of produced water may vary from 1 to ±300 000 mg/L, while the sulphate concentration in produced water is lower than that of seawater. Although the sulphate concentration in some producing areas can be high due to oil enhancing recovery, such as water injection (Jimenez *et al.*, 2018).

Naturally occurring radioactive material (NORM), is derived from radioactive ions, predominantly radium, which co-precipitate with different types of scale from produced water. Barium sulphate is considered the most common scale co-precipitated. There have been accounts of a strong correlation between barium and radium isotope concentrations (Neff *et al.*, 2011). According to Fakhru'l-Razi *et al.* (2009), the concentration of heavy metals in produced water is dependent on the age of the wells and the geology of the formation. Table 2.3 contains the most abundant minerals and heavy metals found in produced water, along with their general concentration ranges in oil and gas fields worldwide.

Table 2.3*: Minerals and heavy metals in seawater and produced water (sourced from* Al-Ghouti *et al.*, 2019; Zheng *et al.*, 2016; Neff *et al.*, 2011; Fakhru'l-Razi *et al.*, 2009)

Parameter	Concentration [mg/L]		
	Oil Field Produced Water	Gas Field Produced Water	Seawater
Aluminium	310 – 410	< 0.500 – 83	-
Arsenic	< 0.005 – 0.300	0.004 – 151	-
Barium	1.3 – 650	9.65 – 1740	22 – 80
Boron	5 – 95	< 56	4.450
Bromide	46 – 1200	150 – 1149	87
Cadmium	< 0.005 – 5	< 0.020 – 1.210	4 – 23
Calcium	13 – 25800	9400 – 51300	416

Chromium	0.020 – 1.100	< 0.030	No Data
Copper	< 0.002 – 1.500	< 0.020 – 5	20 – 500
Iron	< 0.100 – 100	39 - 1100	1.800
Lead	0.002 – 8.800	< 0.200 – 10.2	20 – 81
Lithium	3 – 50	18.600 – 235	0.170
Magnesium	8 – 6000	0.900 – 4300	1294
Manganese	< 0.004 – 175	0.045 – 63	No Data
Nickel	0.020 – 0.300	< 0.080 – 9.200	No Data
Potassium	24 – 4300	149 – 3870	387
Silver	< 0.001 – 0.150	0.047 – 7	No Data
Sodium	132 – 97000	520 – 120 000	10 760
Strontium	0.020 – 1000	No Data	0.008
Zinc	0.010 – 35	< 0.020 – 5	0.300 – 1.400
Ammoniacal nitrogen	9.7 – 600	-	-
Bicarbonate	0 – 800	-	-
Carbonate	0.5 – 7851	< 0.1 – 47	-
Chloride	46 100 – 141 000	1400 – 190000	19 353
Nitrate	0.05	-	-
Nitrite	10 – 300	-	-
Phosphate	0 – 3.5	-	-
Sulphate	2 – 1650	< 0.1 – 47	2712
Sulphite	10	-	-

2.4.3.3. Oil and Grease

According to Amakiri *et al.* (2022), oil and grease refer to a broad category of substances that are not soluble in water such as fats, oils, waxes, and other similar substances and are the constituents of offshore produced water that have gained considerable attention and evolved into crucial regulatory criteria for state legislators, whereas other dangerous substances are widely overlooked. There are both dissolved and dispersed oil in produced water. The amount of dissolved oil in produced water depends on production volume, the technique used to artificially lift the hydrocarbons, the well's age and the most dominant hydrocarbons in the oil. The amount of dispersed or suspended oil in produced water depends on the oil's density, the shear history of the droplet, the quantity of oil precipitation and the interface tension between the water and oil (Amakiri, *et al.*, 2022; Nasiri *et al.*, 2017; Fakhru'l-Razi *et al.*, 2009).

2.4.3.4. Treatment Chemicals

Treatment chemicals are required to aid hydrocarbon fluid recovery, stimulate oil, water and gas separation, maintain steady production, as well as prevent and mitigate operational issues such as corrosion, wax appearance, emulsions, foaming, scale formation, hydrate formation, asphaltene build up and bacterial growth (Al-Ghouti *et al.*, 2019). A number of these chemicals are oil-soluble instead of water-soluble and hence remain in the oil phase. While the water-soluble chemicals remain in the produced water and are disposed of with it. When compared to oil-soluble chemicals at the same concentration, water-soluble chemicals have no detrimental effects in the aqueous phase. Nonetheless, at high concentrations, certain production chemicals can enhance the partitioning of oil components into the aqueous phase in addition to precipitating and accumulating in marine sediments (Neff *et al.*, 2011; Fakhru'l-Razi *et al.*, 2009). Table 2.4 provides a list of chemicals commonly used during oil and gas production, either in pure forms or as active ingredients in a solvent or co-solvent, which partitions themselves into phases depending on the solubilities in oil, gas or water.

Table 2.4*: Typical treatment chemicals used in oil and gas operations (Amakiri et al., 2022; Zheng et al., 2016; Igunnu and Chen, 2014; Knutsen, 2011; Neff et al., 2011)*

Chemical	Purpose	Phase Association	Injection Point	Average Concentration Range
Corrosion Inhibitor	Prevents corrosion related damage in hydrocarbon production equipment	Oil	Outlet water of all the phase separators	25 - 100 ppm
Scale Inhibitor	Prevents excessive barium and strontium sulphate scale in the seawater treatment and water injection system due to changes in temperature and/or pressure while water flows	Water	Inlet to all phase separators, inlet to hydrocyclone and inlet to seawater treatment system	3 - 10 ppm
Emulsion Breaker	Encourages separation of oil and water from emulsions	Oil	Inlet to all the phase separators	10 - 200 ppm
Biocide	Kills-off any bacteria and prevents bacterial growth within the seawater treatment and water injection system	Water	Produced water slop tanks and inlet to seawater treatment system	10 - 200 ppm
Methanol	Prevents the formation of crystals that form at high pressure and low temperature flows where water and natural gas are present (hydrates)	Gas	Outlet of gas dehydrator contactor and outlet of gas compression system	15 - 75 L/MMscf
Coagulants and Flocculants	Promotes the aggregation of suspended particles in the water stream	Water	Inlet to hydrocyclone and outlet of induced gas flotation units	25 ppm

Foam Inhibitor	Prevents the foaming of crude while undergoing depressurisation to release entrapped gas in the production separators	Oil	Inlet to high-pressure phase separator	30 - 35 ppm
Solids Inhibition	Asphaltene, wax and naphthenate inhibitors prevent solids build-up in the production system that leads to a production rate decline, higher viscosity, oil/water separation issues and flow assurance problems, resulting in fouling of pipelines and processing equipment	Oil	Inlet to all the phase separators	< 3 ppm
Oxygen/Chlorine Scavenger	Removes oxygen and chlorine from the seawater to inhibit corrosive action in the water treatment and injection system pipework and also helps prevent the growth of aerobic bacteria in the reservoir	Water	Inlet and outlet of seawater deaerator column	10 - 15 ppm
pH Controller	Reduces the glycol degradation and protects the entire gas dehydration system against rapid corrosion	Water	Within the gas dehydration system	n/a
Water Clarifier	Removes oil and entrapped solids from produced water	Water	Inlet to hydrocyclone	25 pm

2.4.3.5. Solids

Amakiri *et al.* (2022), states that there are both natural and artificial solids present in produced water, with natural solids being reservoir materials like plankton, silt, algae, clay, sediment and sand; and artificial solids originating from external intervention, such as wax, corrosion debris, drill mud and cement fines. Dissolved solids have the potential of causing serious problems during oil production, like clogging flow lines (Igunnu and Chen, 2014*).*

2.4.3.6. Dissolved Gas

Dissolved gases commonly found in produced water include oxygen, carbon dioxide, hydrogen sulphide and nitrogen, which at large quantities can be corrosive and toxic. These gases are naturally occurring, formed by bacterial activity or by chemical reactions within the water (Amakiri *et al.*, 2022; Igunnu and Chen, 2014).

2.4.4. Characteristics of Produced Water

Abbas *et al.* (2021) enunciates that the composition of produced water is a combination of dissolved and particulate organic and inorganic chemicals with its characteristics highly dependent on the quality of crude oil, the origin of wastewater pollutants and the operating conditions. Thus, a wide domain of contaminants in varying quantities have been observed within various research papers (Abbas *et al.*, 2021; Dawoud *et al.*, 2021; Olajire, 2020; Nasiri and Jafari, 2017; Zheng *et al.*, 2016; Igunnu and Chen, 2014 and Fakhru'l-Razi *et al.*, 2009). Table 2.5 summarises a range of produced water characteristics from different oil fields in the world.

Table 2.5*: Characteristics of oil and gas fields hydrocarbon produced water (adapted as a range from Abbas et al., 2021; Dawoud et al., 2021; Olajire, 2020; Nasiri and Jafari, 2017; Zheng et al., 2016; Igunnu and Chen, 2014 and Fakhru'l-Razi et al., 2009)*

Parameter	Oil Field Produced Water	Gas Field Produced Water
Density [kg/m^3]	1014 – 1140	67 – 38000
Conductivity [µS/cm]	1725 – 58600	4200 – 586000
Surface Tension [dynes/cm]	43 – 78	–
Total Organic Carbon - TOC [mg/L]	0 – 1500	491 – 120000
Biochemical Oxygen Demand - BOD [mg/L]	1.2 – 21820	8 – 5484
Total Petroleum Hydrocarbons - TPH [mg/L]	0.39 – 45	–
Chemical Oxygen Demand - COD [mg/L]	380 – 9000	2600 – 120000

Total Dissolved Solids - TDS [mg/L]	8.2 – 80740	–
Total Suspended Solids - TSS [mg/L]	1.2 – 1000	2600 – 36000
pH	4.3 – 10	3.1 – 7.0
Oil and Grease / Oil-in-Water [mg/L]	2 – 565	–
Benzene, Toluene, Ethylbenzene and Xylene - BTEX [mg/L]	0.39 – 35	25 – 400
Polycyclic Aromatic Hydrocarbons - PAH [mg/L]	0.04 – 3	–
Base/neutrals [mg/L]	< 140	–
Total polar [mg/L]	9.7 – 600	–
Turbidity (NTU)	32	–
Higher Acids [mg/L]	< 1 – 63	–
Phenols [mg/L]	0.001 – 23	–
Volatile Fatty Acids [mg/L]	2 – 4900	–

Produced water from oil fields has a density ranging of 1014 to 1140 kg/m^3, a surface tension of 43 to 78 dynes/cm and a pH range of 4.3 to 10, which is similar to the parameters of seawater (i.e., density of 1025 kg/m^3, surface tension of 70 to 77 dynes/cm and pH of 7.5 to 8.4), therefore providing evidence that produced water indeed originates from seawater (Zheng *et al.*, 2016). In gas fields, water injection is not used, resulting in produced water that is a blend of formation water and condensed water. Produced water from gas fields have a smaller volume compared to those in oil fields but are more acidic. Additionally, the concentration of volatile components in produced water from gas fields is greater than in oil fields (Fakhru'l-Razi *et al.*, 2009).

2.4.5. Environmental Trade-offs in Energy Production

While energy production is essential for societal development and maintaining high standards of living, it inevitably leads to environmental trade-offs, particularly evident within the oil and gas industry. The need for efficient energy production must be balanced against the potential for ecosystem disruption, emissions of greenhouse gases and other environmental impacts directly involved with oil and gas production (Costa-Campi *et al.*, 2017). This balance is precarious, as the processes involved specifically in treating and disposing of produced water can have significant ecological footprints, for instance, improper disposal of produced water can lead to soil and water contamination, posing risks to human health and biodiversity (Gamwo *et al.*, 2022). Additionally, the treatment of produced water often requires considerable energy inputs, further exacerbating the environmental impacts associated with energy production.

Governments and regulatory bodies play a crucial role in mediating these trade-offs through policies and regulations to mandate the use of best practices in produced water management, such as promoting the use of cleaner and more efficient technologies, thus, ensuring energy production does not come at an unacceptable environmental cost (Gamwo *et al.*, 2022; Jimenez *et al.*, 2018).

2.4.6. Produced Water Management

The management of produced water is influenced by a range of important factors, with the primary focus on the legal requirements, the environmental impact, potential water-related hazards and the feasibility and economic viability of available options (Gray, 2020). Typically, the efficient management of produced water begins with initiatives aimed at minimising the total volume of produced water generated in oil and gas fields. As detailed in Table 2.6, some of the options available to oil and gas operators for managing produced water include: 1) Prevention of water from rising to the processing platform; 2) Injection of produced water into reservoirs; 3) Discharge of produced water to the environment; 4) Recycling of water for oil and gas operations; and 5) Recycling of water for household and industrial use.

Table 2.6*: Produced water management alternatives (Gray, 2020; Arthur et al., 2005)*

Management Route	Description
Prevention of Produced Water Generation	Considered the most decorous alternative as it eliminates water coming with oil and gas to the surface. However, it is rarely possible. It entails using polymer gels to block water from coming with the oil and gas during exploration or using downhole water separators
Environmental Discharge	Requires treatment of the water to reduce certain contaminants to meet the region's discharge regulations
Reuse in Oil and Gas Operations	Entails the treatment of water to meet the quality criteria needed for reservoir injection, fracturing, drilling, stimulation, workover and oil washing operations
Reuse in Domestic and Industrial Applications	Entails significant treatment of the water to meet the quality criteria required for irrigation, animal and even human consumption

2.4.6.1. Produced Water Treatment

Produced water treatment is a necessary step for reducing the concentration of hazardous constituents prior to reusing or discharging to sea in order to safeguard the integrity of the final usage, as well as the marine environment (Gray, 2020; Nasiri and Jafari, 2017). The objective of the produced water treatment is to eliminate solids and colloids and reduce the content of free, dispersed and emulsified oil, the concentration of algae, bacteria, microorganisms, dissolved organics and inorganics such as salts and minerals (Nasiri and Jafari, 2017).

Onsite water reuse technologies enable the industry to reduce the water demand and supplement limited water supplies (Gray, 2020; Jiménez *et al.*, 2018). Treated produced water has the potential to be used within and beyond oil fields, including activities like water and steam flooding, drilling and hydraulic fracturing within oil fields, as well as broader applications like crop irrigation, wildlife consumption, aquifer recharge and non-oil field power generation. However, non-oil field reuse faces numerous challenges such as technical, regulatory, economic, environmental and social aspects that can limit viability (Gray, 2020).

Selection of the treatment method is a challenge and to meet the specified objectives, most operators apply more than one standalone treatment (Amakiri *et al.*, 2022). Primary treatment methods are physical treatments particularly concerned with the elimination of suspended hydrocarbon components and solid particles from produced water and are not suitable for removing dissolved compounds. Secondary treatments are used to achieve up to 90% oil removal, eliminate soluble organic compounds and dissolved aromatic hydrocarbons such as BTEX. Tertiary treatments methods are categorised as polishing methods and focus on salt removal (mainly nitrates and phosphates), reduction of turbidity and concentration of ultra-small droplets of oil and other impurities such as PHAs (Amakiri *et al.*, 2022; Abbas *et al.*, 2021; Dawoud *et al.*, 2021; Olajire, 2020). Table 2.7 represents the legend of Table 2.8, which summarises the working principle of selected produced water treatment, as well as the efficiency of each technology in removing certain contaminants, namely phenol, BTEX, PAH, inorganic compounds, TSS and O&G.

Table 2.7: *Legend of ranking system used for the contaminants removal efficiency per technology*

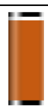
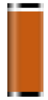
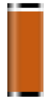
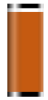
	> 95%		80 - 95%		60 - 80%		40 - 60%		< 40%

***Table 2.8**: Working principle of selected produced water treatment and its efficiency in removing certain contaminants (Amakiri et al., 2022; Jimenez et al., 2018; Nasiri and Jafari, 2017; Mousa and Al-Hasan, 2017; Igunnu and Chen, 2014; Fakhru'l-Razi et al., 2009; Arthur et al., 2005)*

	Principle	**Contaminants Removal Efficiency**					
		Phenol	**BTEX**	**PAH**	**Inorganic Compounds**	**TSS**	**O&G**
Hydrocyclone	Separation based on density differences between heterogeneous phases, influenced by the centrifugal force field created through the pressurised tangential feed inlet.						
Coagulation/Flocculation	Removal of suspended oil droplets, particles and colloids by the addition of chemical agents resulting in solid particles coalescencing in the form of large flakes						
Gravity Separator	Separation relies on variations in densities among the heterogeneous phases, influenced by gravity and facilitated by the spinning motion generated by the moving parts of the equipment						
Flotation	Tiny gas bubbles are injected as a transport medium into the water to separate absorbed particulates and oil droplets						
Adsorption	Surface phenomenon that occurs as a result of the interaction forces between the surface of the adsorbent, the solvent and the dissolved species in the water					N/A	

Activated Sludge	Uses highly concentrated microorganisms to degrade organics and remove nutrients from wastewater					N/A	
Electrodialysis	Separation of dissolved ions from water through electrically charged ion exchange.					N/A	
Freeze-Thaw / Evaporation	Temperature swings are used to alternately freeze, thaw and evaporate produced water to separate dissolved solids						
Macro Porous Polymer Extraction	Water flows into macro porous polymer particles where hydrocarbons with an affinity for the extraction solvent are retained and removed.					N/A	
Membrane Technology – **Microfiltration**	Cross-flow or dead-end filtration, driven by pressure or vacuum						
Membrane Technology – **Ultrafiltration**							
Membrane Technology – **Nanofiltration**							
Membrane Technology – **Reverse Osmosis**							

Advanced Oxidation	**Ozonation**	Decontamination by O_3 in water treatment that can occur by the direct reaction between O_3 and the dissolved organic species, via a reaction between the generated hydroxyl radicals and the dissolved pollutants or a combination of both					N/A	
	Fenton	Decontamination by the addition of a catalytic oxidative mixture that contains an oxidant (usually H_2O_2) and a catalyst (usually ferrous salts)					N/A	
Photocatalytic Oxidation		Chemical process that utilises light and a photocatalyst to promote oxidation reactions					N/A	

2.4.6.2. Produced Water Reuse

2.4.6.2.1. Produced Water Reuse Within the Oil and Gas Industry

The reinjection of produced water is a widespread practice in the oil and gas industry aimed at maintaining hydraulic pressure and enhancing resource recovery. This approach is both environmentally responsible and cost-effective, as it reduces the need for produced water disposal and aids in complying with disposal regulations. Table 2.9 provides the standard regulatory targets for produced water reinjection to be achieved by oil and gas platform offshore (Jiménez *et al.*, 2018).

***Table 2.9**: Regulatory limits for water injection (adapted from Jiménez et al., 2018)*

Component	Maximum Limit (ppm)
SO_4^{2-}	100
TDS	1000 – 10 000
TSS	10
Oil and Grease	42

According to USEIA (2016), as a successful example, in the USA's Permian Basin, the Apache Corporation efficiently treated produced water for drilling purposes, significantly reducing water management costs by recycling up to 90% of produced and non-potable water. In Canada, Tourmaline Oil Corp. gained regulatory approval for a large containment pond to store and reuse produced water for hydraulic fracturing, achieving a recycling rate of over 95% for flowback water from gas operations Gray (2020). In the Canadian oil sands sector, operators have developed an advanced produced treatment method where a steam-assisted gravity drainage pilot plant recycles 90% of the water used to generate steam and facilitate bitumen extraction (Alberta Innovates, 2017).

2.4.6.2.2. Produced Water Reuse Outside the Oil and Gas Industry

There is a growing trend in the oil and gas sector to recycle produced water for various downstream purposes, including irrigation, livestock watering, aquifer storage, municipal use and industrial applications. Maintaining high water quality is vital, particularly for irrigation, where factors such as pH, chloride levels, nitrate levels and boron concentration must meet crop health and safety standards. Regulatory requirements from International Maritime Organisation (IMO), United Nations Environmental Programme (UNEP), Water Framework

Directive (WFD), Environmental Protection Agency (EPA), among others, primarily revolve around parameters such as the concentration of oil and gas, toxicity, suspended solids and overall water quality to protect the environment and public health (Jiménez *et al.*, 2018).

Fig. 2.7 displays the concentrations of key water quality components (Na^+, Cl^-, $SO4^-$, TDS) across different uses as set by the US-EPA and US Department of Agriculture Natural Resources Conservation Service. Each cell reveals the specific concentration in g/m^3, highlighting the variation in water quality requirements and challenges across these applications (Al-Ghouti *et al.*, 2019).

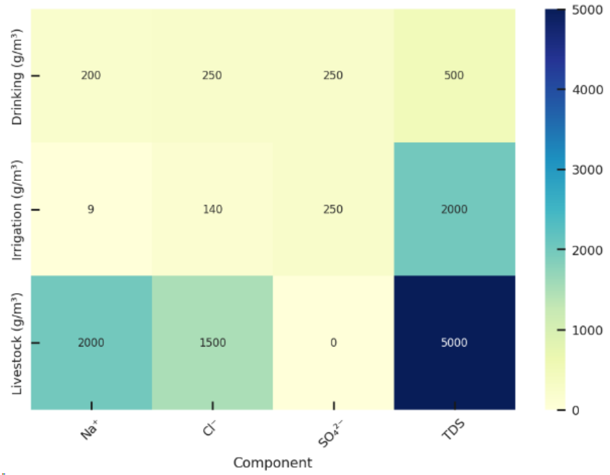


***Figure 2.7**: Heatmap of water quality parameters for drinking, irrigation and livestock use (adapted from Al-Ghouti et al., 2019)*

The necessity for water treatment varies based on its intended reuse and logically applications such as drinking or agricultural use demand more advanced purification than for reuse within the oil and gas industry. The choice of treatment method and intended reuse is also influenced by cost considerations. Bagheri *et al.* (2018) highlighted that the cost of traditional treatment methods like hydrocyclone and gravity separation are more cost-effective for water reinjection (\$0.509/m^3) when compared with sophisticated and costly techniques like mechanical vapor compression for water reuse outside of the oil and gas industry (\$3.808/m^3).

Successful examples of produced water reuse outside of the oil and gas industry include use for irrigation in Australia and the United States, particularly for salt-tolerant crops, where low-salinity produced water is treated before application. In Kern County, California, treated produced water with low salinity is blended with fresh water sources for irrigation (Gray, 2020). In Australia's Bowen Basin, moderately saline produced water from coal bed methane operations undergoes reverse osmosis treatment and is used for large-scale forestry, legume plantations and other agricultural purposes (Mallants *et al.*, 2017). According to Gray (2020), in Oman, produced water with moderate salinity is treated in constructed wetlands, benefitting wildlife habitats and reducing carbon emissions associated with deep well disposal equipment. Australia's electrical power industry supplements the cooling water feed with treated produced water, while in various industrial and commercial applications, produced water is used for tasks like vehicle washing, firefighting and road dust control (Gray, 2020).

2.4.6.3. Environmental Discharge

In the conventional processes of the oil and gas sector, produced water has been traditionally subjected to gravity-based separation methods and subsequently released into the environment. This discharge occurs either on offshore platforms or through reinjection into onshore soil. However, this long-standing practice, though compliant with certain standards, has the potential to result in contamination of soil, surface water and subterranean water sources (Gray, 2020; Hedar, 2018; Igunnu and Chen, 2014).

Fig. 2.8 displays the highest allowable daily limits of oil-in-water concentrations in produced water discharged into the marine environment from the top 20 oil producing countries in 2022, which varies across different countries demonstrating how environmental, economic and technological factors, along with international pressures, shape national policies on environmental impacts from oil and gas operations.

***Figure 2.8**: Maximum allowable oil-in-water concentration limits in produced water discharges by the top 20 oil producing countries in 2022*

China's stringent limit of 10 mg/L oil-in-water suggests a strong emphasis on environmental protection driven by sensitive marine ecosystems and public environmental consciousness. In contrast, countries like Iraq, UAE and Libya, with much higher limits of 100 mg/L, prioritise their economic reliance on the oil and gas industry due to technological and/or economic constraints in water treatment processes (Amakiri *et al.*, 2022; Jimenez *et al.*, 2018; Igunnu and Chen, 2014; Neff *et al.*, 2011). Wealthier nations such as Norway (20 mg/L), indicate their capacity to invest in sophisticated water treatment solutions, balancing environmental protection with industry needs.

2.4.7. Environmental Footprint of Discharged Produced Water

Presently, nearly 30% of the world's produced water is treated and utilised for hydraulic fracking, water injection and steam generation. The remainder is still treated for discharge directly into the marine environment (Hedar, 2018). The vast amount of hydrocarbon produced water discharged to the surrounding aquatic environment, the complexity of some dangerous chemicals and the lack of understanding of likely long-term and ecological impacts have elevated produced water to the top of the list of concerns and research (Amakiri *et al.*, 2022; Nasir *et al.*, 2017). Untenable methods of handling and disposal of produced water can have some detrimental effects on the environment because the impact in a specific environment is dependent on the physical, chemical and biological composition of that environment. The toxicity and organic loading of the waters can broadly categorise the magnitude of the environmental impact (Hedar, 2018; Gazali *et al.*, 2017).

According to Gazali *et al.* (2017) dissolved inorganic ions, metals and radionuclides abundant in produced water at concentrations moderately higher than seawater, are not a concern when discharged to the ocean. Nevertheless, if the produced water is discharged into shallow marine waters, certain elements may aggregate in sediments near the discharge, potentially affecting bottom living biological species and since salinity as given by TDS is higher than seawater, could result in aquatic devastation by influencing the hardness of the water (Olajire, 2020). Sodium, a prevalent component in produced water, can lead to soil degradation by altering soil textures and causing erosion. High sodium levels compete with essential nutrients for plant roots, potentially causing nutrient deficiencies and poor soil structure. Produced water salts have wide-ranging effects on soil, water quality and ecosystems, contributing significantly to environmental toxicity (Hedar, 2018).

Organic compounds can be present in produced water in the form of dispersed oil and non-hydrocarbon substances. Dispersed oil is defined as small and definite droplets suspended in water that do not precipitate on the seabed but rise to the water's surface, whereas non-hydrocarbon organic materials are dissolved in water. Organic materials raise the BOD and COD of the produced water, as well as the toxicity, which can be expressed as acute or chronic (Hedar, 2018). Some produced water contains chemicals that are extremely toxic to delicate marine organisms even at low concentrations. When discharged within coastal waters, or in an area with low water turbulence and prevailing speeds, the chemical concentrations can remain high for a long enough period to cause permanent ecological damage (Abbas *et al.*, 2021; Hedar, 2018; Gazali *et al.*, 2017; Neff *et al.*, 2011).

According to Olajire (2020), due to the extraordinary volatility and rapid dissolution in seawater, BTEX are typically neglected when considering the possible hazardous effects on aquatic ecosystems. On the contrary, the release of PAH has been emphasised in environmental pollution regulations, due to a variety of adverse effects in fish, including disruption to genetic, fertility, reproductive and maturation systems (Bakke *et al.*, 2013). Moreover, based on studies, there is a moderate risk of acute toxicity to the marine ecosystem related to the discharge of produced water, with the predominant risk factor indicated as the concentration of the constituents present in the water, instead of the total discharge amount containing potentially critical toxic elements (Knutsen, 2011). Produced water from gas fields contain higher levels of BTEX compounds and are roughly ten times more toxic than produced water from oil fields (Hedar, 2018).

Some production treatment chemicals, when discharged in high concentrations in produced water, can be toxic. Water soluble chemicals are less toxic in the water phase compared to oil soluble chemicals at similar concentrations, but can increase the presence of oil compounds in water at high concentrations. These chemicals may also accumulate in marine sediments (Hedar, 2018; Neff *et al.*, 2011).

Overall, the environmental effects of produced water depend on the chemical composition, which can be harmful to marine life due to its toxicity. Research has shown a low risk of immediate harm to the marine environment from released produced water, with the primary concern being the concentration of harmful substances rather than the total volume of produced water discharged. Nonetheless, the discharge volume remains important to consider because it is linked to production and drilling operations (Olajire, 2020).

2.4.8. Produced Water Treatment Market

The expansion of the produced water treatment market is driven by various factors, including the rising need for oil and gas and escalating exploration and production endeavours (Grand View Research, 2023). Governments worldwide enforcing strict environmental regulations and water quality standards, mandate oil and gas companies to deploy efficient systems for produced water treatment, aiming to mitigate any environmental consequences (Amakiri *et al.*, 2023). Market growth is also steered by the increasing embrace of cutting-edge technologies, which not only enhance treatment efficiency but also lower operational expenses. In addition, the growing consciousness and corporate commitment to sustainable practices contributes to market expansion, with companies actively seeking eco-friendly treatment technologies to diminish their ecological footprint (Grand View Research, 2023; Spherical Insights, 2023).

According to Grand View Research (2023), the produced water treatment market encounters challenges primarily stemming from the substantial capital investment needed to establish advanced treatment infrastructure, especially for small and medium-sized oil and gas companies. The intricate nature of produced water composition adds to the difficulty in devising standardised treatment solutions. Moreover, the volatility in oil and gas prices, coupled with economic uncertainties, can influence investment decisions and budgets for projects related to produced water treatment. In a broader sense, the limited awareness and comprehension of the advantages of produced water treatment in specific regions pose an additional obstacle to market growth (Grand View Research, 2023; Spherical Insights, 2023). Spherical Insights (2023) states that the global produced water treatment market size was valued at USD 7.2 billion in 2022 and is expected to reach USD 13.74 billion by 2032, at a compound annual growth rate of 6.2% during the forecast period of 2023 to 2032 (Fig. 2.9).

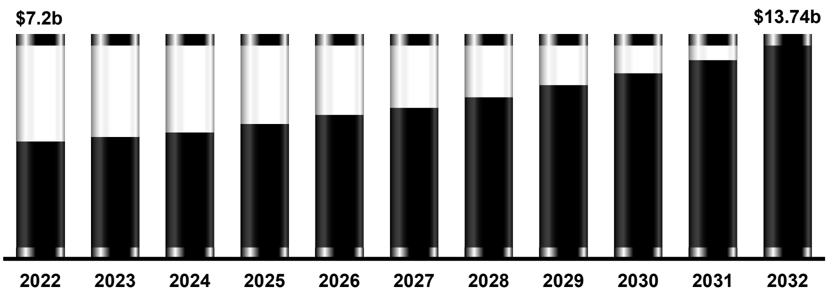


Figure 2.9: *Produced water treatment market growth projection from 2022 to 2032 (Spherical Insights, 2023)*

The produced water treatment industry is categorised according to the type of treatment into physical, chemical and biological treatment (Grand View Research, 2023; Spherical Insights, 2023). In 2022, the physical treatment segment, encompassing hydrocyclones, gas flotation and adsorption, constituted ±43% of the global revenue share (Grand View Research, 2023). In the same year, the chemical treatment segment contributed ±29% to the global revenue share, benefiting from its versatility and effective approach in addressing diverse contaminants found in produced water (Spherical Insights, 2023; Grand View Research, 2023). The biological treatment segment held a ±28% share of the global revenue in 2022. With the ongoing growth in the oil and gas industry and the adoption of biological wastewater treatment methods, this segment is projected to experience a compound annual growth rate of 8.0% until 2032 (Spherical Insights, 2023)

In offshore extraction facilities, where space is limited, compact physical and chemical treatment technologies are preferred (Grand View Research, 2023). As of 2022, the onshore segment held the majority, claiming over ±65% of the market share, with an anticipated compound annual growth rate of ±8% until 2032 (Grand View Research, 2023; Spherical Insights, 2023). This dominance is attributed to onshore locations contributing significantly to global production, with numerous extraction activities occurring on land. Furthermore, regulatory compliance and environmental concerns related to onshore operations tend to be more stringent, necessitating effective produced water treatment to minimise environmental impact (Spherical Insights, 2023).

2.5. Conclusion

This chapter highlighted the key challenges in managing produced water within the oil and gas industry. A significant gap remains in the lack of universally effective and economically viable treatment methods, largely due to the complex and variable nature of produced water. Existing technologies often fall short in terms of efficiency, cost-effectiveness and minimizing environmental impact. Moreover, there is limited understanding of the long-term environmental consequences of discharging produced water into marine environments, particularly concerning trace contaminants, underscoring the need for continuous research and monitoring. The diversity of regulatory standards across regions further complicates multinational operations, suggesting a strong need for more harmonised global regulations. Looking ahead, the industry should prioritise investment in research and development to drive innovation in treatment technologies. Strengthened collaboration among industry stakeholders, researchers, and policymakers is essential to develop clear regulations, share best practices, and promote the adoption of advanced solutions. Ultimately, placing greater emphasis on environmental stewardship is critical to reducing ecological harm. Addressing these gaps through comprehensive, innovative, and collaborative efforts is fundamental to balancing operational efficiency with environmental responsibility, paving the way toward more sustainable energy production.

CHAPTER 3
COLD EYE REVIEW OF PRODUCED WATER DEOILING: HYDROCYCLONE AND INDUCED GAS FLOTATION

César, S.D., De Jager, D. and Njoya, M., 2025. The Role of Hydrocyclone and Induced Gas Flotation Technologies in Offshore Produced Water Deoiling Advancements. *Petroleum Research*, 10(2), pp.342-351.

"Let the wise listen and add to their learning, and let the discerning get guidance."
Proverbs 1:5

CHAPTER 3: COLD EYE REVIEW OF PRODUCED WATER DEOILING: HYDROCYCLONE AND INDUCED GAS FLOTATION

3.1. Introduction

Produced water, the largest waste stream from oil and gas extraction, consists of a complex mix of formation water, injected flowback, and condensation water (Jimenez *et al.*, 2018; Gamwo *et al.*, 2022). Its composition varies widely depending on geology, reservoir characteristics, and extraction methods (Fakhru'l-Razi *et al.*, 2009). As wells mature, produced water volumes increase, raising operational costs and environmental risks (Lusinier *et al.*, 2019). Effective deoiling before discharge is crucial to meet strict environmental regulations and prevent ecological damage such as surface slicks and harm to marine life (Cordes *et al.*, 2016; Olajire, 2020). Public and media scrutiny of oil spills further underscores the importance of advanced treatment technologies.

In response, induced gas flotation (IGF) and hydrocyclone technologies have greatly improved deoiling efficiency and reduced equipment footprint (Richerand, 2012; Eyitayo *et al.*, 2023). While widely studied individually, their integration into dual-stage systems like hydrocyclone-induced gas flotation (HIGF) remains less explored, despite its potential to boost oil removal and reduce energy use. By combining rapid bulk separation from hydrocyclones with fine droplet removal by flotation, HIGF systems offer a compact, energy-efficient solution well suited to offshore operations.

The objectives of this chapter were to examine the historical development and technological advancements of hydrocyclone and IGF systems since the 1940s; compare their efficiency, energy consumption, and waste generation; analyse the benefits of integrating HIGF systems for offshore produced water treatment; and highlight global regulatory variations while underscoring the need for stricter environmental standards to minimise ecological impacts.

3.2. Total Oil and Grease

Total Oil and Grease (TOG), commonly referred to as oil-in-water (OIW), includes a variety of hydrocarbons such as aliphatic hydrocarbons, aromatics, waxes, fats, and other organic compounds present in produced water (Amakiri *et al.*, 2022). These contaminants primarily originate from crude oil and pose significant environmental risks, necessitating effective treatment before produced water is discharged or reused (Oliveira-Junior and Pereira, 2020; Abbas *et al.*, 2021). The concentration of TOG in produced water typically ranges from 2 to

565 mg/L, with variations arising due to factors such as geological location, reservoir age, formation depth, and hydrocarbon recovery methods (Fakhru'l-Razi *et al.*, 2009; Siagian *et al.*, 2018; Olajire, 2020).

Younger reservoirs often produce less water with higher oil content, while older wells generate more water with lower hydrocarbon concentrations (Zheng *et al.*, 2016). Enhanced oil recovery (EOR) techniques, such as water flooding or steam injection, can further increase the presence of emulsified hydrocarbons, complicating the treatment process (Olajire, 2020). Regulatory standards for TOG concentrations in produced water vary by region, with limits ranging from 10 to 100 ppm, with stricter controls in regions like the European Union and China (Igunnu and Chen, 2014; Jimenez *et al.*, 2018). TOG concentrations are also influenced by the design and operation of separation equipment, as factors like retention time, pressure, and flow rate can affect treatment efficiency (Bram *et al.*, 2020). Additionally, environmental factors such as temperature play a role, with higher temperatures increasing hydrocarbon solubility and lower temperatures causing wax precipitation (Zheng *et al.*, 2016). Managing TOG effectively is crucial to meet regulatory compliance and minimise environmental impact, especially in offshore operations where space and resources are limited.

3.3. Offshore Produced Water Deoiling

3.3.1. Historical Perspective of Produced Water Deoiling

In the six decades between 1940 and 2000, water treatment technology evolved significantly from basic to advanced methods, as depicted in Fig. 3.1, representing a timeline of the produced water deoiling.

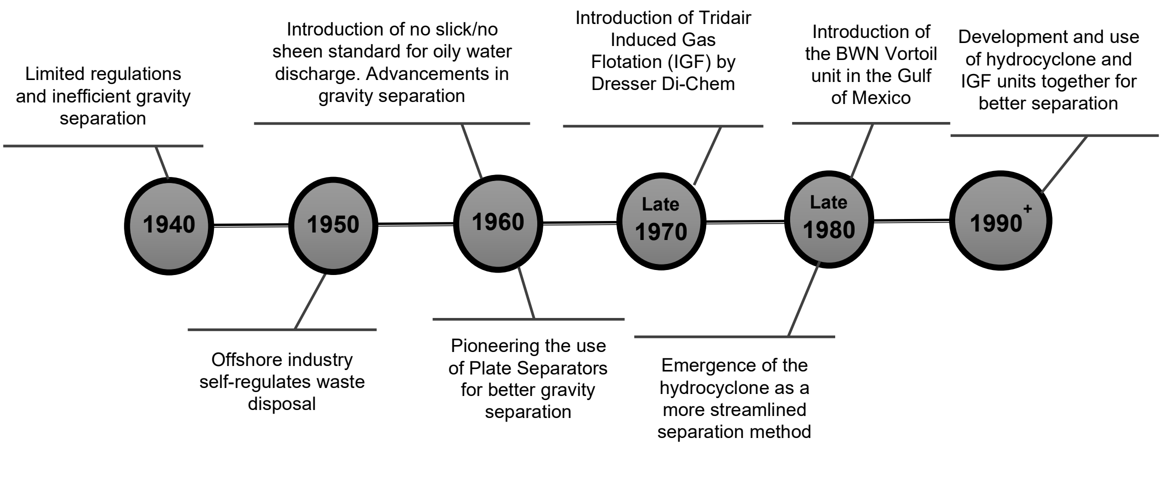


Figure 3.1: *Historical perspective on produced water deoiling (adapted from Richerand, 2012)*

In the 1940s and 1950s, the offshore industry self-regulated waste disposal, resulting in poor water treatment with inefficient gravity separation systems that required significant space. By the 1960s, regulatory standards introduced a no slick/no sheen requirement for oily water discharge, enforced through self-imposed guidelines and visual inspections. Despite continued use of gravity separation tanks, increasing oil production and limited offshore platform space led to challenges in handling larger volumes of produced water, resulting in reduced retention times and higher oil concentrations in discharged water (Richerand, 2012).

Technological advancements emerged in the late 1960s, with innovations like induced gas flotation and hydrocyclones. These technologies offered more efficient and space-saving alternatives to gravity separation, using gas bubbles and centrifugal force for oil-water separation. According to Eyitayo *et al.* (2023) and Richerand (2012), the evolution of these technologies was influenced by several key drivers, such as:

- Increasingly stringent discharge regulations which required more effective removal of oil and grease from produced water to prevent environmental harm.
- High-profile oil spills, which raised public awareness and media scrutiny, emphasising the need for better produced water management practices.
- Advances in engineering and materials science that enabled the development of more efficient and compact treatment technologies.
- The expansion of offshore oil production, which increased the volume of produced water, necessitating more effective treatment solutions.

By the 2000s, induced gas flotation and hydrocyclones became predominant methods for offshore deoiling of produced water, improving efficiency, space utilisation and compliance with environmental regulations. These technologies continue to evolve, incorporating new materials and designs to address emerging environmental challenges (Richerand, 2012).

3.3.2. Induced Gas Flotation

Flotation is a process employed for the removal of absorbed particles and oil droplets from produced water, utilising gas bubbles as a transport medium (Amakiri *et al.*, 2022). This technology has found extensive use in offshore produced water treatment due to its advantages, including the simultaneous removal of oil and suspended solids (Liu *et al.*, 2021). Induced gas flotation is particularly effective, achieving oil removal efficiencies of up to 90%

or more, effectively reducing oil concentrations to below 25 ppm under optimal conditions (Piccioli *et al.*, 2020).

The gas flotation process, illustrated in Fig. 3.2, involves three primary stages. First, bubbles are generated; second, agglomerates form through contact; and finally, attachment occurs between the bubbles, the oil droplets and the suspended solids due to density differences forming an aerated solution (Al-Ghouti *et al.*, 2019; Gamwo *et al.*, 2022). These sequential steps lead to the formation of foam on the water surface, which is subsequently removed, leaving clarified water to be collected below the flotation zone (Jimenez *et al.*, 2018; Al-Ghouti *et al.*, 2019).

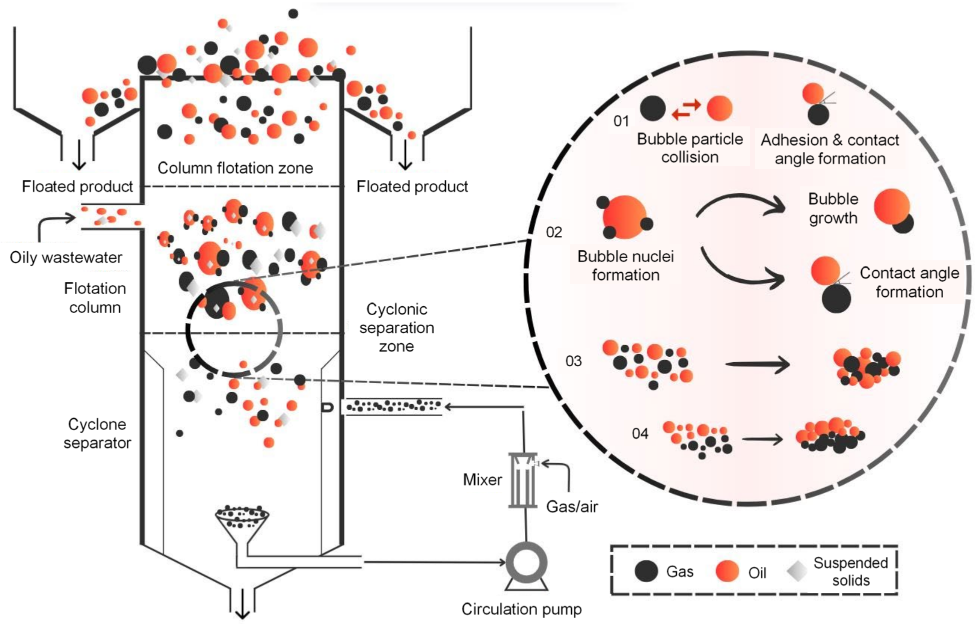


Figure 3.2*: Operating principle of the flotation technology (adapted and merged as a combination of Liu et al., 2021; Amakiri et al., 2022)*

Enhancing collision and adhesion of the gas in the flotation process can lead to improved separation efficiency (Saththasivam *et al.*, 2016). However, it is important to note that flotation comes with drawbacks, including high power consumption due to the need for increased energy to pressurise the system for effective gas circulation. Moreover, the sludge generated during flotation necessitates removal and proper disposal (Amakiri *et al.*, 2022). Saththasivam *et al.* (2016) validated that gas flotation is able to reduce oil concentration from 1000 ppm to 10 ppm in produced water and is able to provides a high removal efficiency at a larger throughput or a shorter retention time for any given flow rate. Particles of 25 µm can be removed by a dissolved gas flotation process and when using coagulation as a pre-treatment step, contaminants of 3 to 5 µm can be removed as well (Al-Ghouti *et al.*, 2019).

According to Liu *et al.* (2021), Stokes' Law is applicable for calculating final floc flotation, but separation efficiency depends on the collision and adhesion of bubbles and substances, as well as floc formation. These processes involve the three-phase interaction of solid, liquid and gas. Under the influence of a flotation agent, bubbles adhere to pollutants, forming initial flocs that continue to grow through catching, wrapping and bridging until they cannot withstand fluid shear or reach the separation area (Liu *et al.,* 2021; Amakiri *et al.*, 2022).

Two different types of flotation exist based on the method of bubble generation and the resultant bubble size, namely dissolved gas flotation, which is achieved by means of a gas compressor; and induced gas flotation, which makes use of mechanical, hydraulic or sparging systems (Gamwo *et al.*, 2022; Liu *et al.*, 2021; Abbas *et al.*, 2021). Dissolved gas flotation uses a gas compressor to dissolve gas into the liquid under high pressure, forming microbubbles that adhere to particles and rise to the surface due to reduced density in contrast, IGF directly injects gas into the flotation chamber, producing larger bubbles, which can slightly reduce efficiency compared to smaller bubbles used in DGF (Jimenez *et al.*, 2018).

The efficiency of the flotation process depends mainly on the contaminants to be removed, the liquid phase's density differences, the temperature of the inlet fluid and the size of the oil droplets. The flotation process has been reported to work well for the treatment of produced water with high and low total organic carbon (TOC) concentrations, as well as water with oil, grease and particulates of less than 7% solids (Guerra *et al.*, 2011; Al-Ghouti *et al.*, 2019). A pilot plant study by Mastouri *et al.* (2010) examined the effect of temperature and impeller speed on IGF performance for oil removal from produced water with total dissolved solids (TDS) ranging from 5 to 300 g/L. The study found that IGF could achieve up to 90% oil removal efficiency at high temperatures (80°C) in a single flotation cell without chemicals (Mastouri *et al.*, 2010).

Recent advancements in flotation unit designs, such as multistage configurations and internal swirl mechanisms, have significantly improved gas separation by enhancing the coalescence of oil droplets. These innovations achieve this by creating lower terminal velocities and increasing the frequency of collisions. Furthermore, the addition of separate inlet chambers enhances flotation efficiency by preventing counter current flow and increasing the residence time for bubbles and droplets, leading to more effective separation. Compact Flotation Units operate with capacities ranging from 10 to 900 m^3/h and have a retention time of less than one minute. They generate bubbles between 200 and 300 μm using stable, independent systems, effectively reducing oil concentrations to below 25 ppm and achieving up to 90% removal efficiency. These advancements make CFUs indispensable for modern offshore oil and gas operations, providing efficient and flexible solutions for water treatment (Piccioli *et al.*, 2020)

3.3.3. Hydrocyclone

Hydrocyclones make up more than 90% of current deoiling facilities in offshore oil and gas installations (Durdevic *et al.*, 2017). The technology dates back to 1891, but it became widely used for liquid-liquid separation in 1980. This was due to the challenges of reduced platform space and greater depths in the offshore oil and gas industry (Richerand, 2012). The University of Southampton in the United Kingdom developed this approach, using g-force to separate oil, sand, and water effectively. Richerand (2012) noted that the technology's initial field trials occurred in the Gulf of Mexico, marking it as a groundbreaking method. It was poised to replace standard separation equipment on offshore platforms.

As technology advanced, the hydrocyclone became a compelling solution for enhanced separation. It was particularly well-suited for limited platform space and dynamic offshore structures. The hydrocyclone is insensitive to orientation, mechanically robust, and operates as a pressure vessel. It alleviated concerns about platform movement when filled with liquid (Richerand, 2012; Bram *et al.*, 2020). Furthermore, using liners reduced the equipment's footprint. Liners were introduced inside the vessel instead of the manifold (Richerand, 2012; Abuhasel *et al.*, 2021).

The hydrocyclone structure, shown in Fig. 3.3, consists of a cylindrical chamber connected to a conical body. This leads to the cone apex at the bottom outlet (Jiménez *et al.*, 2018; Amakiri *et al.*, 2022). In operation, a de-oiling hydrocyclone has a mixture of water and oil entering tangentially. This creates two distinct vortices within the cyclone (Bram *et al.*, 2020). The centrifugal force causes denser fluids, like water, to move towards the wall, forming the underflow. Meanwhile, less dense oil migrates towards the inner axis and exits as overflow (Liu *et al.*, 2021; Ghafoori *et al.*, 2022). This gravity-driven separation relies on the density difference between the fluids. Hydrocyclones use rotating flow to enhance the acceleration field, reaching 2000 to 3000 times gravitational acceleration (Bram *et al.*, 2020).

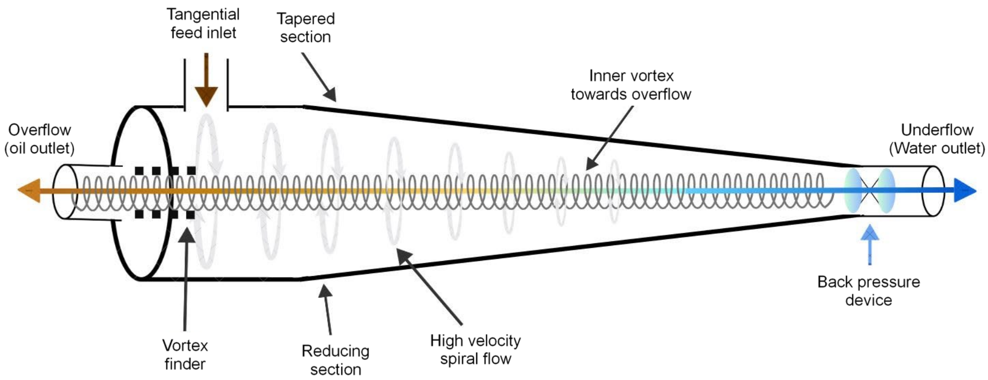


***Figure 3.3**: Illustration of the Modus operandi of deoiling hydrocyclone (retrieved as a combination from Bram et al., 2020; Liu et al., 2021; Amakiri et al., 2022)*

Hydrocyclone dynamics involve the three components of fluid velocity: tangential, axial and radial. Tangential velocity acts perpendicular to the cyclone's wall, generating centrifugal force to separate particles based on density. Radial velocity directs lighter particles towards the overflow and denser particles to the cyclone wall, facilitating separation (Adewove, 2020). Saidi *et al.* (2012) notes that in liquid-liquid hydrocyclones, separation of large droplets begins in the cylindrical section, shifting towards the underflow as droplet size decreases. Studies indicate that centrifugal and drag forces in hydrocyclones are influenced by particle size distribution (Saidi *et al.*, 2012), while some, like Zhang *et al.* (2017), highlight the significance of pressure gradient forces emphasising that larger particles are mainly affected by outward drag and centrifugal force acting radially outwards. Smaller particles experience inward drag, moving towards the cyclone's centre (Zhang *et al.,* 2017).

The terminal slip speed of oil droplets that are dispersed in water can be estimated by Stokes law, with the substitution of the acceleration factor by centripetal acceleration (T^2/r), which is directed towards the hydrocyclone centre axis as shown in Eq. 3.1, where $\Delta\rho$ represents the density difference between the oil and water (kg/m^3), D_d is the oil droplet diameter (m), μ is the dynamic viscosity of water (Pa.s), T is tangential speed (m/s) and r is the radial position inside the hydrocyclone (Bram *et al.*, 2020).

$$U_d = \frac{\Delta\rho D_d^2 T^2}{18\mu r} \qquad \text{(Eq. 3.1)}$$

According to Amakiri *et al.* (2022), under the effects of centrifugal force, radial velocity gradient force, fluid resistance, centrifugal buoyancy and magnus force, hydrocyclones may reduce oil in water concentration from produced water to as low as 20 ppm. Hydrocyclones have the capacity of achieving a 99% removal efficiency for oil droplets higher than 50 µm with the efficiency dropping significantly for droplets of 5 µm or smaller (Igunnu and Chen, 2014; Nasiri and Jafari, 2017; Liu *et al.*, 2021; Amakiri *et al.*, 2022; Ghafoori *et al.*, 2022). A description of centrifugal force (F_c), centrifugal buoyancy (F_b), fluid resistance (F_S), Magus force (F_M) and radial velocity gradient force (F_a) is given in Table 3.1.

***Table 3.1**: Types of effects applied in oil particles during the deoiling process (Adewoye, 2020; Liu et al., 2021)*

Type of Effect	Mathematical Formula	Description
Centrifugal Force	$F_c = m\frac{{u_t}^2}{r} = \frac{\pi d^3 \rho {u_t}^2}{6r}$	Driving force for radial sedimentation
Centrifugal Buoyancy	$F_b = \frac{\pi d^3}{6}\frac{dP}{dr} = \frac{\pi d^3 \rho {u_t}^2}{6r}$	Force caused by the fluid pressure difference between the inside and outside in the radial direction
Fluid Resistance	$F_S = 3\pi\mu d v_r$	Force caused by the relative motion between the droplet and the continuous phase
Magus Force	$F_M = k\rho d^3 \omega v_t$	Lift force perpendicular to the relative velocity direction between the droplet and the continuous phase generated by the droplet rotation
Radial Velocity Gradient Force	$F_a = \frac{\pi}{d}\rho d^3 v_r \frac{dv_r}{dr}$	Force generated by the gradient of the velocity fluctuation in the radial direction
Legend	where m is the mass of droplet (kg), r is the radius (m), u_t is the tangential velocity of the droplet at radius r in (m/s), d is the drop diameter (m), ρ is the droplet density (kg/m^3), P is the pressure at the radius r (Pa), μ is the viscosity of two-phase flow (Pa.s), v_r the relative velocity between the droplet and the continuous phase in the radial direction (m/s), k is the constant coefficient related to droplet size; ω is the rotational angular velocity of the droplet; v_t is the relative tangential velocity between the droplet and the continuous phase (slip velocity) and v_r is the radial velocity of the continuous phase.	

To maintain processing capacity on offshore platforms, multiple cyclone tubes known as liners are installed within a horizontal vessel with hydrocyclones. Efficient operation requires optimizing each tube, but the fixed processing capacity of hydrocyclones is a significant limitation. In the offshore petroleum industry, adjusting equipment capacity involves plugging or unplugging liners, which fails to address real-time changes in produced water volume (Liu *et al.*, 2021).

3.3.3.1. Performance Criteria

Achieving optimal efficiency in a deoiling hydrocyclone is contingent on a well-designed hydrocyclone, along with the right balance of flow rate and flow split. A proper flow rate, indicative of an adequate energy quantity, is essential for establishing a robust gravity field within the hydrocyclone. Additionally, an appropriate flow split is crucial for effective separation, as emphasised by Liu *et al.* (2021).

3.3.3.1.1. Flow Rate

The flow rate versus efficiency relationship in Fig. 3.4 is a common characteristic of deoiling hydrocyclones. When the flow rate reaches the value of minimum flow rate (Q_{min}), the efficiency of the hydrocyclone is at a maximum. If the flow rate is increased to beyond maximum flow rate (Q_{max}) the efficiency will decrease as the retention time is too low for the separation to finish before the mixture leaves the hydrocyclone. The decline in efficiency beyond Q_{max} is due to either a significant increase in droplet break-up due to excessive shear-forces and turbulence or due to a lack of sufficient pressure gradient required to propel the separated oil-core through the overflow. This inadequacy arises from reduced pressure at the Hydrocyclone axis during high flow rates (Husveg *et al.*, 2007).

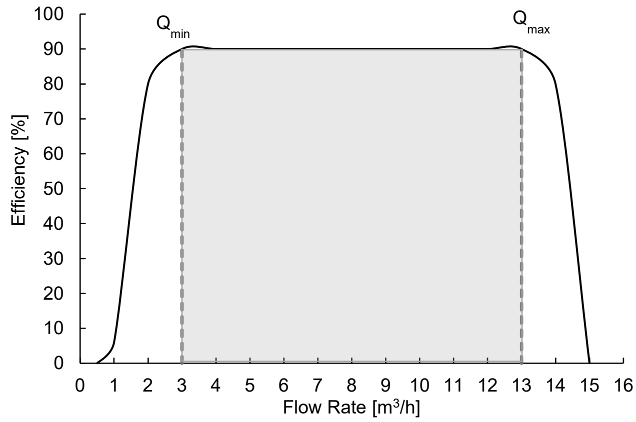


Figure 3.4*: Relationship between the efficiency and flow rate in a hydrocyclone (adapted from Husveg et al., 2007)*

3.3.3.1.2. Flow Split

Maintaining a minimum flow split is essential to ensure that only a minimal amount of oil exits through the underflow. The flow split is defined in Eq. 3.2, where Q_o and Q_i represent the overflow flow rate (m^3/s) and inlet flow rate (m^3/s), respectively. In situations with a small flow split, where the overflow flow rate (i.e., Q_o) and the underflow flow rate (i.e., Q_u) are almost the same, it becomes convenient to use the approximation expressed in Eq. 3.3 (Husveg *et al.*, 2007; Bram *et al.*, 2020):

$$F_S = \frac{Q_o}{Q_i} \quad \text{(Eq. 3.2)}$$

$$F_S = \frac{Q_o}{Q_u} \quad \text{(Eq. 3.3)}$$

Achieving and maintaining an optimal flow split in hydrocyclones is essential for effective oil and water separation. While historical flow splits exceeded 5%, modern deoiling hydrocyclones commonly maintain a range of 2–3% (Husveg *et al.*, 2007). Initially, increasing the flow split enhances separation efficiency, but it eventually plateaus with marginal improvement. Striking a balance is crucial to prevent oil loss through the underflow (Durdevic *et al.*, 2017; Bram *et al.*, 2020). The relationship between flow split and efficiency as described in Fig. 3.5, remains constant for different flow rates, serving as a fundamental factor in hydrocyclone operational control (Durdevic *et al.*, 2017).

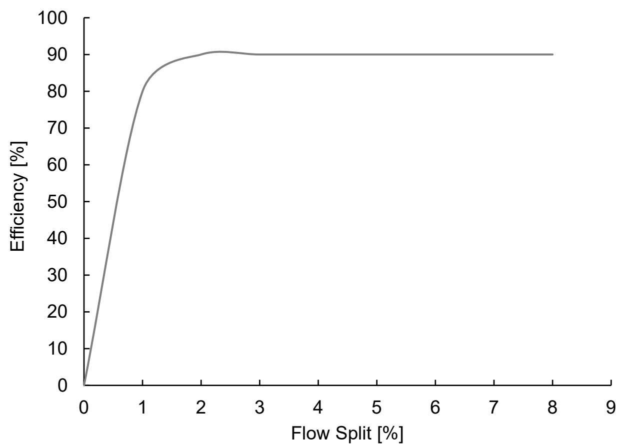


Figure 3.5*: Relationship between the efficiency and flow split in a hydrocyclone (adapted from Husveg et al., 2007)*

3.3.3.2. Operational Control

Deoiling hydrocyclones assume a critical role in navigating the complexities of dynamic operations marked by varying water production rates (Durdevic *et al.*, 2017). According to Husveg *et al.* (2007), it is imperative to ensure that the flow rate remains within the defined range of Q_{min} and Q_{max}, with concurrent adjustments to the flow split in response to rate fluctuations. According to Bram *et al.* (2020), the controller output to the hydrocyclone underflow control valve often incorporates the use of differential pressures corresponding to Q_{min} and Q_{max} as limits in flow rate control, as illustrated by the lower dP_u curve in Fig. 3.6.

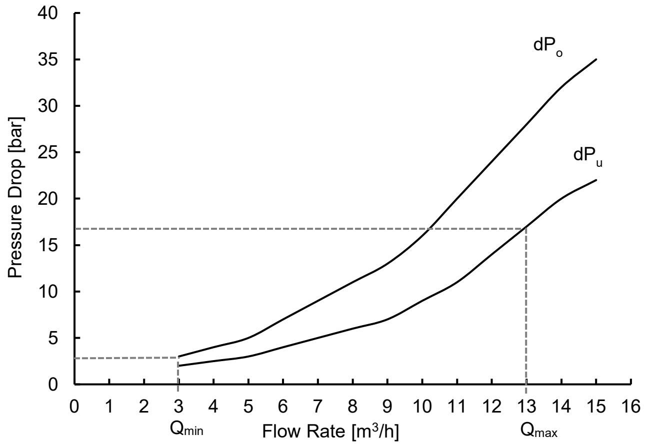


Figure 3.6: *Relationship between the flow rates and the pressure drop in a hydrocyclone (adapted from Husveg et al., 2007)*

Fig. 3.6 illustrates the relationship between inlet-to-overflow and inlet-to-underflow pressure drops (dPo and dPu, respectively) and inlet flow rate under a constant flow split (Husveg *et al.,* 2007). An integral element in this control mechanism is the pressure drop ratio (PDR), as expressed in Eq. 3.4.

$$PDR = \frac{dP_o}{dP_u} = \frac{P_i - P_o}{P_i - P_u} \quad \text{(Eq. 3.4)}$$

Where, P_i (Pa) is the inlet pressure, P_o (Pa) is the overflow pressure and P_u (Pa) is the underflow pressure. Maintaining a consistent PDR as throughput varies, is critical for stabilising the flow split, particularly at flow rates aligned with the efficiency plateau between Q_{min} and Q_{max} (Fig. 3.4), which is the preferred operational range for hydrocyclones (Liu *et al.*, 2021). According to Husveg *et al.* (2007), PDR influences the axial pressure gradient toward the overflow, with a roughly linear relationship between the flow split percentage and the PDR. Typically, deoiling hydrocyclones are configured with the PDR ranging between 1.5 and 3, with the specific value dependent on the size of the hydrocyclone (Husveg *et al.*, 2007).

The direct correlation between hydrocyclone pressure drop and flow rate, conveniently render flow split to be monotonically increasing with PDR. This relationship between PDR and flow split enables control solutions to use PDR obtained through three pressure measurements instead of relying on flow rate measurements, which are less reliable and more expensive than pressure measurements (Bram *et al.*, 2020).

When evaluating the hydrocyclone separation performance, two metrics are commonly used: 1) The concentration reducing separation efficiency (ε_{red}), which describes the percentage of oil-in-water concentration reduction from inlet to underflow as given in Eq. 3.5; and 2) The oil removal efficiency (ε_{oil}), which describes the volumetric percentage of oil introduced at the inlet that exits the overflow as given in Eq. 3.6. C_i, C_o and C_u are the volumetric oil-in-water concentration at the inlet, overflow and underflow, respectively, and Q_i, Q_o and Q_u the flow rates at the inlet, overflow and underflow, respectively (Kharoua *et al.*, 2010; Bram *et al.*, 2020).

$$\varepsilon_{red} = 1 - \frac{C_u}{C_i} \qquad \text{(Eq. 3.5)}$$

$$\varepsilon_{oil} = \frac{C_o Q_o}{C_i Q_i} = 1 - \frac{C_u Q_u}{C_i Q_i} \qquad \text{(Eq. 3.6)}$$

3.3.3.3. Types of Deoiling Hydrocyclones

Deoiling hydrocyclones are critical in the oil and gas industry for separating oil droplets from produced water. They are categorized based on inlet geometry, swirl tube configuration, and flow direction, each impacting their efficiency and suitability for various applications (Van Campen, 2014; Adewoye, 2020; Ekechukwu *et al.*, 2024)

3.3.3.3.1. Inlet Geometry

(1) Tangential inlet hydrocyclones: Enhance oil-water separation efficiency by creating a strong rotational flow that increases centrifugal forces. Ideal for fluctuating feed conditions like offshore platforms, they accommodate varied flow rates and minimise turbulence, reducing wear and maintenance. (2) Axial inlet hydrocyclones: optimise the separation of fine oil droplets by allowing fluid to enter in an azimuthal direction. These hydrocyclones are effective for fine separation but require precise control of feed conditions to maintain efficiency.

3.3.3.3.2. Swirl Tube Configuration

(1) Traditional hydrocyclones: feature conical parts that gradually reduce diameter along the swirl tube to enhance the separation process. These are suitable for general oil-water separation tasks in both onshore and offshore facilities and offer a balance between efficiency and cost-effectiveness, making them a popular choice for many operators. (2) Cylindrical hydrocyclones: maintain a constant diameter throughout the swirl tube, affecting separation

dynamics by lacking the typical conical shape. While these may offer lower separation efficiency compared to traditional designs, their compact form factor is beneficial in space-limited

3.3.3.3.3. Flow Direction

(1) Co-current hydrocyclones: allow both oil and water phases to exit through the downstream side, facilitating straightforward, continuous separation without phase redirection. These are used in applications requiring seamless integration and offer reliable performance across a wide range of operating conditions. (2) Counter-current hydrocyclones: facilitate distinct separation paths by allowing the water to exit downstream and the oil to exit upstream. These are used when enhanced oil and water separation is necessary, especially in systems prioritizing oil recovery.

3.3.3.4. Effect of Geometrical Parameters in Hydrocyclones

The geometrical parameters of hydrocyclones, including the overflow section, inlet section, underflow diameter, overall size, and cone angle, play a critical role in determining separation efficiency. Each parameter affects the fluid dynamics within the hydrocyclone, which in turn impacts its ability to separate oil from water

3.3.3.4.1. Overflow Section

The relationship between the vortex finder and the underflow diameter has traditionally been viewed as a key design parameter. However, Shah *et al.* (2006) challenged this assumption, suggesting that the vortex finder and underflow orifice should be considered independently. This insight is supported by Ghodrat *et al.* (2016) and Ni *et al.* (2019), who emphasised the importance of the vortex finder's diameter in optimizing separation efficiency. Smaller overflow diameters, as noted by Zhao *et al.* (2019), typically improve separation by reducing turbulence and allowing for more controlled fluid flow. The optimal vortex finder length, as highlighted by Chakraborty *et al.* (2018), is critical in maximizing efficiency, but overly long vortex finders can reduce overall efficiency, as confirmed by computational simulations from Tian *et al.* (2019). In general, optimizing the vortex finder size and length is essential for achieving higher separation efficiency in hydrocyclones.

3.3.3.4.2. Inlet Section

The inlet diameter and geometry significantly influence the initial flow conditions, which are vital for effective separation. Larger inlet diameters tend to enhance performance by facilitating smoother fluid entry, as observed by Erikli and Okay (2015). Studies by Vieira and Barroso (2014) and Tang *et al.* (2018) support this, highlighting that the inlet design directly affects the cyclone's ability to generate the necessary centrifugal forces for separation. Dual inlets, as proposed by Osei *et al.* (2015) and Al Kayiem *et al.* (2014), are shown to improve separation efficiency by reducing turbulence and ensuring more uniform flow. Overall, improvements in inlet geometry, particularly the use of dual inlets, can significantly enhance separation performance by stabilizing flow conditions and reducing turbulence.

3.3.3.4.3. Underflow Diameter

The underflow diameter is a critical parameter that influences the separation of oil and water. Ni *et al.* (2019) stressed the difficulty in predicting its effects due to its complex interaction with feed concentration and flow dynamics. Studies by Ghodrat *et al.* (2016) found that larger underflow diameters improve efficiency for high feed concentrations, while smaller diameters are more effective for lower concentrations. However, as underflow diameters increase, efficiency can drop due to larger low-pressure zones forming within the hydrocyclone, as noted by Rakesh *et al.* (2014). In summary, optimizing the underflow diameter is essential for maintaining separation efficiency, particularly in dynamic operational conditions with varying feed concentrations.

3.3.3.4.4. Hydrocyclone Size

The size of the hydrocyclone, particularly its length-to-diameter ratio, has a direct impact on separation efficiency. Brar *et al.* (2015) demonstrated that increasing the length up to an optimal ratio enhances performance by increasing the residence time and allowing more thorough separation. However, beyond this optimal point, efficiency declines due to higher pressure drops. The cylindrical-conical ratio is also important, with the diameter of the cylindrical section being a key factor in determining the size of particles that can be effectively separated (Adewoye, 2020). Larger hydrocyclones may struggle with the separation of smaller particles, as noted by Endres et al. (2012), highlighting the need for balance between hydrocyclone size and particle size.

3.3.3.4.5. Cone Angle

The cone angle is a fundamental design parameter that influences the velocity profile within the hydrocyclone. Cilliers (2000) recommended a cone angle between 15° and 30°, where smaller angles provide better efficiency for finer particle separation and larger angles work better for coarser particles. Vieira and Barroso (2014) and Saidi *et al.* (2013) observed that cone angles larger than 30° reduce separation efficiency due to increased radial velocity, which drags more particles into the overflow. Jiang *et al.* (2011) added that larger cone angles reduce pressure within the cyclone, further impacting separation. Various innovative designs, such as convex cones and tapered shapes (Vakamalla and Mangadoddy, 2019), show promise for improving cut sizes and recovery rates, though their impact on finer particles remains limited. Overall, selecting the appropriate cone angle is crucial for optimizing the separation of different particle sizes, particularly in produced water treatment applications.

3.3.4. Hydrocyclone-Induced Gas Flotation Systems (HIGF Systems)

In the realm of produced water treatment, hydrocyclones play a pivotal role as a primary separation device, often supplanting the need for traditional technologies like corrugated plate interceptors or skimmer vessels (Richerand, 2012). The evolution of hydrocyclones into a practical solution, functioning as pressure vessels and incorporating liners within the vessel to reduce equipment size, has been driven by technological advancements (Liu *et al.*, 2021). To further enhance the effectiveness of gas flotation, there is a recognised need to intensify and increase the probability of three-phase collisions between oil, gas and solid particles post-floc formation.

A common approach in offshore oil and gas, especially in floating facilities, involves synergistically combining hydrocyclone and gas flotation technologies in series as primary and secondary treatment methods, respectively. The combination of these technologies exploits the strengths of both: hydrocyclones provide rapid initial separation, while gas flotation ensures the removal of smaller oil particles that may have escaped the primary stage (Ekechukwu *et al.*, 2024; Gangwar *et al.*, 2024). This integration not only improves the overall oil removal efficiency but also reduces the equipment footprint, which is critical in offshore settings where space is at a premium (Gangwar *et al.*, 2024). The following key aspects highlight the advantages and applications of the HIGF system:

- Sequential Treatment: The hydrocyclone handles the bulk removal of larger oil droplets and solids, reducing the load on the induced gas flotation unit. This sequential approach enhances overall efficiency and extends the operational life of the induced gas flotation unit (Gangwar *et al.*, 2024).

- Compact Design: Which is particularly beneficial for offshore platforms where space is limited (Ekechukwu *et al.*, 2024). HIGF systems can be designed to fit within the constrained space while maintaining high treatment efficiency.
- Energy Efficiency: Hydrocyclones require minimal energy, as they rely primarily on the feed pressure to create the necessary centrifugal force. Induced gas flotation, while requiring energy for gas generation, can benefit from the reduced load provided by the hydrocyclone, leading to overall energy savings (Amakiri *et al.*, 2022; Ghafoori *et al.*, 2022).
- Enhanced Performance: The use approach of using primary and secondary treatment, yields higher oil removal efficiencies and lower residual oil concentrations, which is ideal for applications where stringent environmental regulations must be met (Gangwar *et al.*, 2024).

3.3.5. Qualitative and Quantitative Assessments of Hydrocyclone and Gas Flotation Systems

Table 3.2 outlines the features, advantages, disadvantages and waste streams associated with hydrocyclone and gas flotation technologies in the treatment of produced water. From separation mechanisms to energy consumption, chemical use and waste stream composition, the comparison provides insights for selecting the most suitable technology based on specific water characteristics and treatment requirements.

Table 3.2: *Comparative analysis of hydrocyclone and gas flotation technologies for produced water treatment (adapted as a combination of Igunnu and Chen, 2014; Nasiri and Jafari, 2017; Jimenez et al., 2018; Liu et al., 2021; Amakiri et al., 2022; Gamwo et al., 2022; Ghafoori et al., 2022)*

Criteria	Hydrocyclone	Gas Flotation
Separation Mechanism	Separation of oil occurs freely through the centrifugal force created by the pressurised tangential input of the incoming stream.	Oil particles undergo separation from water by adhering to induced gas bubbles, causing them to ascend to the surface for recovery.
Feasibility	Effective for treating all varieties of produced water, regardless of total dissolved solids (TDS), organic content and salt concentrations.	Suitable for treating produced water characterised by elevated total organic carbons (TOC) and particulate content, with solids concentration exceeding 7%.
Efficiency	90-95% for oil droplets >20 μm	80-85% for oil droplets between 5 to 25 μm
Energy Consumption	Low, relies on feed pressure	Moderate to high, requires gas generation
Chemical Use	None.	May require coagulants
Pre/Post-Treatment	Pre-treatment is unnecessary, but post-treatment may be necessary to eliminate additional contaminants from the outlet water.	Post-treatment is unnecessary, though coagulation may be needed as a preliminary treatment step.
Waste Stream	Slurry containing suspended particles.	Oil with suspended solids and gases.

Advantages	• Operates without the need for chemicals and energy. • Achieves high recovery of treated water. • Capable of lowering oil and grease concentrations to 10 ppm. • Uses compact modules. • Effectively handles produced water with elevated oil content.	• Nearly 100% water recovery. • No need for post-treatment. • Absence of moving parts. • High efficiency attributed to coalescence. • Simple and user-friendly to operate. • Exhibits robustness and durability. • Provides separation effects on suspended solids.
Disadvantages	• Inlet blockage by solids and scale formation may result in additional cleaning costs. • Disposal is necessary for the secondary waste produced. • Energy is needed to pressurise the inlet. • Incapable of removing solids, phenols, and aromatic hydrocarbons. • Susceptible to fouling. • High maintenance costs. • Limited range of optimal operating parameters. • The effectiveness of separation is influenced by blockage and wear of one or more cyclone tubes.	• Unsuitable for high-temperature feed water. • Disposal of sludge is necessary for managing generated solids. • Results in the generation of a significant amount of gas. • Separation is Influenced by the retention time. • Secondary pollution (i.e., scum), due to the use of chemical agents. • The efficiency of separation is affected by the volume of the injected gas.

Fig. 3.7 provides a graphical representation of the influent produced water's oil-in-water and TSS concentration ranges that can be effectively treated by a hydrocyclone as primary treatment, and induced gas flotation as the secondary treatment method. Polishing methods are usually employed for water treated for reuse (Larson, 2018).

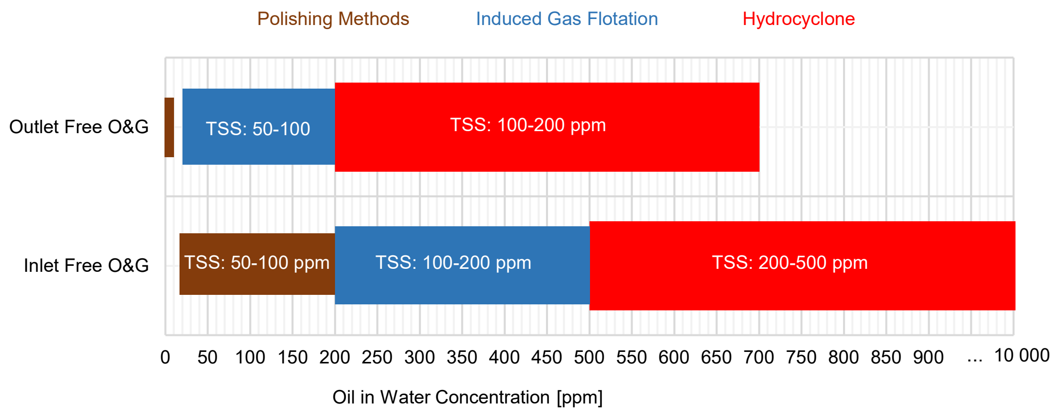


Figure 3.7*: Influent and effluent oil and grease concentration by treatment method (adapted from Larson, 2018)*

Amakiri *et al.* (2022) outlined key criteria for evaluating produced water treatment technologies as shown in Table 3.3. The hydrocyclone and flotation technologies were scrutinised based on various performance parameters and both exhibited similar ratings. However, the hydrocyclone outperformed flotation in terms of robustness, receiving a higher score, while both showed equal consistency ratings. In terms of flexibility to treat a diverse range of wastewater, the hydrocyclone demonstrated intermediate flexibility, whereas flotation had a comparatively lower score. These assessments provide nuanced insights into the comparative performance of hydrocyclone and flotation technologies in wastewater treatment applications, which could serve as a basis for technology selection throughout the oil and gas industry.

Table 3.3*: Qualitative assessment of hydrocyclone and gas flotation technologies (Amakiri et al., 2022)*

Criteria	Hydrocyclone [Primary Treatment]	Flotation [Secondary Treatment]
Waste Generation	●●○○	●●○○
Chemical Demand	●○○○	●○○○
Energy Demand	●●○○	●●○○
Adaptability for Easy Upgrades	●○○○	●○○○
Durability in Tough Conditions	●●●●	●●○○
Consistency / Low-Failure Operation	●●●○	●●●○

Effective in the Treatment of Diverse Water Compositions	●●○○	●○○○

●○○○ Low ●●○○ Average ●●●○ High ●●●● Very High

3.4. Produced Water Regulations for Environmental Discharge

In offshore operations, it is standard practice to release treated produced water into the sea. The primary treatment goal is to decrease both dissolved and suspended oil to acceptable levels depending on location and minimise the toxic effects on marine fauna and flora (Gamwo *et al.*, 2022). Oil and grease within offshore produced water has garnered significant attention and evolved into pivotal regulatory standards for lawmakers, while other hazardous substances have been largely disregarded (Dwyer, 2016). The concentration of dissolved oil in produced water is contingent on the production volume, the enhanced oil recovery method, the age of the well, as well as the characteristics of the predominant hydrocarbon in the oil. While the concentration of the suspended oil is influenced by the density of the oil, the historical shear forces acting on the droplets, the amount of oil precipitation and the interfacial tension between the water and oil (Nasiri *at al.*, 2017; Amakiri *et al.*, 2022).

The discharge of oil and grease into marine environments has far-reaching ecological implications. It poses a significant threat to marine life, as these contaminants are toxic to organisms, hindering their respiration, feeding and reproductive processes (Nasiri *et al.*, 2017; Amakiri *et al.*, 2022). According to Hedar (2018), the settling of oil and grease on the seafloor leads to habitat degradation, affecting benthic species like worms and clams. These pollutants enter the food chain, accumulating in marine organisms and posing health risks to humans who consume contaminated seafood. Additionally, the introduction of these contaminants can disturb the balance of marine ecosystems, potentially leading to shifts in species composition and diversity, further exacerbating the ecological disruption (Gazali *et al.*, 2017).

Growing environmental awareness has prompted regulatory bodies in many countries to impose stricter limits on the discharge of oil and grease (Jimenez *et al.,* 2018). This underscores the significance placed on safeguarding aquatic ecosystems. China stands out with the strictest regulations, setting a minimum oil in water discharge allowance of 10 ppm, emphasising a commitment to marine ecosystem protection. Conversely, higher allowances of 100 ppm are often associated with oil-producing nations like Egypt, Morocco, Tunisia and the Syrian Arab Republic, reflecting economic considerations in detriment to potentially higher environmental risks. The allowances exhibit wide variability globally, reflecting diverse levels of environmental protection measures. Regional clusters, like European countries with similar allowances around 30 ppm, highlight the influence of regional agreements aligning with the European Union's commitment to stringent environmental standards (Neff *et al.*, 2011; Igunnu and Chen, 2014; Jimenez *et al.*, 2018*;* Amakiri *et al.,* 2022).

Evolving regulations reveal a recent trend of tightening standards, driven by a heightened awareness of oil and grease's detrimental impact on marine life (Gamwo *et al.*, 2022). According to Jimenez *et al.* (2018), this shift is reinforced by increasing transparency in reporting and monitoring efforts to responsibly manage oil and grease disposal for the preservation of marine environments. In addition to legislation, many water-stressed countries with oilfields are looking for ways to supplement their limited freshwater resources by focusing on efficient and economical methods to treat produced water, so that it can be channelled to agricultural and industrial uses (Igunnu and Chen, 2014).

3.5. Future Research Directions

The treatment of produced water using hydrocyclones and induced gas flotation systems has experienced significant advancements. However, continuous research and development are crucial to address existing limitations and enhance the efficiency and applicability of these technologies. Future research need to focus on optimising hydrocyclone design by investigating the impact of varying cone angles, inlet configurations and vortex finder dimensions to improve fluid dynamics and separation performance. Additionally, measuring and benchmarking hydrocyclone models against various droplet sizes can enhance accuracy (Ni *et al.*, 2019; Bram *et al.*, 2020).

Exploring the development of wear-resistant and anti-fouling materials for hydrocyclone and induced gas flotation's construction can significantly improve efficiency and automation. This is particularly vital for produced water treatment on offshore platforms, ensuring these systems remain green, efficient, secure and capable of handling challenges posed by deep and ultra-deep-water exploration (Liu *et al.*, 2021). Research needs also to be conducted on verification of operational control systems in full-scale hydrocyclone optimisation programs to ensure they meet performance criteria and maintain efficiency. Developing reliable sensors and monitoring systems that provide real-time data on deoiling performance is essential and by utilising artificial intelligence and machine learning algorithms to predict system behaviour and optimise operating conditions could significantly improve reliability (Durdevic *et al.*, 2017; Tian *et al.*, 2018; Liu *et al.*, 2021).

Further research on optimisation of gas flotation techniques by increasing the collision and adhesion efficiencies of oil droplets and gas bubbles is another important area. This could be achieved by developing new means of generating smaller gas bubbles, promoting oil droplet coalescence with coagulants and enhancing fluid path designs to maximise bubble-droplet interactions (Saththasivam *et al.*, 2016; Liu *et al.*, 2021)

3.6. Conclusion

Produced water treatment remains a critical environmental and operational challenge in the oil and gas industry. This chapter highlighted major advancements in hydrocyclone and induced gas flotation technologies, which offer effective solutions for offshore produced water management. Hydrocyclones efficiently remove larger oil droplets (>20 μm) with low energy use but are less effective for smaller droplets. IGF systems excel at removing fine droplets (down to 5 μm) but require more energy and are prone to fouling. Combining both in hybrid hydrocyclone-induced gas flotation (HIGF) systems merges their strengths, enhancing oil removal while reducing energy consumption.

Global variations in oil-in-water discharge limits (10–100 ppm) underscore the need to further improve treatment efficiency to meet stricter standards. Future research should focus on optimizing designs, improving energy use, and integrating advanced control systems using AI and machine learning to ensure compliance and reduce ecological impacts.

CHAPTER 4
PHYSICOCHEMICAL CHARACTERISATION OF OILFIELD PRODUCED WATER AND ITS TREATMENT CONSTRAINTS

César, S.D., De Jager, D. and Njoya, M., 2025. Physicochemical Characterisation of Oilfield Produced Water and its Treatment Constraints. *Petroleum Research*. (Article in Press)

"Blessed are those who hunger and thirst for righteousness, for they shall be filled."

Matthew 5:6

CHAPTER 4: PHYSICOCHEMICAL CHARACTERISATION OF OILFIELD PRODUCED WATER AND ITS TREATMENT CONSTRAINTS

4.1. Introduction

Produced water is the primary waste stream from oil and gas operations, presenting major environmental and operational challenges because of its complex and highly variable chemistry. It includes naturally occurring formation water from the reservoir as well as injection water used to boost recovery and support well operations (Olajire, 2020; Nasiri and Jafari, 2017; Zheng *et al.*, 2016). Its contaminants including hydrocarbons, heavy metals, salts, and naturally occurring radioactive materials (NORMs), present major ecotoxicological risks and complicate treatment and disposal (Jimenez *et al.*, 2018).

Globally, produced water volumes rose from about 139 billion barrels in 2017 to 170 billion barrels in 2022, driven by intensified oil and gas activities (Energy Institute, 2023). While regions such as Nigeria, Malaysia, India, China, and the Gulf of Mexico have been well studied (Amakiri *et al.*, 2023; Razak *et al.*, 2022; Jiang *et al.*, 2022), there is a notable lack of publicly available data on produced water from Angola's offshore oilfields. Most existing information is confined to internal industry reports, limiting broader analysis.

Angola, as one of Africa's leading oil producers with significant offshore operations, generated about 5 billion barrels of produced water in 2022 alone (Energy Institute, 2023; César, 2022). However, detailed studies on its produced water composition remain scarce.

This chapter addresses this gap by characterizing untreated produced water from Angola's offshore Block 15 over a 50-day sampling period, analyzing more than 30 key constituents. Its objectives were to provide a detailed profile of the physicochemical and organic composition, assess biodegradability and the feasibility of biological treatment and highlight implications for treatment strategies and regulatory compliance. By establishing this foundational reference, the chapter aims to guide future research and inform water management efforts in Angola's offshore oil sector, ultimately supporting the development of more effective treatment solutions and promoting sustainable produced water management in the region.

4.2. Materials and Methods

4.2.1. Site Description

The study was conducted at an undisclosed FPSO operating in the Angolan offshore Block 15 region that has been in operation for a decade (Fig. 4.1). After extraction from the offshore wells, the hydrocarbon mixture is dosed with an emulsion-breaking chemical and routed to a three-phase separator. The separated oil is washed and dehydrated, the gas is compressed for use as fuel gas or gas injection and the water is duly treated for overboard discharge into the marine environment.

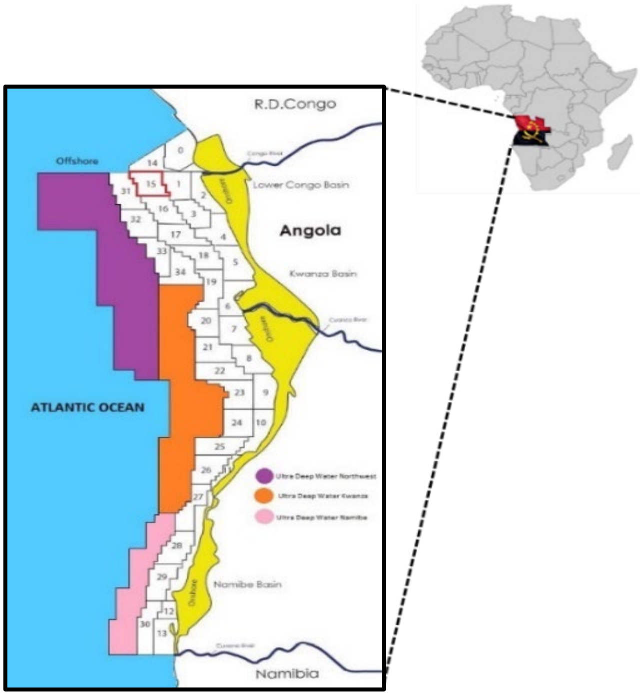


***Figure 4.1**: Angolan offshore oilfield blocks (adapted from Cauxeiro et al., 2022)*

4.2.2. Sampling

Untreated produced water samples were collected from the high-pressure three-phase separator's water outlet at 6:00 AM, 12:00 PM, and 6:00 PM over a 50-day period. This timeframe was selected to capture diurnal variations in production operations, separator performance and chemical dosing cycles. Sampling times were chosen to ensure that both peak and off-peak operational conditions were represented, providing a comprehensive dataset for produced water characterisation. The 50-day duration aligns with similar studies in offshore environments, allowing for a robust assessment of variability in produced water composition. Samples were collected in chemically clean glass and polyethylene bottles, which were rinsed with produced water before filling and sealed with Teflon-lined lids to prevent contamination.

The specific bottles were chosen based on the type of analysis as follows:

- Glass bottles of 1 L: Used for Benzene, Toluene, Ethylbenzene and Xylene (to prevent the loss of volatile organic compounds), Polycyclic Aromatic Hydrocarbons, Total Organic Carbon, Dissolved Organic Carbon (to avoid contamination and adsorption of organic compounds), Total Oil and Grease, Total Petroleum Hydrocarbons (for chemical inertness and resistance to solvents), and Biochemical Oxygen Demand over five days (for 5-day storage).
- polyethylene bottlesof 500 mL: Used for pH, Conductivity, Dissolved Oxygen (measured immediately without long-term storage), turbidity (unaffected by container material), alkalinity (for titration-based methods), Total Dissolved Solids, Total Suspended Solids, Chemical Oxygen Demand, phenol, anions, and metal concentrations (for acidified samples to prevent metal contamination).

For in-house laboratory analysis, samples were immediately analysed upon arrival. For external laboratory analysis, samples were transported under refrigerated conditions (4°C) every Mondays, Wednesdays and Fridays.

4.2.3. Produced Water Analyses

Produced water analyses were conducted using standardised analytical techniques, ensuring compliance with regulatory and scientific standards. pH, conductivity, and dissolved oxygen (DO) were measured using a Hach® HQ4300 portable multi-meter with PHC20101, CDC40101, and LDO10105 probes, following EPA methods 9040C, 2510B, and 360.2, respectively. Turbidity was determined using a Hach® 2100Q turbidimeter (EPA 180.1), while alkalinity was quantified through titration (EPA 2320B).

Gravimetric analysis was employed for Total Dissolved Solids (TDS) and Total Suspended Solids (TSS) using 0.15 µm filters (EPA 2540C and 2540D). Chemical Oxygen Demand (COD) and phenol concentrations were determined using a Hach® DR3900 spectrophotometer with EPA 410.4 (COD) and EPA 420.1 (phenols). Biochemical Oxygen Demand (BOD_5) was measured with a Hach® HQ4300 meter using probe LBOD101 (EPA 5210B), ensuring accurate quantification of biodegradable organic matter.

Total Organic Carbon (TOC) and Dissolved Organic Carbon (DOC) were analysed using a Shimadzu® TOC-V CSH analyser (EPA 415.3, 5301). Total Oil and Grease (TOG) and Total Petroleum Hydrocarbons (TPH) were measured via n-hexane liquid-liquid extraction followed by infrared spectroscopy (EPA 1664B), with silica gel treatment to separate non-polar fractions.

Benzene, Toluene, Ethylbenzene and Xylene (BTEX) analysis was conducted using a Tekmar–Dohrmann® Velocity XPT purge-and-trap system coupled with an Agilent® 7890 gas chromatograph (GC) and an Agilent® 5977A mass spectrometer (MS) (EPA 8260C). Total Polycyclic Aromatic Hydrocarbons (PAHs) were quantified by liquid-liquid extraction followed by GC-MS analysis (EPA 3510C and 8270C). TOC and DOC (the latter using a 0.45 µm filter) were analysed using a Shimadzu® TOC-V CSH analyser in line with EPA 415.3 and 5301. Metal concentrations were measured in acidified samples by means of inductively coupled plasma (ICP) - optical emission spectrometry (OES) using an Agilent® 5800 ICP-OES, based on EPA 200.7. Anions such as bromide, chloride and sulfate were determined with a Thermo Scientific® Dionex ICS 3000 SP ion chromatography system, following EPA 300.1.

4.3. Results and Discussion

4.3.1. Inorganic Profile

Table 4.1 represents the concentrations of metals and anions in 150 produced water samples, averaged across three daily measurements over a 50-day period. The concentration of aluminum (Al) in the produced water averaged 3.07 mg/L, ranging from a minimum of 2.00 mg/L to a maximum of 4.74 mg/L. Calcium (Ca) concentration, averaging 13.24 mg/L, was relatively stable with a standard deviation of 2.86 mg/L and a range of 10.40 mg/L to 18.97 mg/L. Iron (Fe) concentration averaged 4.10 mg/L and ranged from 1.20 mg/L to 6.59 mg/L and the nickel (Ni) concentration exhibited a wide range (28.00 mg/L to 621.24 mg/L), with high variability reflected in the 25th and 75th percentiles of 141.93 mg/L and 456.15 mg/L, respectively. Elevated concentrations of nickel suggest contamination from corrosion of nickel-based alloys used in production equipment (Neff *et al.*, 2011). The sodium (Na) concentration was exceptionally high, averaging 1729.09 mg/L with a range of 1167.77 mg/L to 2100.00 mg/L. The high standard deviation of 325.14 mg/L and median concentration of 1796.11 mg/L suggest either seawater intrusion or the inherent high salinity of the reservoir formation (Jiang *et al.*, 2022).

***Table 4.1**: Results of comprehensive elements analyses on the produced water from Angola block 15*

Elements/Ions	Average [mg/L]	Range [mg/L]
Aluminum	3.07 ± 1.03	2.00 – 4.74
Barium	2.63 ± 0.8	1.42 – 3.82
Calcium	13.24 ± 2.86	10.40 – 18.97
Chromium	1.80 ± 0.72	0.80 – 3.00
Copper	1.37 ± 0.68	0.60 – 2.58
Iron	4.10 ± 1.67	1.20 – 6.59
Potassium	214.12 ± 121.53	45.00 – 415.22
Magnesium	44.99 ± 8.5	30.66 – 61.00
Sodium	1729.09 ± 325.14	1167.77 – 2100.00
Nickel	314.07 ± 179.33	28.00 – 621.24
Lead	0.66 ± 0.54	0.00 – 1.88
Silicon	67.06 ± 17.08	39.00 – 101.31
Strontium	35.63 ± 23.22	14.00 – 83.08
Zinc	0.67 ± 0.31	0.14 – 1.22
Bromide	101.46 ± 20.91	71 - 137.91
Chloride	1206.72 ± 538.33	400 - 2146.84
Sulfate	151.49 ± 66.87	64 - 271.91

Bromide ranges from 71 to 137.91 mg/L, showing moderate variability. Chloride, with concentrations between 400 and 2146.84 mg/L, exhibits the highest variability, indicating significant fluctuations. Sulfate levels range from 64 to 271.91 mg/L, displaying moderate variability with occasional peaks.

4.3.2. Physicochemical Profile

The statistical results of the physicochemical and organic characteristics of the analysed oilfield produced water are presented in Table 2 and the graphical representation of the results are illustrated in Fig. 4.2 and Fig. 4.3.

***Table 4.2**: Statistical results of the physicochemical characteristics of the analysed produced water*

	Mean	Min	Max	25th Percentile	50th Percentile	75th Percentile
pH	6.4 ± 1.1	4.8	8.2	5.5	6.3	7.1
Conductivity [mS/cm]	28 ± 3	24.8	32.7	25.4	28.7	30.6

DO [mg/L]	4.3 ± 0.6	3.1	5.5	3.8	4.2	4.8
Turbidity [NTU]	140 ± 84	29.0	276.5	50.7	148.9	204.1
Alkalinity [mg/L $CaCO_3$]	325 ± 169	81.0	577.0	168.1	335.2	478.9
TDS [mg/L]	14 256 ± 1 376	11 948	16 582	13 242	14 117	15398
TSS [mg/L]	346 ± 55	314.0	502.8	314.0	314.0	351.5
COD [mg/L]	471 ± 49	394.0	526.0	423.3	466.2	526.0
BOD_5 [mg/L]	13.2 ± 3.9	9.8	20.4	9.8	11.1	16.3
Phenol [mg/L]	3.4 ± 0.4	2.7	4.1	3.1	3.3	3.8
BTEX [mg/L]	4.8 ± 1.4	2.8	7.0	3.6	4.4	5.9
PAH [mg/L]	0.56 ± 0.35	0.2	1.2	0.2	0.56	0.91
TOC [mg/L]	555 ± 200	248.0	918.5	415.2	577.4	678.6
DOC [mg/L]	526 ± 201	210.0	863.6	362.9	475.9	697.7
TOG [mg/L]	96 ± 71	19.0	206.7	19.0	111.7	160.5
TPH [mg/L]	54 ± 28	12.0	105.0	27.2	55.3	75.4

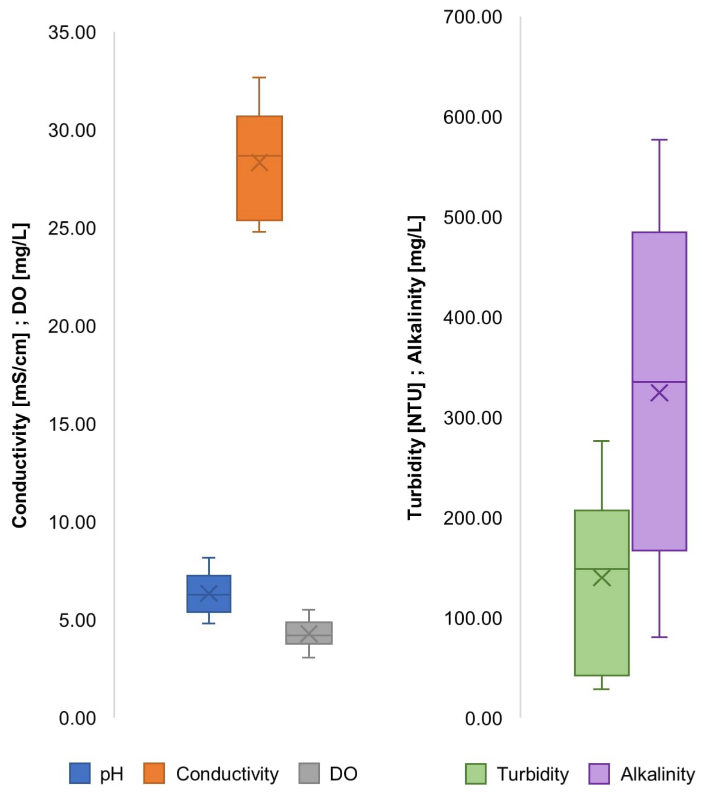


***Figure 4.2**: pH, conductivity, dissolved oxygen, turbidity and alkalinity in produced water*

The pH measurements ranged from 4.81 to 8.17 with an average of 6.36, indicating a slightly acidic to neutral environment overall. The conductivity ranged from 24.8 mS/cm to 32.67 mS/cm, with a mean of 28.34 mg/L and the dissolved oxygen (DO) values ranged from 3.09 mg/L to 5.52 mg/L, with a mean of 4.3 mg/L. Turbidity and alkalinity ranged from 29 NTU to 276.47 NTU and 81 mg/L $_{CaCO_3}$ to 577.04 mg/L $_{CaCO_3}$, with an average of 139.93 NTU and 324.7 mg/L $_{CaCO_3}$, respectively.

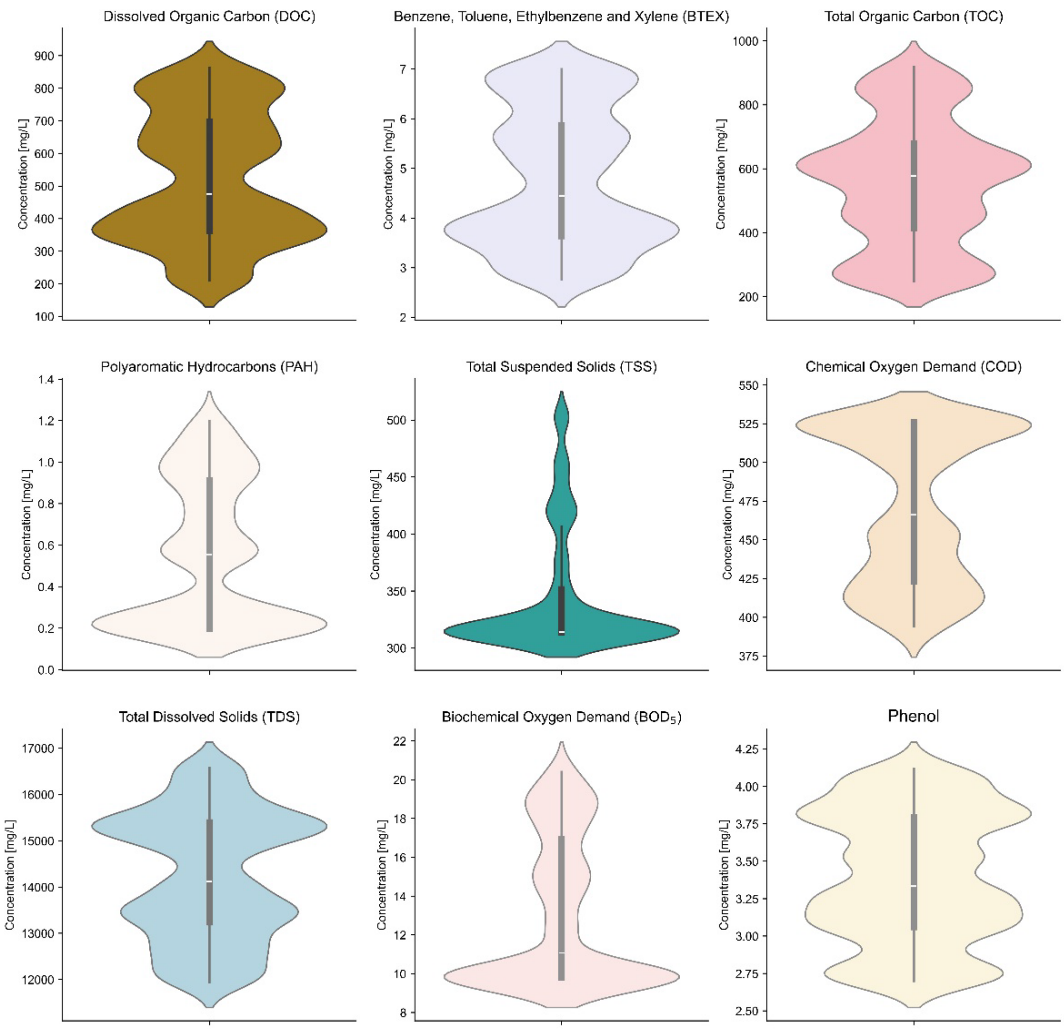


***Figure 4.3**: TDS, TSS, COD, BOD$_5$, phenol, PAH, BTEX, TOC and DOC in produced water*

Total dissolved solids (TDS) exhibited an average concentration of 14,256 mg/L and a median of 14,117.07 mg/L. The values ranged from a minimum of 11,948.38 mg/L to a maximum of 16,582.19 mg/L, with a spread of 4,633.81 mg/L. Total suspended solids (TSS), which may include drifting particles such as sand, silt, algae, plankton, and sediments, ranged from 314.00 mg/L to 502.81 mg/L, with an average of 345.89 mg/L. The distribution was slightly right-skewed and exhibited heavier tails than a normal distribution, suggesting occasional high-concentration outliers.

Chemical oxygen demand (COD) varied between 394.02 mg/L and 526.00 mg/L, with an average of 470.82 mg/L, indicating a significant presence of organic and inorganic contaminants. Phenol concentrations ranged from 2.70 mg/L to 4.12 mg/L, averaging 3.40 mg/L, highlighting the presence of toxic organic compounds requiring effective treatment. Biochemical oxygen demand (BOD_5) values ranged from 9.80 mg/L to 20.40 mg/L, with a mean of 13.20 mg/L, suggesting relatively low levels of biodegradable organic matter.

Total organic carbon (TOC) levels ranged from 248.00 mg/L to 918.45 mg/L, with an average of 554.58 mg/L. The wide distribution of TOC, reflected in the high standard deviation and range, suggests substantial variability in organic content. The combined concentration of benzene, toluene, ethylbenzene, and xylene (BTEX) ranged from 2.77 mg/L to 7.00 mg/L, with an average of 4.79 mg/L, indicating a relatively low presence of these harmful volatile organic compounds. Polycyclic aromatic hydrocarbons (PAHs) were detected at concentrations between 0.20 mg/L and 1.20 mg/L, with an average of 0.56 mg/L, suggesting the presence of these persistent organic pollutants at low to moderate levels.

Fig. 4.4 represents the distribution of total oil and grease (TOG) and total petroleum hydrocarbons (TPH) concentrations in the produced water with respective trends observed for the set of samples analysed in the form of a scatter plot. Light and dark shades indicate areas with lower and higher densities, respectively.

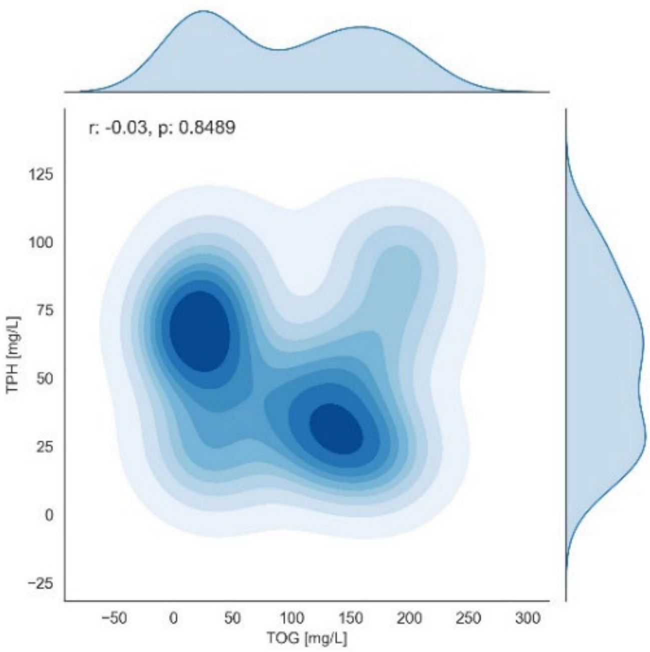


Figure 4.4: *Density plot of Total Oil and Grease vs Total Petroleum Hydrocarbons*

For TOG (average: 96.08 mg/L), the distribution indicates two distinct groups within the data, with one group clustered at a low value of 19.00 mg/L and the second group spread out across higher values. The distribution of TPH is more uniform compared to TOG, with a central tendency around the average value of 54.04 mg/L. The Pearson correlation analysis ($r = -0.03$, $p = 0.849$) between TOG and TPH suggests no statistically significant linear relationship, indicating that these two variables behave independently in Angola’s produced water. This unexpected result raises critical questions about hydrocarbon speciation and why TOG and TPH do not correlate as expected. One possible explanation lies in fractional composition differences, as TOG encompasses both non-polar and polar hydrocarbons, while TPH quantifies only non-polar fractions, excluding oxygenated hydrocarbons such as acids and alcohols, which may be prevalent in Angola’s produced water. Another contributing factor could be oil-water emulsions, where poorly dispersed oil droplets are captured in TOG analysis but may not be fully extracted in TPH measurements, particularly if emulsifiers influence the TOG-to-TPH ratio. Additionally, operational conditions such as hydrocyclone efficiency and separation performance may play a role, with small, stabilised oil droplets persisting in TOG while heavier hydrocarbons are effectively removed, reducing TPH concentrations (ASTM International, 2022)

To contextualise the findings within a broader scope, a comparative analysis was performed using data from offshore oilfields worldwide. The results, illustrated in Fig. 4.5 as a Principal Component Analysis (PCA) biplot, provide insights into the relationships between key water quality parameters. Due to data availability constraints, the analysis was limited to Phenol, COD, BTEX, PAHs, TDS, and Oil and Grease.

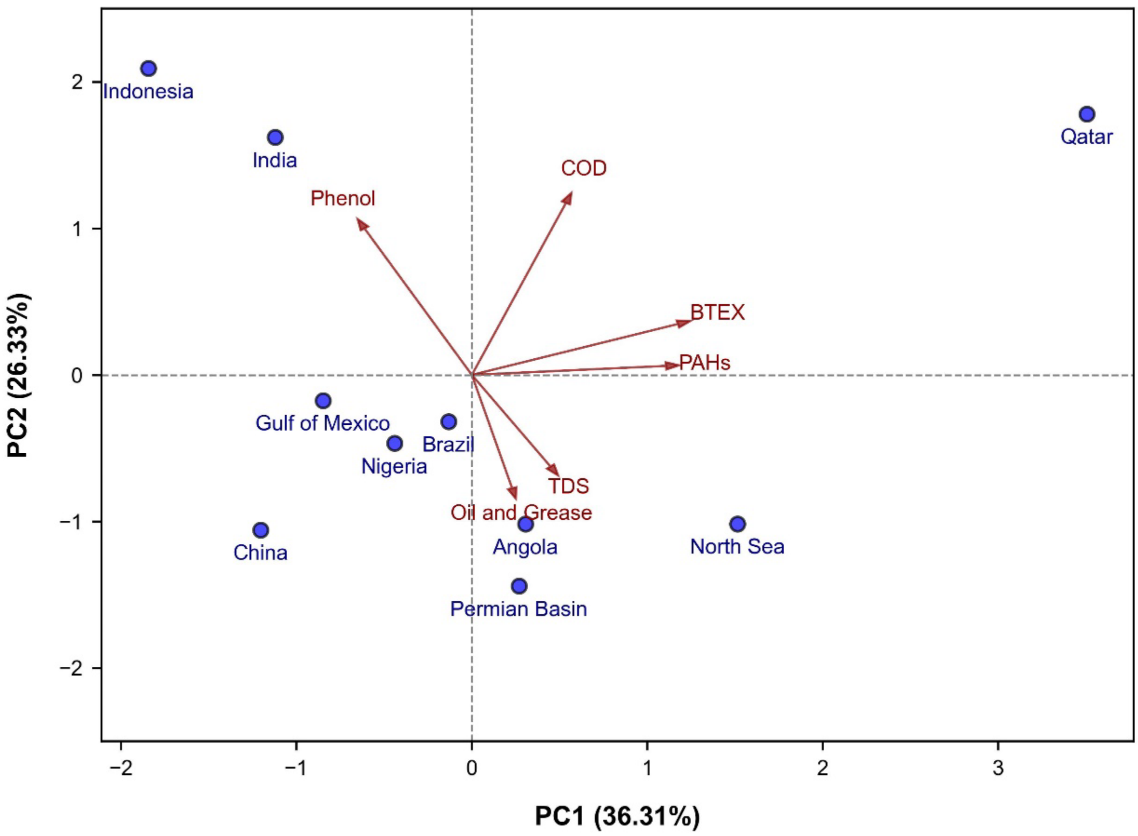


Figure 4.5*: Principal Component Analysis (PCA) of Produced Water Composition (Amakiri et al., 2023; Herawati et al., 2023; Amakiri et al., 2022; Hu et al., 2022; Jiang et al., 2022; Kumar et al., 2022; Effendi et al., 2022; Eldos et al., 2022; Suzuki and Alhadeff, 2022; Gul et al., 2021; Lahlou et al., 2020; Al-Kaabi et al., 2019; Janson et al., 2015; Ahan, 2014; Pampanin and Sydnes, 2013; Neff et al., 2011; Gabardo et al., 2011; Dórea et al., 2007; Lu et al., 2006; Ji et al., 2002; Fillo et al., 1992)*

The PCA biplot provides a comparative assessment of produced water composition across different oil-producing regions by reducing the dataset into two principal components (PC1 and PC2) to highlight dominant patterns. PC1 accounts for 36.31% of the variance, while PC2 explains 26.33%, together capturing 62.64% of the total variability. This indicates that these two components effectively summarise the main differences in produced water characteristics for the selected parameters.

Regions positioned closer together on the biplot exhibit similar produced water profiles, whereas those further apart display distinct compositional differences. Notably, the North Sea and the Permian Basin show strong similarities, characterised by their salinity-dominant profiles. In contrast, India and Qatar occupy opposite ends of the spectrum, reflecting significant differences in composition. Angola, meanwhile, is positioned in an intermediate region, suggesting that its produced water characteristics share similarities with multiple regions while retaining distinct features.

While Angola's salinity (TDS: 14,256 mg/L) is moderate compared to highly saline regions such as the North Sea (229,860 mg/L) and the Permian Basin (128,641 mg/L), it still necessitates some level of desalination, albeit less intensively than these regions. Oil and grease contamination (96.08 mg/L) is a major concern, positioning Angola alongside the Permian Basin (104.9 mg/L) and China (81.43 mg/L) in terms of oil content. This highlights the importance of mechanical separation technologies, such as hydrocyclones and induced gas flotation (IGF), to effectively reduce oil and grease concentrations.

The high total oil and grease (TOG) and total petroleum hydrocarbons (TPH) levels further underscore the necessity of mechanical separation processes to prevent hydrocarbon contamination of marine environments. Without effective removal, residual oil could lead to chronic water pollution, impacting aquatic biodiversity and disrupting food chains. Additionally, Angola exhibits moderate levels of organic contaminants (BTEX: 4.79 mg/L, COD: 470.82 mg/L), suggesting that biological treatment and advanced oxidation processes could be beneficial in managing these pollutants. Although phenol levels (3.40 mg/L) are relatively low, they are not negligible. To ensure compliance with environmental discharge limits, adsorption techniques such as activated carbon filtration or membrane-based separation may be beneficial as a final polishing step in the treatment process.

4.3.3. Biodegradability Assessment

The organic strength of wastewater is commonly assessed using COD and BOD_5. COD measures the total organic matter concentration in the water, while BOD_5 evaluates the biodegradable portion of such organic matter. The biodegradability index (BOD_5 to COD ratio) indicates how readily organic pollutants can be broken down biologically. A ratio greater than 0.6 suggests high biodegradability, between 0.3 and 0.6 indicates moderate biodegradability, and less than 0.3 implies low biodegradability with a significant amount of non-biodegradable matter present (Lacalamita *et al.*, 2024; Choi *et al.*, 2017; Henze *et al.*, 2002).

Due to the empirical correlation between BOD_5 and COD, COD measurements can be used to approximate BOD_5 values for a given sample, although this correlation is specific to each site (Henze *et al.*, 2002). COD testing is advantageous because it delivers rapid results, unlike the five-day BOD_5 test, which only offers retrospective data due to the length of time required to obtain results and does not not enable quick water quality evaluation needed for effective process control (Abdalla and Hammam, 2014; Henze *et al.*, 2002).

The relationship between Influent BOD_5 and COD of the produced water stream exiting the three-phase separator is depicted in a comprehensive pair plot (Fig. 4.6). This plot includes kernel density estimates (KDE) for both BOD_5 (Fig. 4.6A) and COD (Fig. 4.6D), a contour plot overlaid on a scatter plot displaying the Pearson correlation coefficient and p-value (Fig. 4.6C) and a scatter plot with a fitted polynomial regression (Fig. 4.6B).

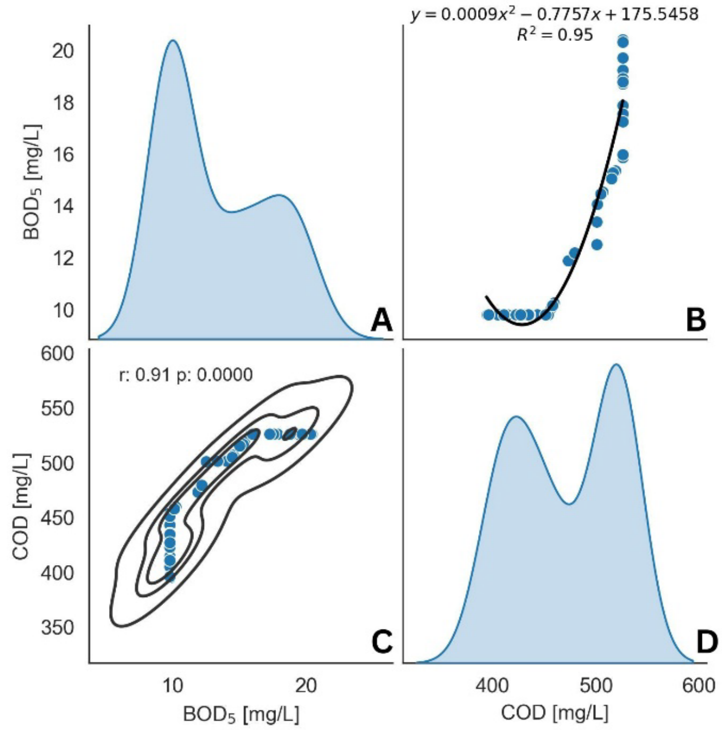


***Figure 4.6**: Analysis of Biochemical Oxygen Demand$_5$ and Chemical Oxygen Demand correlation*

The KDE plots reveal that BOD_5 values clustered around 10 to 18 mg/L, while COD values concentrated around 450 to 520 mg/L, indicating a bimodal distribution for both parameters. A strong positive correlation (r = 0.91) between BOD_5 and COD is evident, with a statistically significant p-value of 0.00, implying that as COD increases, BOD_5 also tends to rise. The quadratic polynomial trendline emerged as the best fit, with an R^2 value of 0.95, suggesting that the relationship, while not perfectly linear, can be well represented by a quadratic function. This polynomial model accounts for 95% of the variance in BOD_5 values based on COD.

Fig. 4.7 illustrates the calculated biodegradability index for the produced water from Angola block 15, which yielded an average of 0.028 within the range of 0.021 to 0.039. The biodegradability index (0.028) of the produced water suggests that conventional biological treatment methods would be ineffective due to the high presence of refractory organics and toxic compounds. The presence of phenol (2.70–4.12 mg/L) further supports this limitation, as phenolic compounds are known to inhibit microbial activity, making biological degradation difficult (Abdalla and Hammam, 2014).

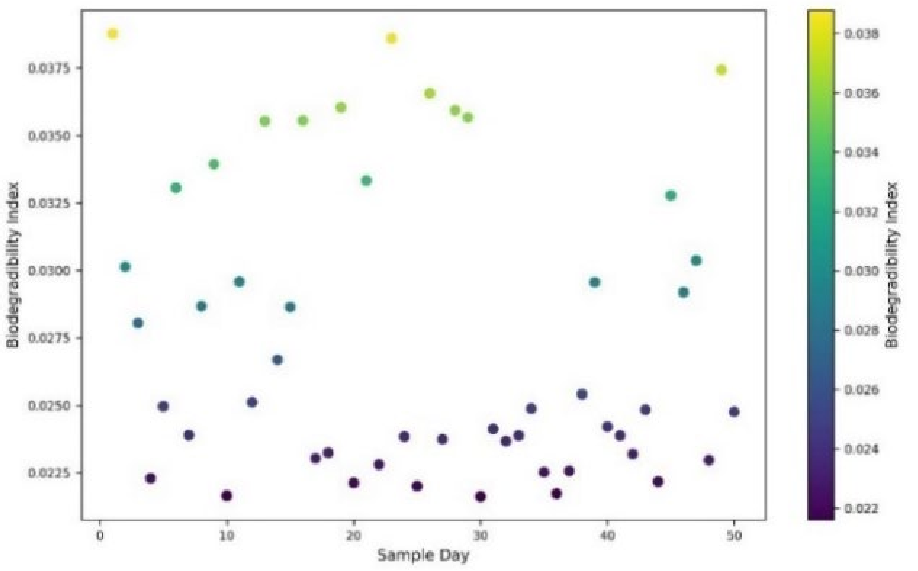


Figure 4.7: *Variation of Biodegradability Index Over 50 days*

4.4. Environmental Challenges and Treatment Strategies

Non-compliance with these regulations can result in severe ecological consequences, including marine habitat contamination and disruptions to fisheries, underscoring the need for strict monitoring and enforcement. Angola's produced water composition, particularly its organic and inorganic pollutant load, differs from other offshore regions, necessitating site-specific treatment strategies rather than a universal approach.

With a TDS concentration of 14,256 mg/L, Angola's produced water exhibits moderate salinity compared to the North Sea (229,860 mg/L) and the Permian Basin (128,641 mg/L). However, high chloride levels (400–2,146.84 mg/L) disrupt marine osmoregulation, affecting survival and reproduction (Jiang *et al.*, 2022). Elevated TDS inhibits biological treatment efficiency due to osmotic stress, requiring desalination methods such as reverse osmosis or thermal distillation for reuse. Additionally, microbial sulfate reduction can generate hydrogen sulfide, increasing corrosion risks and necessitating sulfate removal technologies.

Hydrocarbon contamination remains a significant concern, with TOG (96.08 mg/L) exceeding Angola's discharge limit of 30 ppmw (César *et al.*, 2024), posing risks of bioaccumulation, biodiversity loss, and food chain disruption (Hu *et al.*, 2022). Mechanical separation via hydrocyclones and induced gas flotation is necessary before secondary treatments such as activated carbon adsorption or advanced oxidation to meet discharge standards. Persistent hydrocarbons (TPH: 54.04 mg/L) may require ozonation or Fenton's reagent for further degradation. Organic pollutants, including BTEX (4.79 mg/L) and PAHs (0.56 mg/L), present carcinogenic and mutagenic threats. PAHs persist in sediments, endangering benthic ecosystems (Hu *et al.*, 2022), while a biodegradability index of 0.028 indicates resistance to natural degradation. Effective removal involves zeolite adsorption, membrane filtration and advanced oxidation.

Heavy metals such as nickel (314.07 mg/L) and iron (4.10 mg/L) further exacerbate environmental risks due to their bioaccumulation potential (Al-Ghouti *et al.*, 2019). Often originating from production infrastructure corrosion, these metals highlight the need for corrosion-resistant materials and chemical inhibitors. Techniques like chemical precipitation, ion exchange, and membrane separation play a vital role in lowering metal levels prior to discharge.

High TSS (314.00–502.81 mg/L) and COD (394.02–526.00 mg/L) indicate substantial organic and inorganic contamination. Increased TSS reduces oxygen exchange, while elevated COD suggests oxidizable organic matter that can deplete oxygen in receiving waters (Zhou *et al.*, 2018). Multimedia filtration and gas flotation effectively reduce TSS, while advanced oxidation and biological aerated bioreactors lower COD to acceptable levels. Phenol concentrations (2.70–4.12 mg/L), though lower than other organic pollutants, remain toxic to aquatic life (Al-Ghouti *et al.*, 2019). Derived from extraction chemicals and hydrocarbon degradation, phenols can be removed through activated carbon adsorption, synthetic resins, or enzymatic laccase-based bioreactors.

4.5. Future Research Directions

Since produced water from offshore oil and gas producing platforms is often discharged into the environment, it is essential to treat it properly to meet regulatory standards and prevent pollution. The study emphasises the importance of characterizing petroleum effluents. The following recommendations are proposed to aid a broader characterisation of produced water to determine the best wastewater management within the Angolan context as well as assist in planning optimal treatment practice:

1. Extend the characterisation of produced water to include platforms from other blocks from Angolan offshore waters and conduct studies beyond the initial 50-day period to understand the full variability in quality.

2. Conduct further analysis of organic compounds, including gasoline, diesel, and oil range organics, to thoroughly understand their composition.
3. Investigate the presence of naturally occurring radioactive material (NORM), such as Radium, Uranium, Thorium, Polonium, and Plutonium, due to the limited data on NORM in Angolan produced water.
4. Characterise the role of organic contaminants in aquatic toxicity, particularly their interactions with salinity and heavy metals, using various ecotoxicological models.
5. Explore beneficial reuse applications in Angola which is a water-scarce region, ensuring regulatory compliance and environmental sustainability.
6. Optimizing treatment efficiency through the integration of physicochemical and biological processes while developing region-specific discharge regulations that balance environmental protection with operational feasibility.
7. Further hydrocarbon fractionation analysis to distinguish between non-polar, polar, and semi-volatile fractions, thereby clarifying the observed TOG-TPH relationship.

4.6. Conclusion

This study provides a comprehensive assessment of produced water from Angola’s offshore Block 15, revealing its complex composition of high salinity, hydrocarbons, organic contaminants, and heavy metals. The low biodegradability indicates that conventional biological treatments are inadequate, highlighting the need for advanced processes such as membrane separation, adsorption, and advanced oxidation. While Angola’s produced water shares some similarities with other offshore regions, its unique characteristics demand tailored treatment strategies and strict regulatory oversight. By filling a critical knowledge gap, this study offers a foundational reference for future research, encouraging further analysis across other offshore blocks, deeper exploration of contaminants, and evaluation of reuse opportunities to support sustainable water management in Angola’s oil and gas sector.

CHAPTER 5

ANOMALY DETECTION AND PERFORMANCE EVALUATION IN PRODUCED WATER DEOILING

"Watch and pray so that you will not fall into temptation. The spirit is willing, but the flesh is weak."

Matthew 26:41

CHAPTER 5: ANOMALY DETECTION AND PERFORMANCE EVALUATION IN PRODUCED WATER DEOILING

5.1. Introduction

In offshore oil and gas operations, the treatment of produced water is critical to both regulatory compliance and environmental stewardship. Among the key separation technologies employed are the Hydrocyclone and Induced Gas Flotation (IGF) units, which work to reduce oil-in-water concentrations before discharge. However, their performance is often influenced by fluctuating inlet conditions, equipment aging, and inconsistencies in chemical dosing (Liu *et al.*, 2021).

This chapter sets out to achieve several key objectives aimed at enhancing the understanding and monitoring of produced water treatment performance. First, it aimed to explore the behavior of operational data and discern patterns in unit outputs to characterise system variability. The second objective was to detect and label anomalies in oil-in-water concentrations, which are indicative of abnormal or suboptimal operating states. Additionally, the study seeked to quantify the oil removal efficiency of each treatment stage, establishing a performance baseline under varying conditions.

Building on these insights, the chapter further aimed to develop classification models capable of identifying anomalous events, as well as to compare the diagnostic performance of multiple classification models, guiding the selection of suitable models for integration into digital monitoring frameworks.

5.2. Overview of Anomaly Detection

An anomaly, often referred to as an outlier, is any data point or pattern that deviates notably from expected behavior within a dataset. Such deviations can arise from random noise, like faulty sensor readings or data entry errors, or they may indicate rare but significant events, such as equipment malfunctions or abrupt changes in system dynamics (Shekhar *et al.*, 2021). Detecting these anomalies involves the use of statistical methods and Machine Learning (ML) algorithms to automatically flag unusual patterns as the primary objective is to identify abnormal behavior at an early stage, enabling timely intervention, improving system reliability and maintaining both process stability and regulatory compliance (Jesus *et al.*, 2021).

In industrial applications such as offshore produced water deoiling, anomaly detection is critical. It facilitates the early recognition of process irregularities or measurement errors, ensures the integrity of high-frequency operational data and supports process control by maintaining oil-in-water concentrations within permissible limits.

5.3. Anomaly Detection Methods

Anomaly detection encompasses a broad range of techniques developed to identify unusual patterns or data points that deviate from expected behavior. These methods generally fall into two categories: statistical approaches and ML-based algorithms. Statistical methods offer simplicity and interpretability, making them suitable for preliminary screening, while ML approaches such as ensemble models and kernel-based detectors are more adaptable to high-dimensional or nonlinear data structures. Each technique brings distinct advantages depending on the characteristics of the dataset, the presence of noise and the operational context. Sections 5.3.1 to 5.3.6 outline key methods that have been widely discussed in literature for their relevance to anomaly detection in industrial and data-intensive applications (Schneider and Xhafa, 2022; Geekiyanage *et al.*, 2021; Katragadda *et al.*, 2020; Chandola *et al.*, 2009)

5.3.1. Statistical Detection via the Interquartile Range (IQR)

The interquartile range (IQR) is a classical statistical method for detecting outliers in univariate data. It measures the spread of the central 50% by subtracting the first quartile (Q1) from the third quartile (Q3). Points falling below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ are flagged as anomalies; stricter thresholds (e.g., $3 \times IQR$) can be used for higher sensitivity (Vinutha *et al.*, 2018).

5.3.2. Isolation Forest

Isolation Forest is an unsupervised, ensemble-based method designed for anomaly detection. Unlike distance- or density-based approaches, it isolates anomalies by creating random decision trees. The core idea is that anomalies are few and structurally different, so they can be separated with fewer splits. Each tree recursively selects random features and split points; the number of splits needed to isolate a point (path length) indicates its abnormality. Anomaly scores are derived from the average path length across all trees. Effective in high-dimensional data and with linear computational complexity, Isolation Forest is well-suited for large-scale industrial applications (Fadul, 2023).

5.3.3. Local Outlier Factor

The Local Outlier Factor (LOF) algorithm detects anomalies by comparing the local density of each point to that of its neighbors, making it effective for datasets with varying densities. LOF uses reachability distance and local reachability density (LRD) to measure how isolated a point is. The LOF score is the ratio of the average LRD of a point's neighbors to its own LRD; higher scores indicate stronger outlierness. While LOF excels at finding local anomalies without assuming a global distribution, it is sensitive to the choice of the neighborhood parameter k, which requires careful tuning (Fadul, 2023).

5.3.4. Novelty Detection using LOF

A variant of LOF, Novelty Detection focuses on identifying anomalies in new, unseen data. The model is first trained on a clean reference dataset assumed to be free of outliers, learning the typical local density patterns of normal behavior. During deployment, it evaluates incoming data points by comparing their local densities to this learned structure, effectively flagging deviations as anomalies. By separating training and testing phases, Novelty Detection is well-suited for real-time monitoring and operational systems where historical data is well-characterised but incoming data may contain unexpected shifts. This enables proactive responses and supports continuous operational reliability (Yang *et al.*, 2022).

5.3.5. One-Class Support Vector Machine

The One-Class Support Vector Machine (SVM) is a kernel-based method for unsupervised anomaly detection. It learns a decision function that defines a boundary enclosing most of the normal data in a high-dimensional feature space, classifying points outside this boundary as anomalies. By using kernel functions, it maps input data into a space where a linear hyperplane can effectively separate normal instances from outliers. Trained exclusively on normal data, it is robust to irrelevant features since only support vectors define the boundary. Although computationally intensive, One-Class SVM offers a powerful framework for detecting anomalies in complex, nonlinear systems (Amer *et al.*, 2013)

5.3.6. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a high-performance implementation of gradient boosting, known for its efficiency and accuracy in supervised learning. For anomaly detection, XGBoost can be adapted as a binary classification task, labeling data points as normal or anomalous. It builds an ensemble of decision trees iteratively, with each tree correcting the errors of the previous ones. The model optimises a combined objective function, including a loss term (typically binary cross-entropy) and regularisation terms to prevent overfitting. XGBoost's scalability, parallelisation, and ability to handle imbalanced data make it ideal for large-scale industrial anomaly detection. The workflow involves labeling anomalies, splitting data into training and testing sets, training the model, and using it to predict anomaly likelihood in new data (Anjum *et al.*, 2022).

5.4. Process Description

The produced water treatment under analysis consists of a number of process units as illustrated in Fig. 5.1. The water recovered from the high pressure three phase separator flows under pressure to the Hydrocyclone-Induced Gas Flotation system for processing. The main components of the produced water treatment system are:

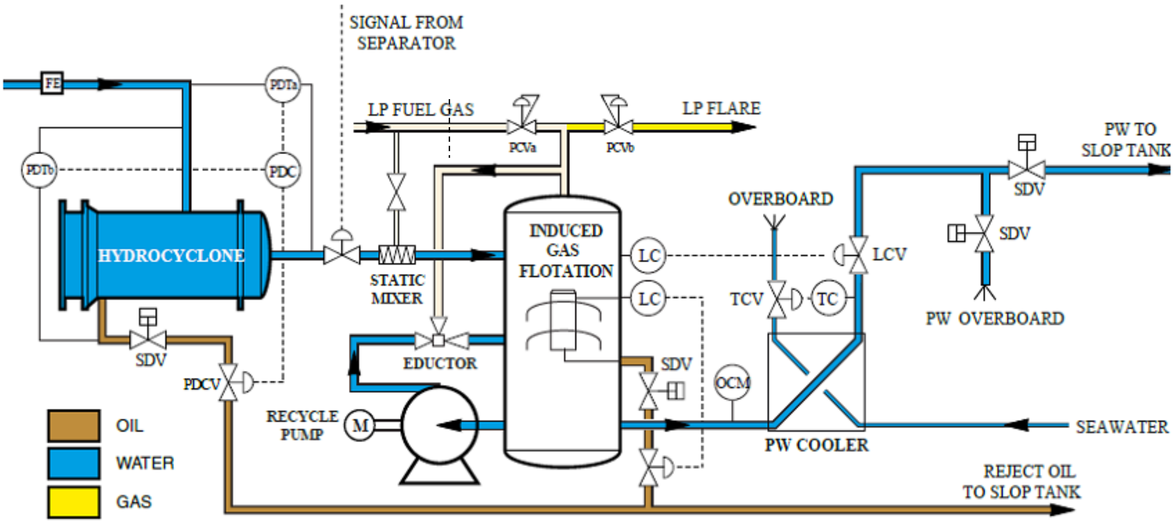


***Figure 5.1**: Process flow diagram of the offshore produced water deoiling system under assessment, illustrating key treatment stages*

Demulsifier is dosed into the hydrocarbon raw fluids upstream of the three-phase separators to assist in breaking emulsions and improving initial oil–water separation. Before water enters the hydrocyclone, a water clarifier is also added to enhance separation efficiency by promoting the coalescence of dispersed oil droplets.

The feed stream of dirty produced water then enters the hydrocyclone, where the separation process takes place. To maintain efficient operation, a differential pressure ratio is controlled between the inlet, oil reject (to slop tank), and clean water outlet (to the IGF unit). The hydrocyclone reject flow rate is managed by a differential pressure ratio control instrument (PDC), which modulates the differential pressure control valve (PDCV) based on signals from two pressure transmitters: PDTb (inlet to oil reject) and PDTa (inlet to water outlet).

After hydrocyclone treatment, produced water is routed to the Induced Gas Flotation (IGF) unit. A continuous recycle stream, driven by a duty recycle pump, maintains internal circulation, while gas bubbles are injected to support flotation. Fuel gas from the topsides low-pressure fuel gas header maintains the vessel pressure at 2.5 barg through PCVa (set at 2.4 barg), with excess pressure vented to the low-pressure flare via PCVb (set at 2.6 barg).

Downstream of the IGF, produced water passes through a cooler. The temperature is controlled by regulating seawater flow on the cold side. An oil-in-water analyser (OCM) installed upstream of the cooler monitors water quality. If the water meets discharge specifications, it is sent overboard via shutdown valve SDVa. If off-spec, SDVa closes, and shutdown valve T24-SDVb opens to divert the stream to the slop tank system.

5.5. Methodology

The dataset used in this analysis originates from high-resolution operational monitoring of an offshore produced water deoiling system, capturing time-stamped measurements across key processing units, specifically the Hydrocyclone and Induced Gas Flotation (IGF) unit, as shown in Fig. 5.1. The dataset, summarised in Fig. 5.2 and Table 5.1, comprises continuous recordings that provide a detailed, granular view of the separation dynamics over time. All statistical analysis, data preprocessing, and machine learning modelling were conducted using *Python 3.10* within a *Jupyter Notebook* environment, employing key libraries including *Pandas, NumPy, Scikit-learn, Matplotlib,* and *XGBoost* to ensure transparency and reproducibility. To preserve data integrity while maintaining the fidelity of anomalous patterns, missing values were forward-filled, a method selected over interpolation to avoid smoothing potentially informative deviations (Geekiyanage *et al.*, 2021).

Feature matrices were then constructed to support classification modeling, incorporating both raw sensor measurements and derived variables. Among the engineering features, were efficiency metrics for the Hydrocyclone and IGF units, enabling the quantification of oil removal performance supporting downstream benchmarking and diagnostic analyses.

5.6. Results and Discussion

Understanding the statistical behavior of process variables is a crucial first step in assessing the stability and efficiency of offshore produced water deoiling systems. Fig. 5.2 presents a boxplot visualisation of the selected operational data (y-axis) from the offshore oil production facility, which consisted of 255 aggregated data points, each representing an average of three consecutive readings, derived from a total of 765 original data entries collected over approximately 64 days of offshore operation.

Variables were arranged along the y-axis, with their statistical distributions shown horizontally. The boxplots convey essential distributional features, including interquartile range, median and the presence of outliers

.

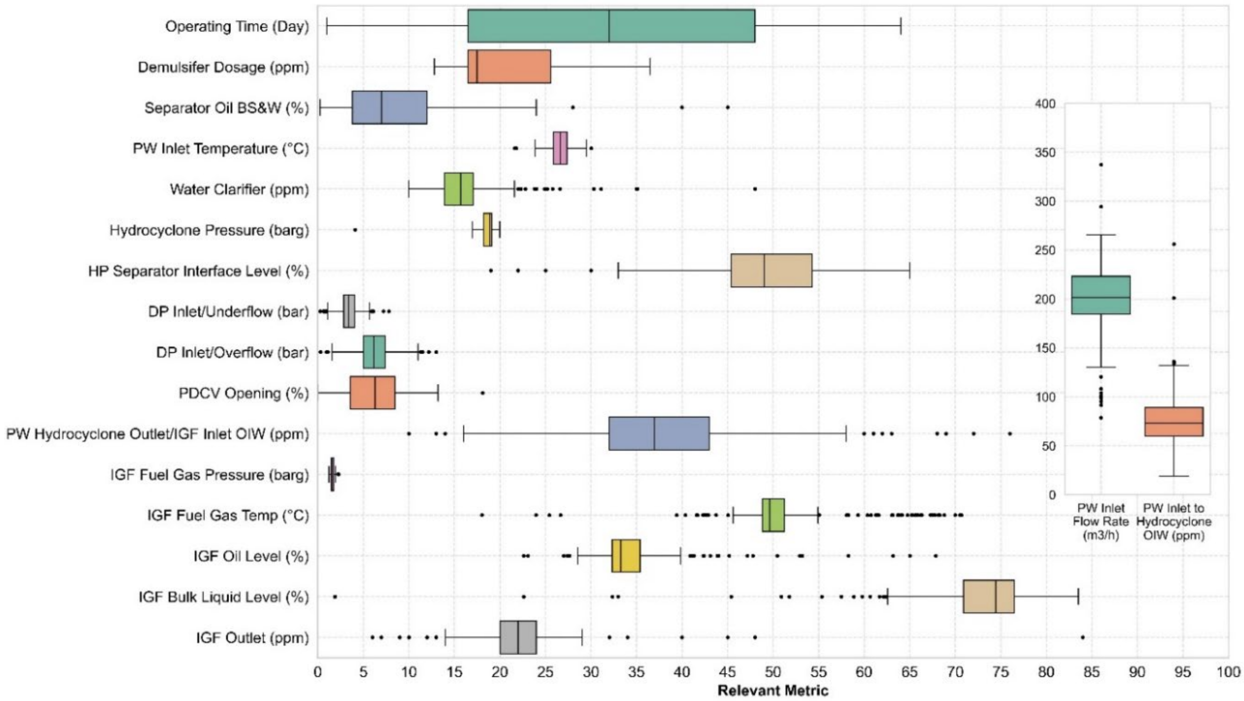


***Figure 5.2**: Critical operational variables across the Hydrocyclone and IGF systems*

To complement the visualisation in Fig. 5.2, Table 5.1 provides detailed descriptive statistics for each operational feature, including the mean, standard deviation, and observed operational range.

Table 5.1: *Descriptive statistics for key operational parameters across the Hydrocyclone and IGF systems*

Parameter	Mean	Range
Demulsifier Dosage (ppm)	20.1±5.4	12.8 - 36.5
Separator Oil BS&W (%)	8.3±6.2	0.27 - 45.0
Hydrocyclone Liners Online (#)	133.0±0.0	133.0 - 133.0
Hydrocyclone Liners Blanked (#)	22.0±0.0	22.0 - 22.0
PW Inlet Flow Rate (m^3/h)	226.5±214.1	78.62 - 2201.1
PW Inlet Temperature (°C)	26.6±1.3	21.62 - 30.0
Water Clarifier (ppm)	16.3±4.4	10.0 - 48.0
Hydrocyclone Pressure (barg)	18.7±1.1	4.1 - 19.98
HP Separator Interface Level (%)	48.8±7.5	19.0 - 65.0
PW Inlet to Hydrocyclone OIW (ppm)	76.4±26.8	19.0 - 256.0
DP Inlet/Underflow (bar)	3.5±1.2	0.27 - 7.8
DP Inlet/Overflow (bar)	6.2±2.4	0.3 - 13.0
PDCV Opening (%)	6.1±3.2	0.0 - 18.1
PW Hydrocyclone Outlet/IGF Inlet OIW (ppm)	39.7±13.9	10.0 - 134.0
IGF Fuel Gas Pressure (barg)	1.6±0.2	1.25 - 2.3
IGF Fuel Gas Temp (°C)	51.0±6.9	18.03 - 70.7
IGF Oil Level (%)	34.9±5.8	22.62 - 67.8
IGF Bulk Liquid Level (%)	71.1±12.4	1.87 - 83.5
IGF Outlet (ppm)	21.5±6.5	6.0 - 84.0

5.6.1. Effect of Anomalies on Oil-In-Water Concentration Patterns

Fig. 5.3 illustrates the kernel density estimations (KDEs) of oil-in-water concentrations at three key sampling points: the PW Inlet to the Hydrocyclone, the Hydrocyclone Outlet/IGF Inlet, and the IGF Outlet. Fig. 5.3A shows the full dataset, encompassing both normal and anomalous observations, while Fig. 5.3B depicts the cleaned distributions after outlier removal. In the presence of anomalies (Fig. 5.3A), the PW Inlet and Hydrocyclone Outlet/IGF Inlet exhibited broad and highly skewed distributions (skewness of 1.72 and 1.97, respectively), coupled with elevated kurtosis values (8.37 and 8.64, respectively), signaling frequent deviations that suggest upstream instability or episodic process upsets. The IGF Outlet distribution shows pronounced skewness (3.64) and exceptionally high kurtosis (33.86), indicating residual sensitivity to upstream fluctuations. Upon anomaly exclusion (Fig. 5.3B), the distributions contracted and become more symmetric. The IGF Outlet, in particular, demonstrates a tightly centered and near-normal profile (a mean of 21.7 ppm and standard deviation of 3.1 ppm), indicating its stable and effective polishing capability. Nevertheless, the PW Inlet continued to exhibit the widest spread (std dev = 19.0 ppm), emphasizing its role as the primary source of variability, likely influenced by the inherent heterogeneity of the produced fluid feed.

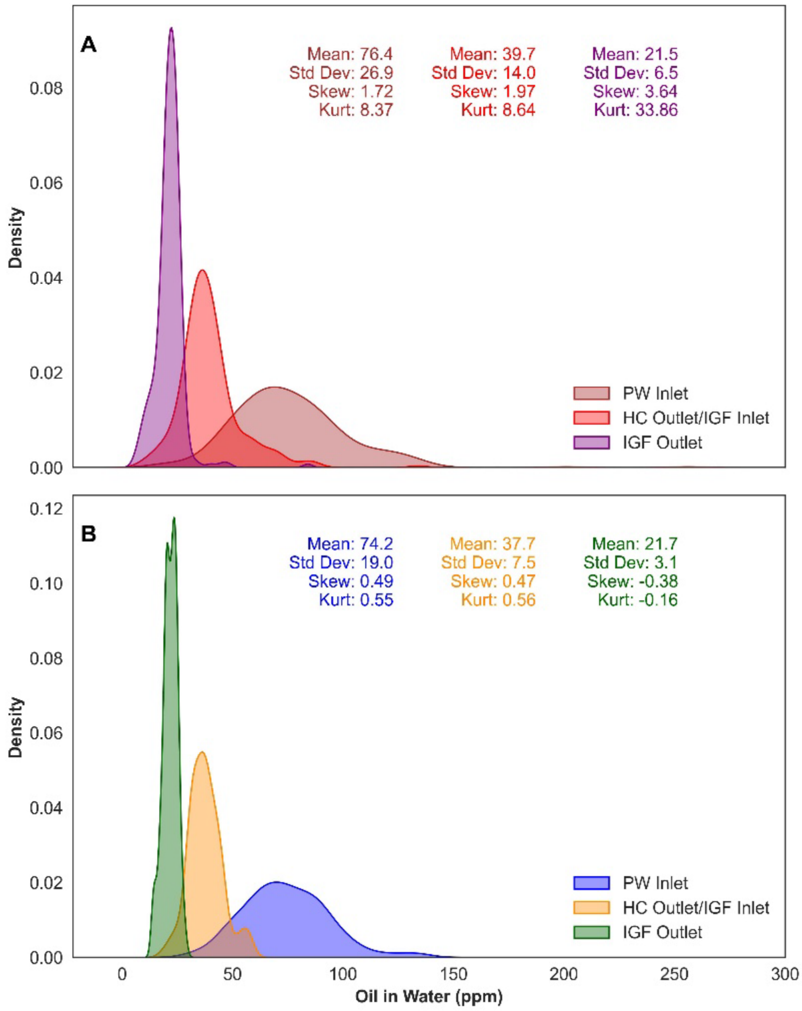


Figure 5.3: *Distribution profiles of oil-in-water concentrations of (A) dataset with anomalies and (B) cleaned data post-anomaly filtering*

Fig. 5.4 provides a time-series view of oil-in-water concentration behavior across the same process locations. Fig. 5.4A captures the full operational profile, including anomalies, where distinct transient spikes are visible, particularly at the PW Inlet, occasionally exceeding 250 ppm. These peaks are echoed downstream at the Hydrocyclone Outlet and IGF Outlet (albeit with dampened amplitude), reflecting partial buffering during upset conditions. Fig. 5.4B captures the operational profiles without the anomalies, offering a clearer view of baseline operations. Here, the Hydrocyclone Outlet and IGF Outlet maintain consistently low oil-in-water concentrations, with the IGF Outlet remaining below 50 ppm throughout, underscoring its role in ensuring discharge compliance.

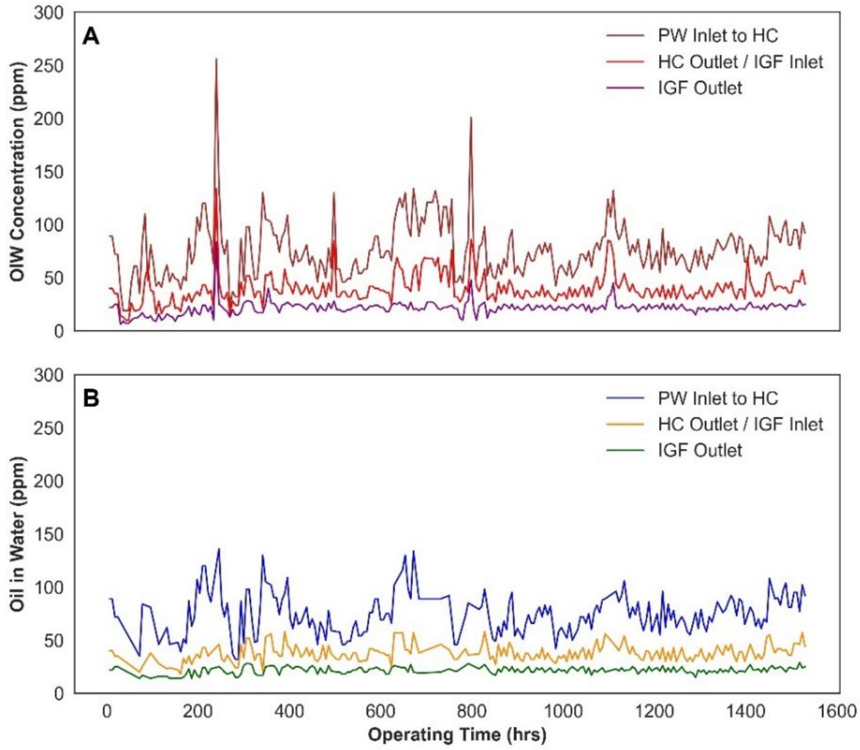


Figure 5.4: *Time series profiles of oil-in-water concentrations of (A) dataset with anomalies included and (B) cleaned data after anomaly removal*

Fig. 5.5 presents scatter plots of oil-in-water concentrations over operating time for the three primary stages of the produced water treatment process: (Fig. 5.5A) PW Inlet to Hydrocyclone, (Fig. 5.5B) Hydrocyclone Outlet / IGF Inlet, and (Fig. 5.5C) IGF Outlet. Each subplot distinctly visualises normal observations and detected anomalies, highlighting deviations in system behavior across the treatment chain. In Fig. 5.5A, the PW Inlet stream exhibited the highest dispersion, with several pronounced outliers exceeding 150 ppm and a maximum outlier surpassing 250 ppm. These anomalies suggest upstream disturbances, such as unstable production conditions, emulsified slugs or suboptimal separation efficiency. Moving downstream in Fig. 5.5B, the number and magnitude of outliers were notably reduced. However, intermittent peaks reaching 120 ppm indicate the Hydrocyclone is only partially effective at mitigating upstream variability. Fig. 5.5C, corresponding to the IGF Outlet, reveals a markedly stabilised trend. Here, the concentration remained tightly clustered around the 20–30 ppm range, and the detected anomalies were relatively low in both frequency and magnitude, rarely exceeding 50 ppm. This sequential suppression of oil-in-water concentration confirms the layered functionality of the treatment process: while the Hydrocyclone provides initial bulk separation, the IGF unit performs as a polishing stage, effectively smoothing residual fluctuations.

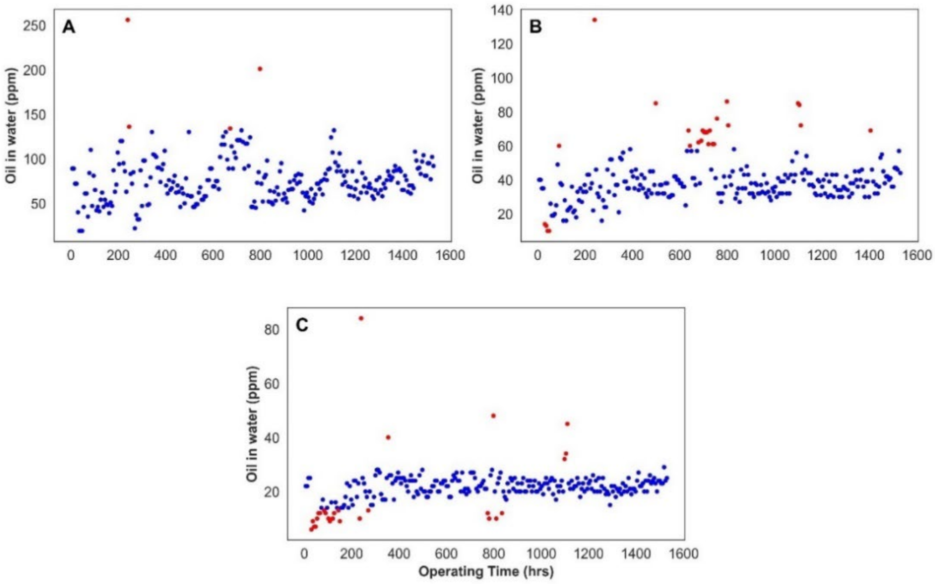


***Figure 5.5**: Anomaly-labeled oil-in-water time series at (A) Hydrocyclone Inlet, (B) Hydrocyclone Outlet/IGF Inlet and (C) IGF Outlet*

5.7. Performance Evaluation of Treatment Units

***Figure 5.6**: Produced water samples showing treatment progression: from left to right, Hydrocyclone Inlet (untreated), Hydrocyclone Outlet/IGF Inlet (partially treated), and IGF Outlet (fully treated).*

Fig. 5.6 shows three produced water samples taken at different stages of the treatment process, clearly illustrating the progressive removal of oil-in-water. The first bottle on the left represents the Hydrocyclone inlet, containing raw produced water with a visibly high concentration of oil. The middle bottle corresponds to the Hydrocyclone outlet/IGF inlet, where partial separation has occurred and oil content is reduced. The final bottle on the right contains the sample of the IGF outlet, showing significantly improved clarity with minimal visible oil traces, indicating effective polishing by the IGF unit. This treatment sequence underscores the cumulative efficiency of the Hydrocyclone and IGF stages in treating produced water before discharge or reuse. Refer to to the values of the operational features *PW Inlet to Hydrocyclone OIW (ppm), PW Hydrocyclone Outlet/IGF Inlet OIW (ppm) and IGF Outlet (ppm)* in Table 5.1.

Fig. 5.7 illustrates the distribution of removal efficiencies for the Hydrocyclone and IGF units, with subplot A representing the raw data including outliers, and subplot B representing the cleaned dataset after anomaly filtering.

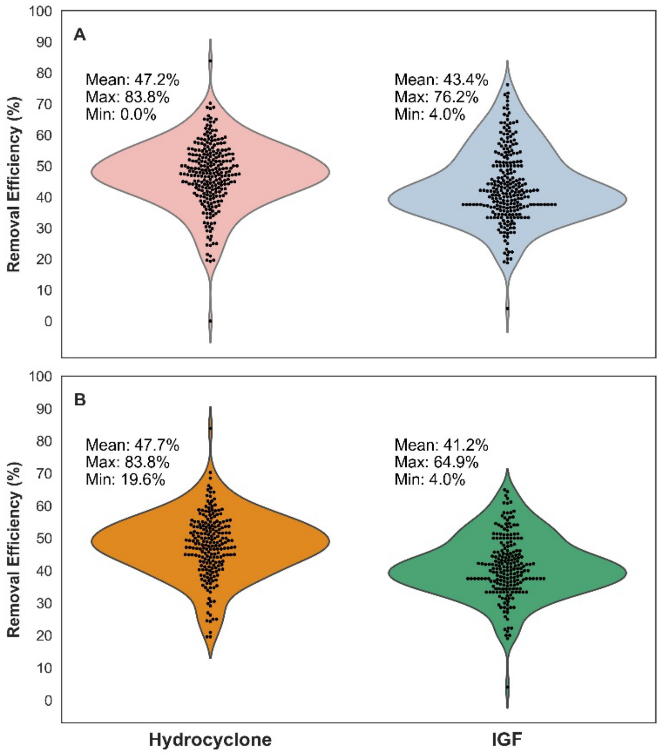


***Figure 5.7**: Hydrocyclone and IGF removal efficiencies using (A) all data and (B) anomaly-filtered data*

In Fig. 5.7A, the Hydrocyclone exhibited a mean removal efficiency of 47.2%, ranging from 0.0% to 83.8%, while the IGF unit showed a mean efficiency of 43.4%, with values spanning from 4.0% to 76.2%. The broad base and tapering tail of the Hydrocyclone's violin indicates a bimodal-like spread, suggesting periods of underperformance intermixed with higher-efficiency operation. Notably, the presence of a 0.0% minimum highlights potential process failures or severe anomalies, such as short-circuiting, phase carryover or measurement artifacts. The IGF unit, in turn, displayed a slightly narrower and more symmetrical distribution but still retained a wide spread, implying sensitivity to fluctuations in feed quality or loading.

In Fig. 5.7B, the distributions shifted toward more consistent performance. The Hydrocyclone's mean efficiency increased slightly to 47.7% and the minimum value improved to 19.6%, suggesting that the extreme low-efficiency events were largely attributable to outliers rather than representative behavior. The violin shape became more concentrated, indicating tighter clustering of efficiency values near the median. The IGF unit showed a mild reduction in mean efficiency to 41.2%, but a significant narrowing of the distribution, particularly on the lower end. The maximum efficiency also dropped slightly, reflecting the removal of positively skewed anomalous peaks. The removal of anomalies resulted in clearer, more actionable insights by eliminating distortive data points and better representing steady-state performance.

Fig. 5.8 presents the overall daily average oil removal efficiency over the operating timeline for the deoiling system, incorporating intra-day variability through standard deviation error bars.

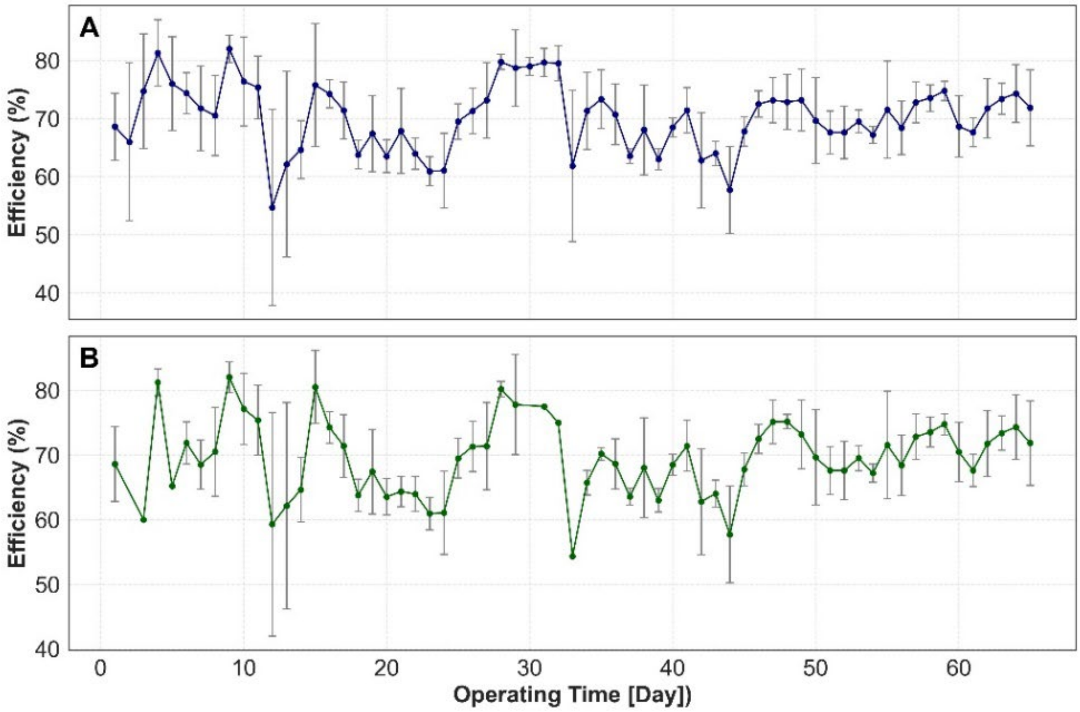


Figure 5.8: *Temporal variation of treatment efficiency. (A) Hydrocyclone unit and (B) Induced Gas Flotation unit*

In Fig. 5.8A, which includes the full dataset with anomalies, efficiency values range from 48.2% to 85.6%, with an average of 66.5%. Pronounced fluctuations occur during the early days (0–15) and mid-range period (25–40), as shown by wider error bars. These suggest short-term disturbances, measurement anomalies, or transient process instabilities that distort performance assessment. Fig. 5.8B, after anomaly filtering, shows a smoother trend with reduced variability. The minimum efficiency improves to 58.0%, the maximum adjusts to 82.7%, and the average slightly increases to 68.2%. The shorter error bars indicate more consistent operation, allowing clearer identification of stable performance periods, notably around days 20–25 and 50–60.

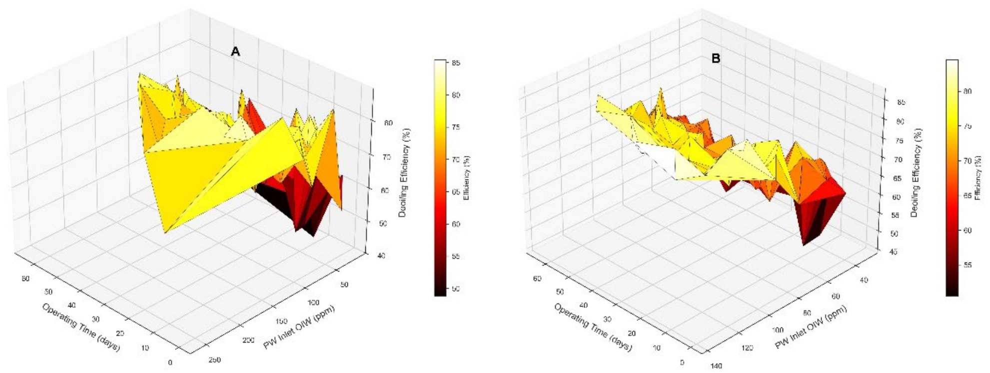


***Figure 5.9**: Deoiling efficiency versus operating time and PW inlet oil-in-water concentration. (A) With anomalies and (B) After anomaly filtering*

Fig. 5.9 shows the corresponding 3D efficiency surfaces before and after anomaly filtering, as functions of operating time and PW inlet oil-in-water concentration. In Fig. 5.9A (with anomalies), the surface reveals pronounced peaks and troughs, reflecting strong variability and scattered local instabilities. In contrast, Fig. 5.9B (after anomaly filtering) exhibits a smoother, more continuous surface. This emphasises the system's stable operational domain and indicates a clearer relationship between inlet oil-in-water concentration, operating time and deoiling efficiency.

Taken together, Fig. 5.8 and Fig. 5.9 highlight the value of anomaly detection and filtering. While the increase in mean efficiency is modest, the reduction in artificial variability and enhanced interpretability significantly improve the accuracy of system performance assessments and provide a stronger foundation for advanced data-driven optimisation and digital twin implementations.

5.7.1. Classification of Anomalies

While statistical detection methods are effective at flagging extreme values, they often fail to capture the complex multivariate relationships that underlie anomalous behavior. To address this limitation, this section introduces ML models developed to classify anomalies based on key operational features. The analysis focuses on the Hydrocyclone and IGF units, with each modeled as a binary classification task. Class imbalance was addressed using appropriate resampling or weighting techniques.

To evaluate the relative effectiveness of various approaches for anomaly classification, both supervised and unsupervised classification models were assessed. The aim was to identify suitable candidates for real-time anomaly detection in produced water deoiling systems. The models tested include: (1) Isolation Forest, (2) LOF, (3) Novelty-based LOF, (4) One-Class

SVM and (5) XGBoost. All classifiers were trained and tested using consistent operational feature sets tailored to the respective Hydrocyclone and IGF systems. Model performance was evaluated based on overall accuracy, confusion matrix structure and recall for the minority (anomaly) class.

5.7.1.1. Hydrocyclone

Fig. 5.10 presents confusion matrices and accuracy metrics for five anomaly detection models applied to the produced water deoiling system, with Fig. 5.10A corresponding to the Hydrocyclone unit and Fig. 5.10B to the IGF unit. For both units, the unsupervised approaches (Isolation Forest, LOF and Novelty LOF), achieved moderate accuracy (ranging from 0.82 to 0.85) with some ability to detect anomalies, typically identifying 3 to 5 true positives while maintaining a reasonable false positive rate. Novelty LOF showed slightly better performance among the unsupervised methods across both treatment units. One-Class SVM performed poorly for both treatment units, with low accuracy (0.52 to 0.53) and a high number of false positives, indicating limited reliability in practical deployment.

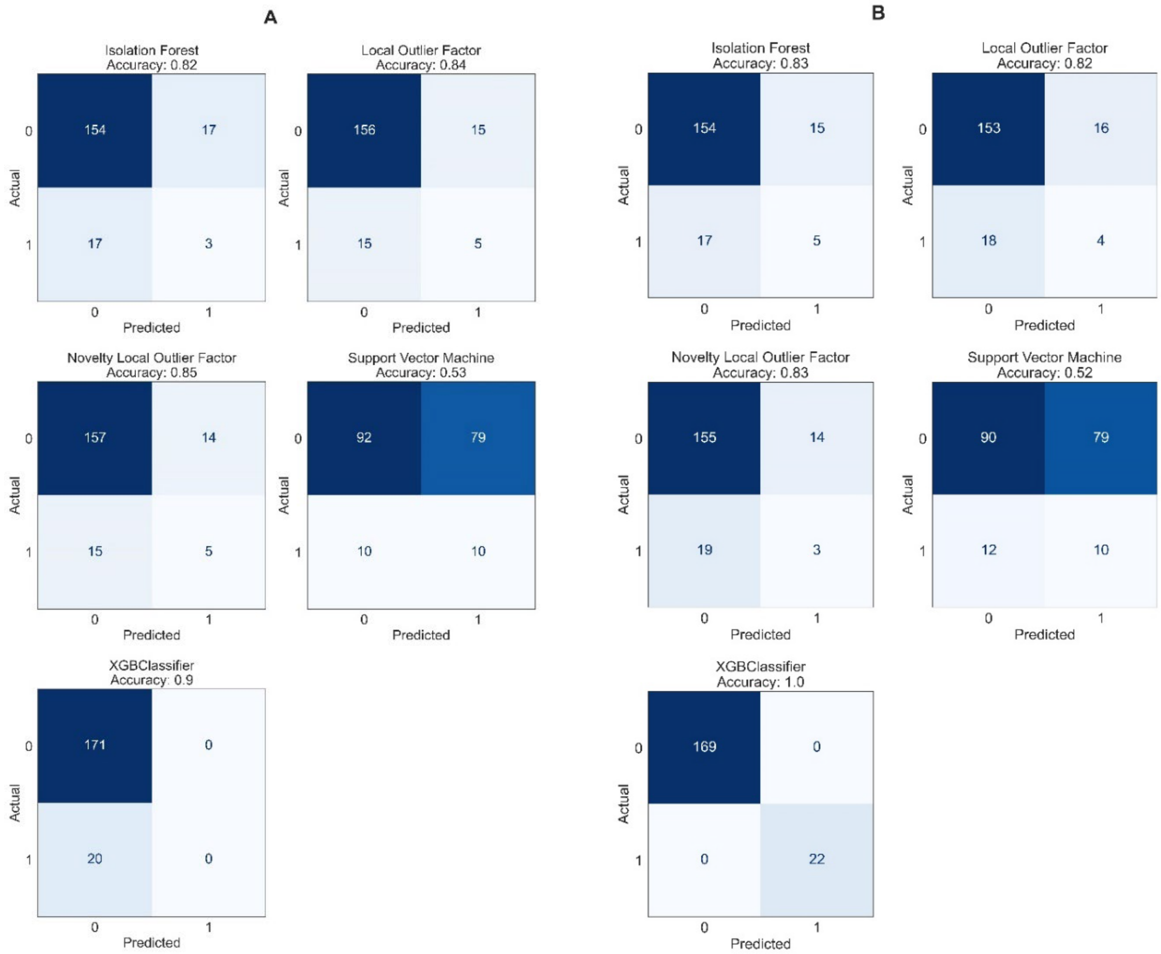


***Figure 5.10**: Confusion matrices for five anomaly detection models applied to the produced water deoiling system for (A) Hydrocyclone and (B) IGF treatment units*

The supervised XGBoost model demonstrated contrasting behavior between the two treatment units. In the Hydrocyclone subsystem (Fig. 5.10A), it achieved the highest overall accuracy (0.90) but failed to detect any anomalies, classifying all samples as normal, highlighting a strong bias toward the majority class due to imbalance (Anjum *et al.*, 2022). However, for the IGF treatment unit (Fig. 5.10B), XGBoost achieved perfect classification, correctly identifying all normal and anomalous samples, with an accuracy of 1.00. This contrast underscores the critical importance of dataset characteristics and balance in model performance, and highlights the potential of supervised learning for Anjum *et al.*, 2022). These quantitative metrics validate the earlier visual interpretations and further reinforce the necessity of applying class imbalance mitigation methods, such as those demonstrated in Fig. 5..10, to unlock the full potential of supervised models in critical anomaly detection tasks.

To complement the confusion matrices results presented in Fig. 5.10, Table 5.2 summarises the full classification reports, including precision, recall, and F1-score for each model across both the Hydrocyclone and IGF treatment units. As observed, the unsupervised models (Isolation Forest, LOF, and Novelty LOF) consistently demonstrated high precision and recall for the normal class (above 0.91), yet struggled with anomaly detection. For example, Isolation Forest produced an anomaly F1-score of 0.11 for the Hydrocyclone and 0.29 for the IGF unit, with similar limitations seen in LOF and Novelty LOF. Despite moderate accuracy values (ranging from 0.83 to 0.87), these models yielded low macro averages due to poor performance on the minority class.

One-Class SVM further exemplified this imbalance, particularly for the Hydrocyclone treatment unit where accuracy dropped to 0.50 and the anomaly F1-score was 0.11, despite a relatively high recall of 0.33. For the IGF unit, recall remained high at 0.53, but precision fell sharply, leading to an anomaly F1-score of 0.17 indicating a high false positive rate and limited practical application consideration.

***Table 5.2**: Performance metrics of anomaly detection models for Hydrocyclone and IGF systems*

	Hydrocyclone				**Induced Gas Flotation**			
	Isolation Forest							
Classes	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
Normal	0.91	0.91	0.91	173	0.93	0.93	0.93	174
Anomaly	0.11	0.11	0.11	18	0.28	0.29	0.29	17
Accuracy			**0.83**	191			**0.87**	191

Macro Average	0.51	0.51	0.51	191	0.60	0.61	0.61	191
Weighted Average	0.83	0.83	0.83	191	0.87	0.87	0.87	191
Local Outlier Factor								
Normal	0.92	0.92	0.92	173	0.93	0.93	0.93	174
Anomaly	0.22	0.22	0.22	18	0.28	0.29	0.29	17
Accuracy			**0.85**	191			**0.87**	191
Macro Average	0.57	0.57	0.57	191	0.60	0.61	0.61	191
Weighted Average	0.85	0.85	0.85	191	0.87	0.87	0.87	191
Novelty Local Outlier Factor								
Normal	0.91	0.92	0.92	173	0.93	0.93	0.93	174
Anomaly	0.19	0.17	0.18	18	0.24	0.24	0.24	17
Accuracy			**0.85**	191			**0.86**	191
Macro Average	0.55	0.55	0.55	191	0.58	0.58	0.58	191
Weighted Average	0.85	0.85	0.85	191	0.86	0.86	0.86	191
Support Vector Machine								
Normal	0.88	0.52	0.65	173	0.92	0.54	0.68	174
Anomaly	0.07	0.33	0.11	18	0.10	0.53	0.17	17
Accuracy			**0.50**	191			**0.54**	191
Macro Average	0.47	0.43	0.38	191	0.51	0.53	0.43	191
Weighted Average	0.81	0.50	0.60	191	0.85	0.54	0.64	191
XGBoost								
Normal	0.91	1.00	0.95	173	0.91	1.00	0.95	174
Anomaly	0.00	0.00	0.00	18	0.00	0.00	0.00	17
Accuracy			**0.91**	191			**0.91**	191
Macro Average	0.45	0.50	0.48	191	0.46	0.50	0.48	191
Weighted Average	0.82	0.91	0.86	191	0.83	0.91	0.87	191

Fig. 5.11 addresses the limitations observed in the XGBoost model by illustrating the performance of an enhanced XGBoost classifier model trained with class imbalance mitigation techniques. Two strategies were applied: the Synthetic Minority Over-sampling Technique (SMOTE) and class weighting. In Fig. 5.11A, the SMOTE-based model successfully identified all anomalies for the Hydrocyclone system with only one false positive, while the class-weighted XGBoost achieved perfect classification, accurately identifying both normal and anomalous samples without error. This demonstrates a substantial improvement over the initial classifier model, which previously failed to detect any anomalies for the Hydrocyclone treatment unit.

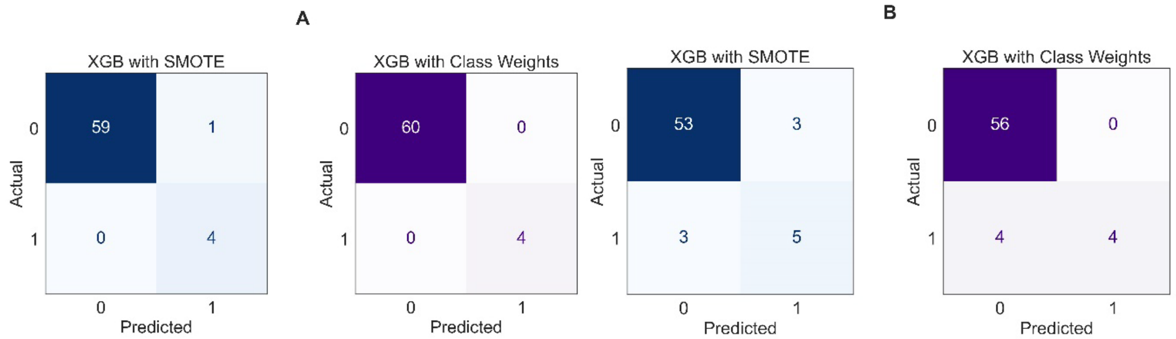


Figure 5.11: *Confusion matrices of XGBoost models using SMOTE and class weights for anomaly detection. (A) Hydrocyclone system. (B) Induced Gas Flotation (IGF) system.*

In Fig. 5.11B, both models showed improved balance between sensitivity and precision for the IGF treatment unit. The SMOTE-enhanced model correctly classified five of the nine anomalies, with moderate false positive and false negative rates (3 each), while the class-weighted model achieved slightly better balance with four true positives and only four false negatives, without misclassifying any normal data points.

5.8. Conclusion

This chapter evaluated the performance and anomaly behavior of offshore produced water deoiling units, focusing on the Hydrocyclone and IGF systems. Operational data analyses revealed substantial variability at the PW Inlet into the deoiling system and progressive oil-in-water reduction across the treatment stages.

Anomalies were detected using both statistical and ML methods. IQR-based filtering helped isolate outliers that distorted efficiency trends, while ML classifiers provided deeper insights into multivariate anomaly patterns. Among unsupervised models, LOF and Novelty LOF offered moderate performance. One-Class SVM showed high false positive rates and XGBoost, though accurate overall, failed to detect anomalies due to class imbalance.

Enhanced XGBoost models with SMOTE and class weighting significantly improved anomaly detection, with the class-weighted version achieving perfect or near-perfect classification. These results highlight the value of preprocessing and imbalance correction in supervised learning workflows.

The combined use of statistical screening, efficiency evaluation, and ML classification establishes a data-driven framework for monitoring the performance of the deoiling units, with strong potential for integration into digital twin platforms for real-time diagnostics and control.

CHAPTER 6

MACHINE LEARNING APPLICATIONS FOR EFFLUENT QUALITY PREDICTION IN WASTEWATER TREATMENT: A BIBLIOMETRIC ANALYSIS

"You will know the truth, and the truth will set you free."

John 8:32

CHAPTER 6: MACHINE LEARNING APPLICATIONS FOR EFFLUENT QUALITY PREDICTION IN WASTEWATER TREATMENT: A BIBLIOMETRIC ANALYSIS

6.1. Introduction

Water, a fundamental resource for human life and industrial activities, is increasingly threatened by hazardous pollutants arising from both anthropogenic activities and natural processes. Each day, substantial volumes of industrial wastewater is generated, much of it containing toxic substances that are often discharged into the environment without adequate treatment (Wang *et al.*, 2023; Ray *et al.*, 2023). Advancing wastewater treatment and reuse is thus essential for safeguarding limited water resources, minimizing environmental harm, and meeting the growing global demand for water. By integrating physical, chemical and biological processes to remove contaminants, wastewater treatment facilitates the safe discharge or reuse of effluent (Zaibel *et al.*, 2022). As a key pillar of sustainable development, wastewater treatment plays a critical role in preserving human and ecological health, particularly in industrial and agricultural settings, and its continued improvement relies on the adoption of innovative, energy-efficient and cost-effective technologies (Malviya and Jaspal, 2021).

In this context, the integration of advanced tools such as machine learning (ML), a prominent subfield of artificial intelligence, offers new opportunities to enhance wastewater treatment systems. Fueled by increased computational power and the rise of big data, ML has attracted significant attention for its ability to continuously improve predictive accuracy and optimise complex processes, leading to its widespread adoption in both industry and environmental management (Ige *et al.*, 2025). By enabling data-driven decision-making and supporting intelligent system design, ML contributes to more efficient, adaptive and resilient wastewater treatment strategies.

The objectives of this chapter are to investigate the key drivers and patterns shaping research on the application of ML in wastewater treatment. Specifically, this chapter aims to identify the leading institutions and countries advancing the use of ML for predicting effluent water quality parameters in wastewater treatment and to determine which journals and publications act as primary hubs for disseminating innovative studies in this area. Furthermore, it seeks to analyse the collaborative networks among countries and institutions to understand the structure and dynamics of research partnerships in this field. Finally, the chapter examines the distribution of these research topics across different countries and academic disciplines, providing insights into the global landscape and thematic focus areas of ML-related wastewater treatment research.

6.2. Machine Learning

ML has emerged as a powerful tool for addressing complex, uncertain, and dynamic challenges. It is particularly valuable for optimizing wastewater treatment plant operations (de Canete *et al.*, 2021). By harnessing large volumes of data and operational knowledge, ML enables precise modeling and control of complex environmental processes (Guo *et al.*, 2015; Ye *et al.*, 2020; Zhao *et al.*, 2020). Consequently, these methods increasingly support decision-making, strengthen management and improve system performance in wastewater treatment facilities (Zhao *et al.*, 2020).

ML techniques can be classified as either supervised or unsupervised. Supervised approaches create predictive models that capture the relationships between input (explanatory) variables and output (response) variables (Corominas *et al.*, 2018; Newhart *et al.*, 2019; Newhart *et al.*, 2022). In contrast, unsupervised methods develop descriptive models to uncover hidden patterns within data when no prior information about potential relationships is available (Newhart *et al.*, 2019).

Table 6.1 provides an overview of the primary ML models applied in environmental and wastewater treatment research, detailing their acronyms, concise definitions and key references. This summary highlights the diversity of approaches available and underscores the growing importance of advanced data-driven methods in tackling complex environmental challenges.

Table 6.1: *Overview of machine learning algorithms commonly applied in wastewater treatment studies*

MODEL	ACRONYM	DEFINITION	REFERENCE
Adaptive Boosting	AdaBoost	Combines weak learners by reweighting misclassified samples to improve performance	Feng *et al.*, 2020
Artificial Neural Network	ANN	Brain-inspired network of connected neurons that learn nonlinear patterns	Ding *et al.*, 2013
Bagging Regressor	BR	Averages multiple models trained on bootstrapped samples to reduce variance	Kadiyala and Kumar, 2018
Bayesian Network	BN	Graph-based model capturing probabilistic dependencies among variables.	Marcot and Penman, 2019
Bidirectional Gated Recurrent	GRU	GRU variant using forward and backward sequences to capture full context.	Yang *et al.*, 2021
Categorical Boosting	CatBoost	Gradient boosting specialised for categorical data with built-in encoding	Bentéjac *et al.*, 2021
Cubist	CB	Rule-based regression tree with linear models in leaves for improved predictions.	Kuhn *et al.*, 2012
Decision Tree	DT	Splits data into branches based on feature rules for easy-to-interpret prediction	De Ville, 2013
Extreme Gradient Boosting	XGBoost	Advanced gradient boosting with regularisation and efficient computation	Bentéjac *et al.*, 2021
Fuzzy Network	FN	Combines neural networks and fuzzy logic to handle uncertainty and learn patterns.	de Campos Souza, 2020

Gaussian Process Regression	GPR	Non-parametric, probabilistic model providing predictions with uncertainty estimates	Schulz *et al.*, 2018
Gradient Boosting	GB	Builds models sequentially, each correcting errors of the last to improve predictions	Bentéjac *et al.*, 2021
K-Nearest Neighbors	KNN	Predicts by averaging or voting among k closest data points	Kramer, 2013
Linear Regression	LR	Captures the relationship between dependent and independent variables through a linear equation	Maulud and Abdulazeez, 2020
Logistic Regression	LogR	Estimates probabilities for classes using a logistic (sigmoid) function	Rymarczyk *et al.*, 2019
Long Short-Term Memory	LSTM	A type of recurrent neural network that uses memory cells to effectively capture and maintain long-term dependencies in sequential data	Van Houdt *et al.*, 2020
Multi-Layer Perceptron	MLP	Feedforward neural network with hidden layers for complex pattern modeling.	Rana *et al.*, 2018
Naïve Bayes	NB	Probabilistic classifier assuming independent features; simple and fast.	Sheth *et al.*, 2022
Partial Least Squares	PLS	Projects data into latent space to handle collinearity and improve prediction	Abdi and Williams, 2012
Random Forest	RF	A bagging-based tree ensemble that enhances precision and controls overfitting	Cutler *et al.*, 2012
Support Vector Machine	SVM	Identifies the best hyperplane that maximizes the margin between different classes	Suthaharan, 2016; Salcedo-Sanz *et al.*, 2014

6.3. Methodology

Bibliometric analysis is a powerful and systematic approach for investigating large-scale scholarly datasets, providing valuable quantitative and qualitative insights. By integrating data mining, processing and visualisation techniques, this method enables the graphical representation of complex interconnections within literature, such as thematic structures, conceptual frameworks, collaborative networks and areas of overlap. It effectively combines elements of statistical analysis, mathematical modeling and graphical design to uncover hidden patterns and relationships (Baarimah *et al.*, 2024; Wu *et al.*, 2023; Bazel *et al.*, 2023).

Due to its ability to map the current state of scientific output, bibliometric analysis supports the advancement of future investigations by critically evaluating the development and trajectory of prior studies, which have led to its extensive adoption across multiple scientific and technical fields (Baarimah *et al.*, 2024; Baarimah *et al.*, 2023; Donthu *et al.*, 2021; Van Eck and Waltman, 2010)

To guarantee a thorough and transparent study selection process, this research followed the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Page *et al.* 2021). Widely acknowledged for its capacity to enhance methodological transparency and reproducibility, PRISMA supports the development of systematic reviews that contribute to evidence-based knowledge. Additionally, the study design incorporated the structured and explanatory approach outlined by Jesson *et al.* (2011), further reinforcing methodological robustness.

This study relied on Scopus as the primary database, given its wide-ranging collection of peer-reviewed sources, consistent bibliometric records crucial for reliable analyses, and strong emphasis on technology-related research (Moya-Anegón *et al.*, 2007). Scopus' detailed indexing across diverse disciplines further ensured a comprehensive and robust foundation for the analysis (Baarimah *et al.*, 2024; Bazel *et al.*, 2023).

In this study, bibliometric analysis was employed to trace the evolution and identify emerging trends in ML applications within wastewater treatment research, with a particular emphasis on the use of effluent water quality indicators as primary target parameters. The literature search was executed on 10th June 2025 and focused exclusively on English-language articles published between 2015 and 2025, thereby capturing a decade of scholarly developments. The search strategy employed a combination of targeted keywords: "*wastewater*", "*water**", "*machine learning**", "*treat**", "*predict**", "*quality*", "*effluent*" and "*algorithm**". The symbol (*)

was applied to include all derivative forms of the root terms, maximizing the inclusiveness of the search results.

The process of study selection, carried out is illustrated in the flowchart presented in Fig. 6.1. An initial pool of 211 records consisting of research articles and conference proceedings was retrieved from the database search. Following an initial abstract screening phase, 24 articles were excluded as they were identified as literature reviews rather than empirical case studies. A comprehensive assessment of the remaining 187 articles led to the exclusion of an additional 77 studies that were deemed irrelevant, specifically those not addressing wastewater treatment or predictive modeling.

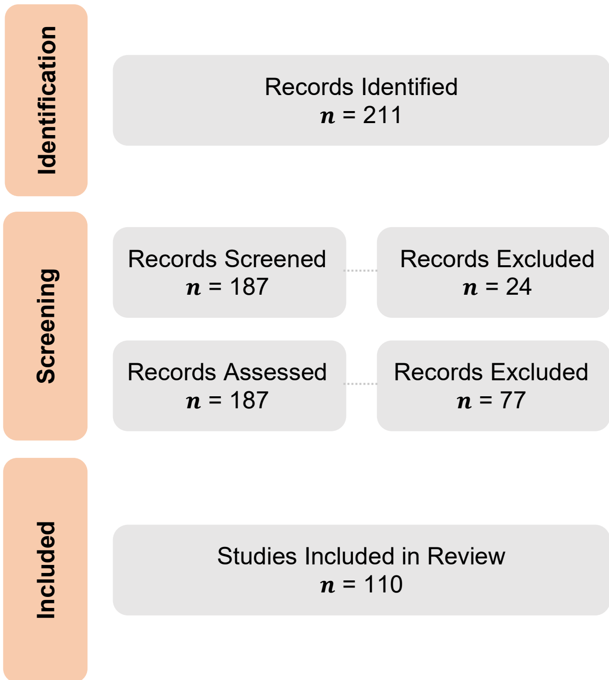


Figure 6.1: *PRISMA flow diagram illustrating the systematic identification, screening and selection process of studies included in this review*

6.4. Results and Discussion

6.4.1. Trend of Annual Publications

A total of 110 studies have demonstrated ML's strong capabilities in wastewater treatment effluent quality prediction, emphasizing ML's versatility and effectiveness in advancing wastewater treatment applications (Liu *et al.*, 2022; Safeer *et al.*, 2022). Fig. 6.2 illustrates the temporal evolution of publications employing ML models for predicting effluent water quality in wastewater treatment systems from 2015 to June 2025.

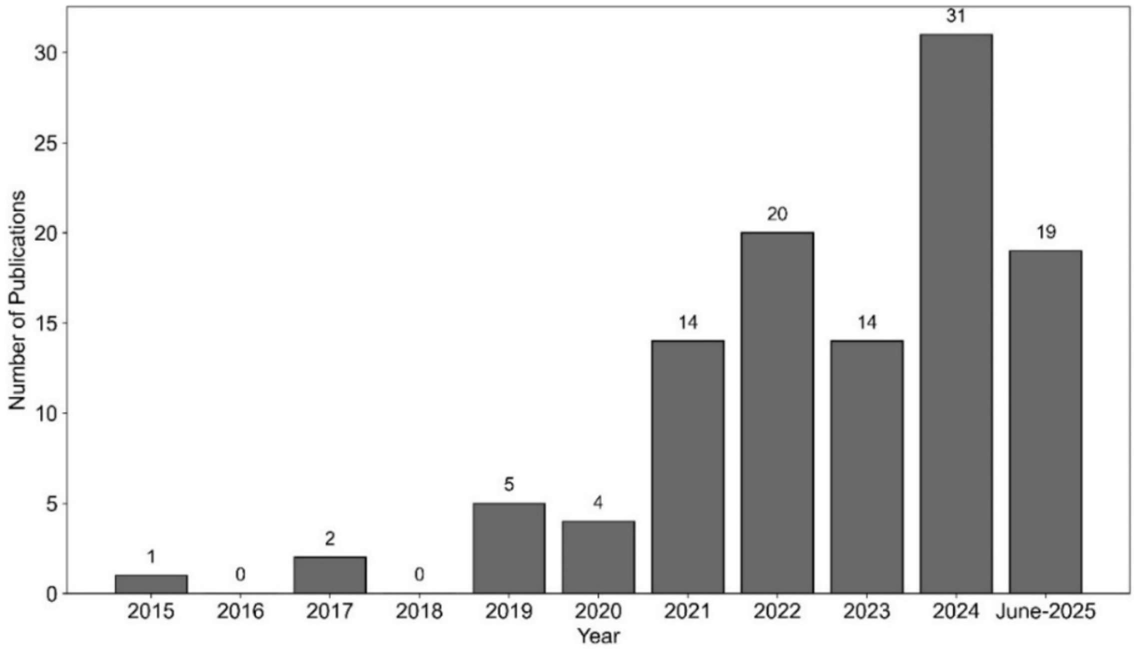


***Figure 6.2**: Annual number of publications employing Machine Learning models for predicting effluent quality in wastewater treatment systems (2015–2025)*

As shown in Fig. 6.2, there was minimal research activity in this area prior to 2019, with publication counts generally remaining below 5 per year. A notable increase begins in 2019, marking the growing recognition of ML methods as powerful tools for enhancing treatment performance and monitoring accuracy. From 2020 onward, the trend shows a marked acceleration, reaching a peak of 31 publications in 2024. While the count for 2025 currently stands at 19, which reflects data collected up to June-2025 and is expected to rise by the end of the year. This strong upward trajectory highlights the rapid adoption and integration of ML approaches in wastewater treatment research. Overall, Fig. 6.2 underscores the field's shift toward data-driven, predictive systems for improved wastewater management, confirming the growing relevance and maturity of ML-based solutions in environmental engineering.

6.4.2. Major Research Focus Areas

The extracted target parameters from the abstracts demonstrated a diverse range of effluent water quality metrics, with some parameters appearing more frequently than others. Solids were the most commonly mentioned parameter, appearing in 38 papers, followed closely by chemical oxygen demand (COD) with 36 mentions. Total Nitrogen (TN) was highlighted in 32 papers. Other frequently referenced variables included total suspended solids (TSS) and biochemical oxygen demand (BOD), each mentioned in 12 papers, as well as total Phosphorus (TP) and Ammonia Nitrogen, which appeared in 11 and 9 papers, respectively.

Additionally, Ammonium Nitrogen (NH_4^+-N) was noted in 5 papers, further underscoring the importance of monitoring nitrogen species in effluent water quality control.

Notably, many studies examined multiple output parameters simultaneously. Although explicit breakdowns of single versus multi-output models were not always provided, the frequent mention of several quality indicators within individual abstracts suggests a tendency toward comprehensive effluent characterisation. This multi-dimensional approach enables a more robust understanding of wastewater treatment performance and offers valuable operational insights for future process optimisation.

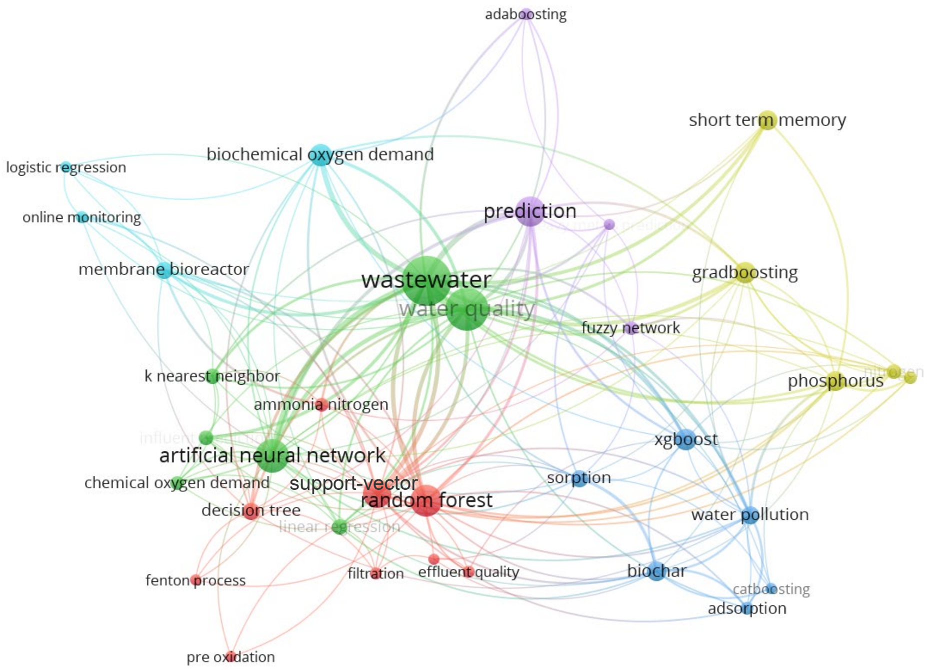


***Figure 6.3**: Network visualisation with author keyword co-occurrence clustering*

Fig. 6.3 presents a keyword co-occurrence network visualised, illustrating the research landscape of ML applications in wastewater treatment management. The network includes only specific keywords that appeared at least three times in the analysed dataset, ensuring focus on the most relevant and impactful terms (Bukar *et al.*, 2023). Central nodes such as wastewater and water quality dominate the map, underscoring their pivotal roles as foundational research themes. Surrounding these, key ML techniques, notably Artificial Neural Network, Random Forest and Support-Vector Machine, emerge as prominent, highly connected hubs, reflecting their widespread adoption for tasks such as effluent water quality prediction, filtration optimisation and pollutant modeling.

The network in Fig 6.3 further reveals thematic clusters; for instance, nutrient control terms like "Phosphorus" and "Ammonia Nitrogen" are closely linked with advanced approaches such as Long-Short Term Memory, Gradient Boosting and Fuzzy Network, indicating recent trends toward deep learning and hybrid methodologies. On the periphery, specialised processes such as Membrane Bioreactor, Fenton Process and Pre-Oxidation suggest targeted ML applications for specific treatment strategies. Additionally, terms like Biochar, Sorption and Adsorption cluster together, emphasising research focused on pollutant removal and resource recovery. Overall, this network illustrates a highly interconnected and multi-disciplinary research landscape, highlighting the growing integration of intelligent, data-driven solutions to advance treatment performance and environmental monitoring.

Fig. 6.4 illustrates a word cloud generated to visually represent the most frequently occurring terms related to ML applications in wastewater treatment and effluent water quality prediction. The size of each term reflects its relative frequency in the analysed dataset, highlighting the prominence of key topics such as *Water Quality*, *Treatment*, *Wastewater*, *Neural Network*, *Random Forest*, *Support-Vector Machine*, *Gradient Boosting* and *Prediction*. To guarantee reproducibility and ensure a consistent visual arrangement, a fixed random seed was applied. The process involved normalising the terms, mapping frequencies and configuring layout parameters, such as orientation preference and relative scaling.

***Figure 6.4**: Keyword co-occurrence word cloud illustrating prominent terms and research focuses in Machine Learning applications for effluent water quality prediction in wastewater treatment*

6.4.3. Geographic Distribution of Publications

Fig. 6.5 presents a global distribution map illustrating the number of articles focused on ML applications in wastewater treatment categorised by country.

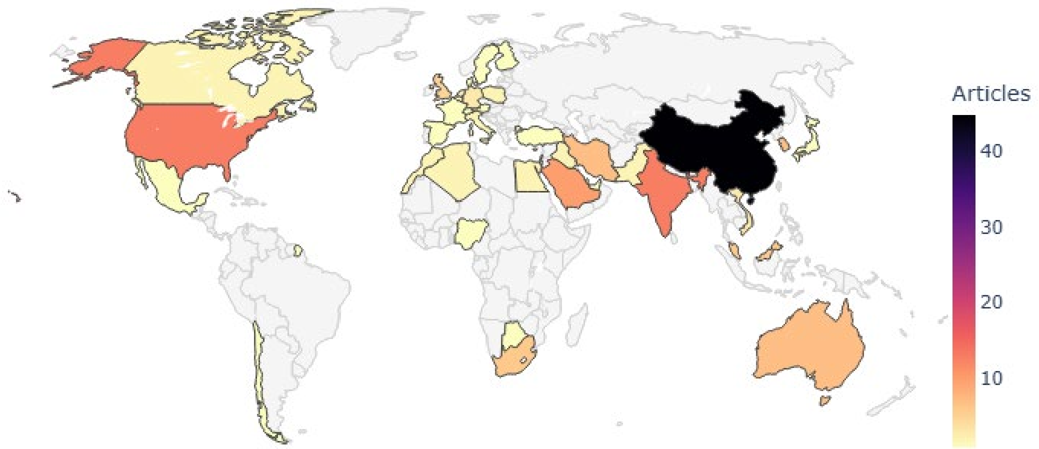


Figure 6.5*: Global distribution of publications on Machine Learning applications in wastewater treatment, with colour intensity indicating article counts per country*

The map highlights China as the dominant contributor, with 45 articles, underscoring its leading role in advancing data-driven effluent water quality prediction within wastewater treatment research. India and the United States follow, each contributing 13 articles, reflecting their significant engagement and strong research capacity in this field. Countries such as Saudi Arabia (10 articles), South Korea, Iran and Australia (each with 7 articles) also show active participation, indicating a broader international interest. Moderate contributions from Malaysia, South Africa and the United Kingdom (ranging from 5 to 6 articles) further underscore the global reach and multi-disciplinary nature of this research area. European countries, including Germany, Poland and the Netherlands, along with nations from Asia, the Middle East and Africa, are also represented, albeit with fewer publications.

6.4.4. Cross-Country Collaboration in Research

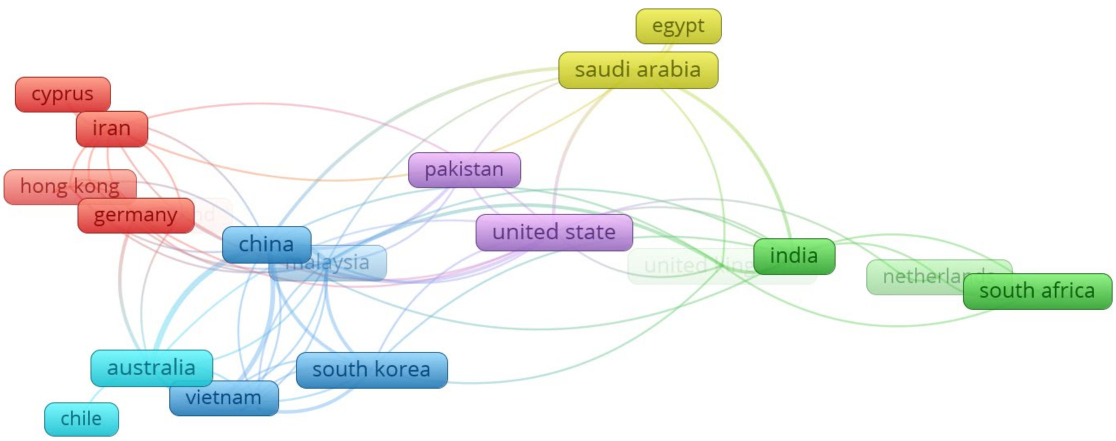


Figure 6.6*: Visual map of country cooperation publications*

Fig. 6.6 illustrates a country collaboration network map that highlights the most active and interconnected countries engaged in research on ML applications in effluent quality prediction for wastewater treatment. The thickness of lines and the size of the frames represent the number of publications and the type of colour represents the clustering of countries that regularly cooperate.

China ranks as the leading hub of collaboration, establishing extensive links with countries such as the United States, Australia, Vietnam and Pakistan, reflecting its central role in global research efforts. The United States follows closely, forming partnerships with India, Australia, China and several other nations, underscoring its influence and broad research outreach. Saudi Arabia and India also demonstrate strong international connections, indicating their growing investment in this interdisciplinary field. Other active collaborators include Australia, Iran, Malaysia, South Africa and Germany, each contributing to regional and global research exchanges. Vietnam, while smaller in scale, appears in multiple collaborative groups, suggesting emerging participation.

6.4.5. Journal-Level Publications Trends

Fig. 6.7 provides the top 10 leading journals and publication sources contributing significantly to the body of academic literature on the application of ML in effluent water quality prediction in wastewater treatment, including the number of articles published by each source.

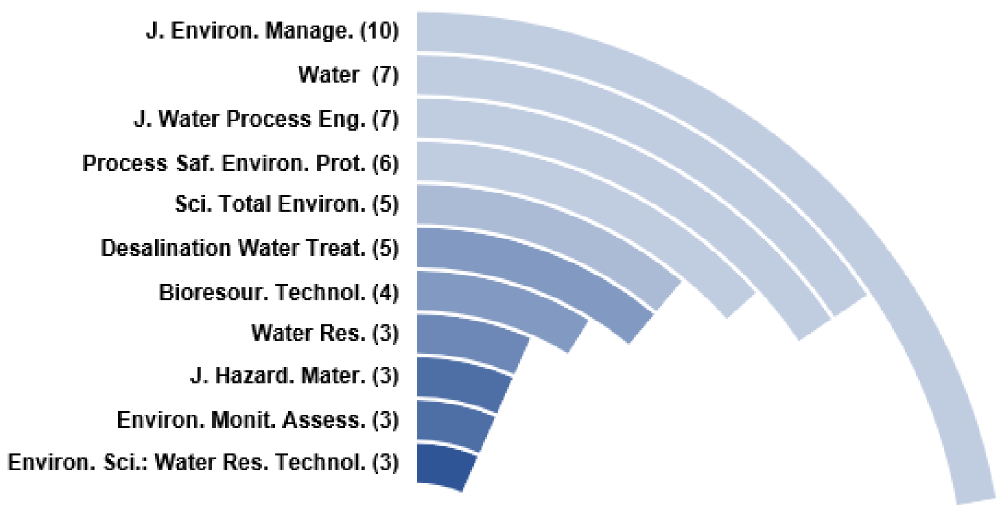


***Figure 6.7**: Distribution of publications on Machine Learning in wastewater treatment across leading journals*

The Journal of Environmental Management hosts the highest number of publications (10), reflecting its central role as a preferred outlet for research in this domain. This is followed by

Water and the Journal of Water Process Engineering, each with 7 publications, emphasising their prominence in disseminating studies related to water treatment and process engineering. Other notable journals include Process Safety and Environmental Protection (6), Science of the Total Environment and Desalination and Water Treatment (5 each), indicating their importance in broader environmental and desalination research contexts.

6.4.6. Top Contributing Universities Engaged in the Research Topic

Among the contributing institutions, Tsinghua University and South China University of Technology lead with 4 publications each, reflecting their strong research focus and continued commitment to advancing water and environmental technologies. Close behind, the University of Johannesburg, Ulsan National Institute of Science and Technology and Nanjing University each contributed 3 publications, underscoring their emerging influence and active engagement in this research domain. Together, these top five universities highlight the global and collaborative nature of scientific efforts addressing wastewater treatment effluent quality prediction and its challenges.

6.4.7. Research Area Distribution of Machine Learning Applications

Fig. 6.8 highlights the evolving disciplinary landscape of ML applications in wastewater treatment research from 2015 to June-2025. Over this period, Environmental Sciences emerged as the predominant research area, reflecting the central focus on environmental monitoring and effluent water quality prediction. Chemical Engineering also maintained a strong presence, underscoring the importance of integrating ML approaches for process optimisation and system design. Notably, recent years (2024 and 2025) show an increasing diversification of contributions from fields such as Energy, Computer Science, Biochemistry and Agricultural and Biological Sciences, suggesting a growing interdisciplinary interest in addressing complex wastewater treatment challenges through data-driven solutions.

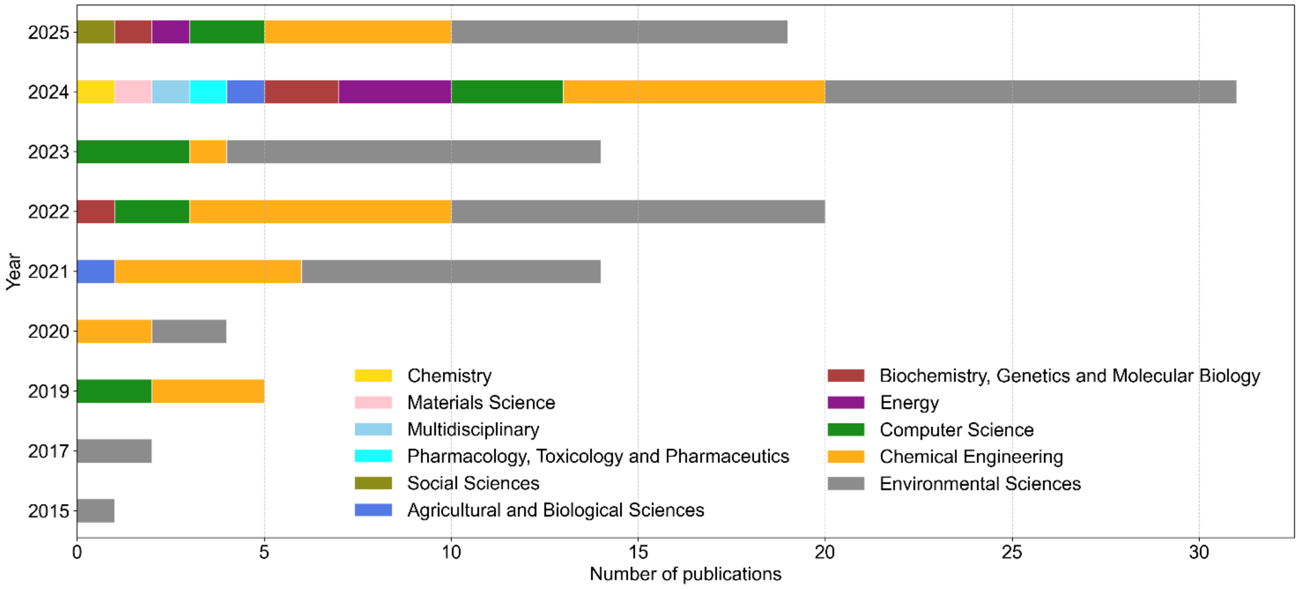


Figure 6.8*: Annual distribution of publications on Machine Learning applications in wastewater treatment by research area (2015 to June-2025)*

The sharp rise in publications observed in 2024 indicates an accelerated adoption of ML methodologies, aligning with broader global trends towards intelligent, sustainable and efficient water management practices.

6.4.8. Distribution and Temporal Trends of the ML Models Used

Fig. 6.9 presents a polar bar plot illustrating the distribution of articles that have employed various ML models for effluent water quality prediction in wastewater treatment. Each bar represents a distinct model, with its length indicating the number of articles citing or utilising that approach, thereby emphasising its relative prominence in the field. The colour gradient provides an additional visual dimension to distinguish the frequency of citations across models. The circular arrangement enhances readability and enables an intuitive comparison of model popularity without favoring a linear ordering.

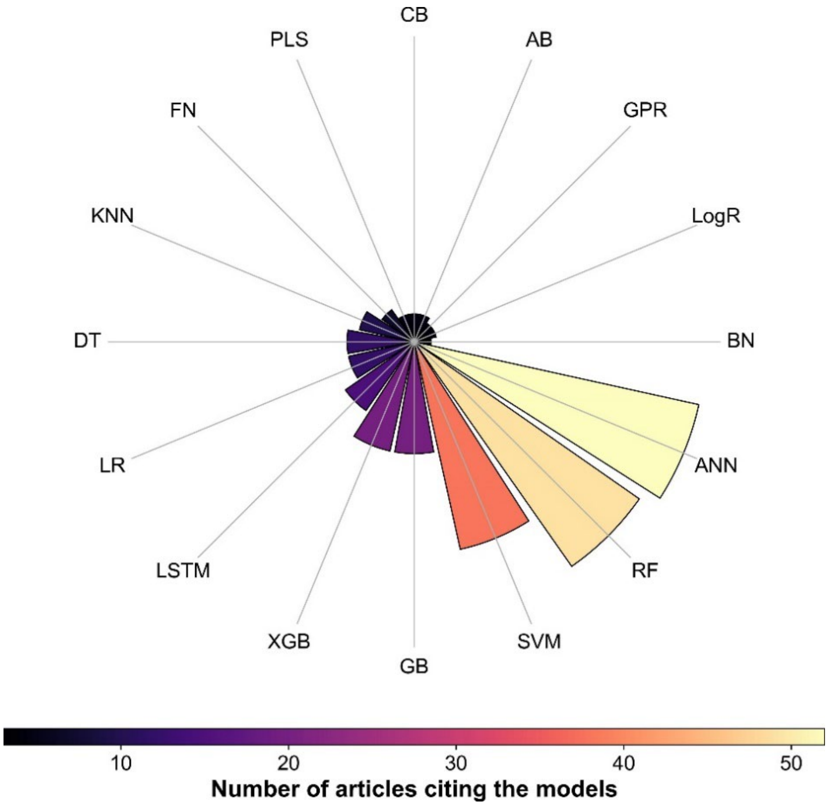


Figure 6.9*: Distribution of articles citing Machine Learning models for effluent quality prediction in wastewater treatment*

The results reveal that Artificial Neural Network (ANN), Random Forest (RF) and Support-Vector Machine (SVM) are the most frequently adopted models, followed by Gradient Boosting and XGBoost, highlighting their robustness and suitability for complex environmental data analysis.

Fig. 6.10 illustrates the temporal distribution of articles citing different ML models for effluent water quality prediction in wastewater treatment research from 2015 to June-2025. Each bubble represents the number of articles that mention a specific ML model in a given year, with the bubble size and colour intensity both reflecting the number of citations. The timeline reveals a marked increase in adoption and research output from 2021 onwards, suggesting a surge of interest and technological advancements during this period. Models like Gradient Boosting, Extreme Gradient Boosting and Long-Short Term Memory also show notable growth, reflecting a trend toward more sophisticated ensemble and deep learning techniques. In contrast, methods such as Bagging Regressor, Cubist and Bayesian Network have more modest citation frequencies, indicating their niche or emerging roles in this research field.

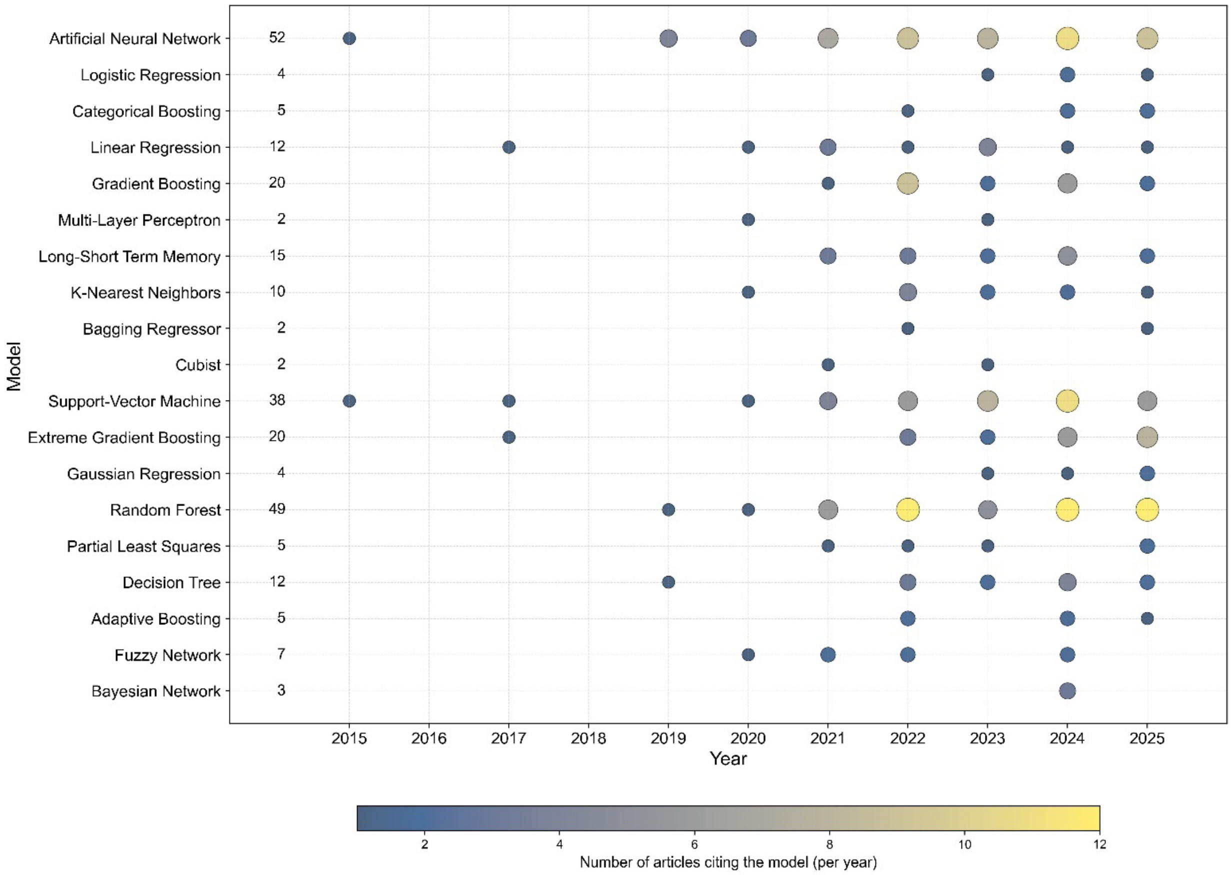


***Figure 6.10**: Temporal distribution of articles citing Machine Learning models for effluent quality prediction in wastewater treatment (2015 to June-2025)*

6.5. Conclusion

This chapter provided a comprehensive bibliometric analysis of ML applications for effluent water quality prediction in wastewater treatment. The findings highlight a sharp rise in publications over the past decade, driven by the growing recognition of ML's potential to enhance treatment performance and monitor accuracy. Environmental Sciences and Chemical Engineering emerged as the leading disciplines, with China, United States and India as major contributors to publications. The strong international collaborations and diverse research focus areas underscore the dynamic, multi-disciplinary nature of this domain. Key models such as Artificial Neural Network, Random Forest, Support-Vector Machine, Gradient and Extreme Gradient Boosting dominate the field, reflecting their robustness in handling complex environmental data. Overall, the results emphasise the crucial role of ML in advancing intelligent, data-driven solutions for sustainable wastewater management.

CHAPTER 7
ENSEMBLE LEARNING MODELS FOR PREDICTING EFFLUENT OIL-IN-WATER OUTLET CONCENTRATION IN OFFSHORE PRODUCED WATER DEOILING

"If you have faith as small as a mustard seed, you can say to this mountain, 'Move from here to there,' and it will move. Nothing will be impossible for you."

Matthew 17:20

CHAPTER 7: ENSEMBLE LEARNING MODELS FOR PREDICTING EFFLUENT OIL-IN-WATER CONCENTRATION IN OFFSHORE PRODUCED WATER DEOILING

7.1. Introduction

Produced water is the largest waste stream from offshore oil and gas operations and poses significant environmental and operational challenges due to its high oil content and complex contaminant profile (Gamwo *et al.*, 2022; Ghafoori *et al.*, 2022; Liu *et al.*, 2021). Effective treatment is essential to meet stringent discharge regulations, protect marine ecosystems, and maintain production efficiency. Hydrocyclone–Induced Gas Flotation (HIGF) systems are commonly employed to reduce oil-in-water (OIW) concentrations before discharge. However, their performance is heavily influenced by nonlinear and dynamic process conditions that challenge conventional control and prediction strategies (Bram *et al.*, 2020; Bram *et al.*, 2018; Durdevic and Yang, 2018; Das and Jäschke, 2018; Hansen *et al.*, 2018; Durdevic *et al.*, 2017).

While Computational Fluid Dynamics (CFD) models have been used to simulate flow behavior within these systems, they are not practical for real-time monitoring or control applications (de Araújo *et al.*, 2020; Vallabhan *et al.*, 2020). Traditional empirical control methods, such as those based on the Pressure Drop Ratio (PDR), often lack adaptability and are limited by model uncertainties, requiring continuous manual tuning and offering limited responsiveness to process fluctuations (Durdevic and Yang, 2018; Das and Jäschke, 2018; Hansen *et al.*, 2018; Durdevic *et al.*, 2017).

Machine learning (ML) has emerged as a powerful alternative, capable of capturing complex nonlinear relationships from historical data without relying on rigid first-principles assumptions (Lampropoulos, 2023; Aghbashlo *et al.*, 2021; Matheri *et al.*, 2021). In wastewater treatment, ML models have demonstrated success in predicting critical effluent quality indicators, enabling optimised treatment performance and early detection of anomalies (Lowe *et al.*, 2022; Nourani *et al.*, 2021; El-Rawy *et al.*, 2021; Guo *et al.*, 2015; Qambar and Al Khalidy, 2022; Wang *et al.*, 2022; Salles *et al.*, 2022; Ly *et al.*, 2022; Saboe *et al.*, 2021). However, applications of ML in offshore produced water treatment, particularly for predicting oil-in-water concentrations in HIGF systems, remain limited.

This chapter aims to fill this gap by exploring the application of interpretable, ensemble-based decision tree machine learning models—including Extreme Gradient Boosting (XGBoost), Gradient Boosting, and Random Forest to predict oil-in-water concentrations at the outlets of

Hydrocyclone and IGF units. Using a historical dataset from an offshore floating production facility, the study develops and evaluates regression models to capture the complex dynamics of produced water treatment. Specifically, the objectives of this chapter were to develop and compare decision tree-based ML models for predicting oil-in-water concentrations at key deoiling stages; to identify and analyse the most influential process variables affecting treatment performance; to assess the models' predictive accuracy and robustness under varying operational conditions; and to demonstrate the potential of ML models to support real-time monitoring, process optimisation, and adaptive control strategies in offshore water treatment operations.

7.2. Produced Water Deoiling: Hydrocyclone-Induced Gas Flotation System

Produced water deoiling involves the removal of free, dispersed and emulsified oil through various physical and chemical treatment technologies, selected based on the feed stream composition and discharge requirements (Gamwo *et al.*, 2022; Ghafoori *et al.*, 2022; Liu *et al.*, 2021). Offshore oil and gas facilities typically employ a combination of Hydrocyclones and IGF units, whereas Hydrocyclones serve as the primary separation step, using centrifugal force to separate oil and water based on density differences (Jiménez *et al.*, 2018; Amakiri *et al.*, 2022), and IGF units function as secondary treatment, injecting fine gas bubbles to capture remaining oil droplets and facilitate their removal by flotation (Olajire, 2020; Nasiri and Jafari, 2017; Igunnu and Chen, 2014). These systems operate under varying pressures, temperatures, and flow rates, making performance prediction and control complex.

The deoiling performance of HIGF systems is highly sensitive to operational parameters, particularly inlet flow rates and pressure conditions (Durdevic *et al.*, 2017). In practice, Hydrocyclone control is often based on Proportional-Integral-Derivative (PID) schemes that act on the pressure drop ratio (PDR), an indirect measure of separation performance. However, these control loops lack systematic tuning methodologies and generally require empirical calibration (Durdevic and Yang, 2018; Das and Jäschke, 2018). While CFD models have been used to analyse flow dynamics, they are not designed for real-time control implementation (Bram *et al.*, 2020; de Araújo *et al.*, 2020; Vallabhan *et al.*, 2020). Furthermore, conventional models often rely on simplifying assumptions that fail to capture the nonlinear and time-varying nature of real systems, resulting in limited adaptability and poor performance during process upsets or transient operations (Lee *et al.*, 2020; Hansen *et al.*, 2018; Bram *et al.*, 2018). These complexities make data-driven prediction tools particularly valuable for maintaining treatment efficiency across varying operational states of produced water treatment.

7.3. Machine Learning in Wastewater Treatment

ML has emerged as a transformative tool in chemical and environmental engineering, offering the ability to model nonlinear, multivariate systems without relying on rigid first-principles formulations (Lampropoulos, 2023; Aghbashlo *et al.*, 2021; Matheri *et al.*, 2021). In wastewater treatment, ML algorithms such as Random Forest (RF), Gradient Boosting Decision Trees (GBDT), Support Vector Machines (SVM) and Deep Neural Networks (DNN) have been widely applied to predict critical effluent quality indicators within wastewater treatment, including chemical oxygen demand (COD), biological oxygen demand (BOD), total suspended solids (TSS), and pH (El-Rawy *et al.*, 2021; Nourani *et al.*, 2021; Guo *et al.*, 2015). These models enable intelligent decision-making by supporting real-time monitoring, optimizing chemical dosing, and detecting process anomalies early, thereby enhancing overall treatment performance and ensuring regulatory compliance (Lowe *et al.*, 2022; Qambar and Al Khalidy, 2022; Wang *et al.*, 2022; Salles *et al.*, 2022; Ly *et al.*, 2022; Saboe *et al.*, 2021).

The selection of an appropriate ML algorithm depends on the volume and quality of available data, process complexity, and operational goals. Each model offers distinct advantages, and when thoughtfully applied, can yield actionable insights that improve system reliability and efficiency (Miller *et al.*, 2023). However, most ML applications to date have focused on municipal and biological treatment systems, with limited application to offshore produced water systems, particularly in predicting and controlling oil-in-water concentrations in Hydrocyclone and Induced Gas Flotation (IGF) units. Offshore environments present dynamic, high-variability conditions that challenge traditional control strategies and create opportunities for intelligent modelling. This study addresses this gap by applying interpretable, decision-tree-based ML models to predict oil-in-water concentrations in produced water in a floating production environment, a novel application aimed at enhancing real-time operational insight, treatment efficiency and regulatory compliance.

7.4. Ensemble-Based Decision-Tree Machine Learning Models

Decision tree-based machine learning models such as XGBoost, Gradient Boosting and Random Forest are well-suited for predicting continuous variables such as oil-in-water concentration, as they can capture non-linear interactions among input parameters and provide interpretability through feature importance rankings. This capability is especially advantageous in produced water treatment systems, where separation performance is sensitive to subtle variations in fluctuating operating conditions (Kaparthi and Bumblauskas, 2020).

While XGBoost and Gradient Boosting construct trees sequentially to minimise residual errors, Random Forest builds trees in parallel using bootstrapped subsets of the data and randomly selected features at each node split. These models are capable of capturing complex data patterns and estimating continuous outcomes, with their performance influenced by data complexity, noise levels, and computational demands (Velthoen *et al.*, 2023, Ali *et al.*, 2023).

XGBoost builds trees sequentially using a regularised objective function, optimised through a second-order Taylor expansion of the loss as shown in Equation 7.1 (Chen and Guestrin, 2016).

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \qquad \text{(Eq. 7.1)}$$

Where $f_t(x_i)$ is the prediction from the t^{th} tree and g_i and h_i are the first-order and second-order gradients of the loss with respect to the prediction and $\Omega(f_t)$ penalises model complexity. This formulation enables robust learning with strong generalisation, especially in noisy datasets. XGBoost supports both L1 (Lasso) and L2 (Ridge) regularisation and handles missing values during training. However, it may require extensive hyperparameter tuning and can be computationally demanding on large datasets (Velthoen *et al.*, 2023, Ali *et al.*, 2023, Lee *et al.*, 2023)

Gradient Boosting builds trees sequentially, where each learner focuses on correcting the residuals of the previous ensemble (Aziz *et al.*, 2020; Si and Du, 2020). The cumulative prediction after M iterations is defined in Equation 7.2.

$$\hat{y}_i^{(M)} = \sum_{m=1}^{M} \gamma_m h_m(x_i) \qquad \text{(Eq. 7.2)}$$

Where $h_m(x_i)$ is the m^{th} weak learner for the i^{th} instance, and γ_m is the learning rate controlling the contribution of each tree. Gradient Boosting is highly flexible and performs well across many domains but is sensitive to overfitting if the learning rate and tree depth are not carefully controlled (Aziz *et al.*, 2020; Bentéjac *et al.*, 2021)

Random Forest constructs trees in parallel using bootstrapped samples and random feature selection, with predictions averaged across all trees as represented in Equation 7.3 (Schonlau and Zou, 2020).

$$\hat{y} = \frac{1}{T}\sum_{T=1}^{T} h_t(x) \qquad \text{(Eq. 7.3)}$$

This ensemble averaging improves generalisation and reduces model variance. RF is particularly effective for noisy, high-dimensional data and tends to be less prone to overfitting. However, as the number of trees increases, interpretability and training speed may be impacted. (Schonlau and Zou, 2020; Speiser *et al.*, 2019).

7.5. Materials and Methods

7.5.1. Equipment and Process Description

A simplified schematic of the system and its control loops is shown in Fig. 7.1. The deoiling system evaluated comprises a horizontal Hydrocyclone followed by a vertical Induced Gas Flotation (IGF) unit, designed to reduce oil-in-water concentrations from levels above 2000 ppm to approximately 30 ppm.

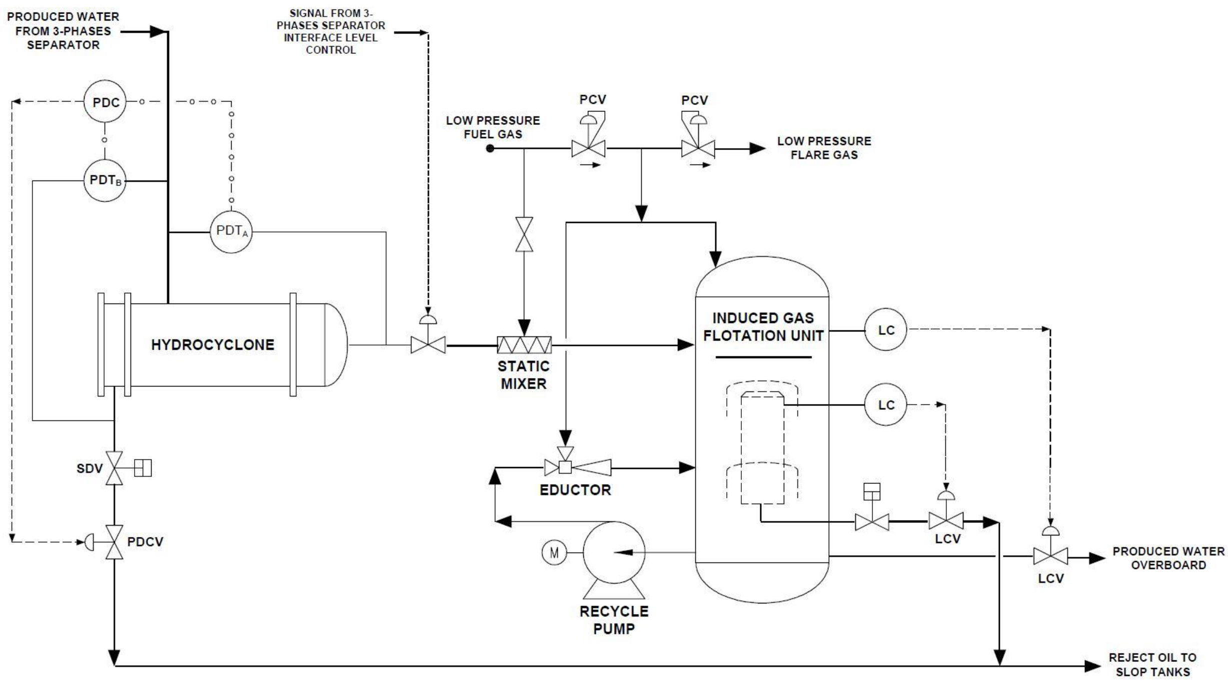


Figure 7.1: *Hydrocyclone–Induced Gas Flotation system schematic with control elements*

The Hydrocyclone is an 883 mm ID × 2012 mm long vessel fabricated from carbon steel with epoxy lining and duplex stainless-steel internals to provide corrosion resistance in harsh produced water environments. It contains 178 liners and is designed to handle flow rates between 212 and 667.2 m³/h, with an oil rejection capacity of 1.6 m³/h. It typically operates at

19 barg, with temperatures ranging from 36°C to 63°C. Produced water, chemically dosed with a clarifier, enters the liners tangentially, inducing a spiral motion that separates phases via centrifugal force: oil is concentrated along the central axis and exits through the overflow, while the treated water is discharged through the underflow.

A key aspect of operational stability is maintaining pressure balance across the system. Stable separation is achieved by controlling the differential pressure ratio (dP ratio) between the inlet-to-oil reject and inlet-to-water outlet. A pressure differential controller (PDC) receives signals from pressure transmitters (PDT_B and PDT_A) and modulates a control valve (PDCV) to maintain a setpoint near 2.0 bar. If the dP ratio increases (e.g., 3 bar), the valve opens to reduce reject pressure; if it decreases (e.g., 1.8 bar), the valve closes slightly to raise it. This feedback mechanism ensures consistent separation efficiency along the Hydrocyclone's axial geometry.

Water from the Hydrocyclone flows into the IGF unit (1828 mm ID × 3900 mm T/T), which operates with flow rates between 33 and 665.6 m³/h. Two centrifugal recycle pumps (each 133 m³/h, 10 barg, duty/standby) sustain internal circulation and drive gas–liquid mixing through an eductor, which draws flotation gas from the vessel headspace. To ensure flotation continuity during upset conditions or eductor malfunction, a static mixer on the inlet line disperses fuel gas from the topside low-pressure (LP) gas header into the incoming stream.

Flotation gas supply is regulated by self-actuated valves: PCVA (2.4 barg) for gas supply and PCVB (2.6 barg) for venting to the LP flare, maintaining vessel pressure near 2.5 barg. This control prevents pressure surges that could destabilise bubble formation and oil recovery. Gas bubbles attach to residual oil droplets, enhancing their buoyancy. Coalesced oil rises to the surface and is skimmed into a centrally located oil bucket, which is level-controlled and directed to the slop tank for disposal or reprocessing.

7.5.2. Model Selection and Data Preparation

Historical operational data from a floating production facility was used to train and validate the models predicting oil-in-water concentrations at the Hydrocyclone and IGF outlets. The final dataset consisted of 255 aggregated data points, each representing an average of three consecutive readings, derived from a total of 765 original data entries collected over approximately 64 days of offshore operation. The dataset was split into training and testing subsets using an 80:20 ratio, with the test data held back during model training and hyperparameter tuning to ensure an unbiased evaluation of generalisation performance.

For the Hydrocyclone outlet oil-in-water prediction, selected features included: ['*Operating Time (Day)', 'Operating Time (hrs)', 'Demulsifier Dosage (ppm)', "Separator's BS&W (%)", 'Hydrocyclone Liners Online (#)', 'Hydrocyclone Liners Blanked (#)', 'PW Inlet Flow Rate (m³/h)', 'PW Inlet Temperature (°C)', 'Water Clarifier (ppm)', 'Hydrocyclone Pressure (barg)', 'HP Separator Interface Level (%)', 'PW Inlet to Hydrocyclone OIW (ppm)', 'DP Inlet/Underflow (bar)', 'DP Inlet/Overflow (bar)', 'PDCV Opening (%)']*. For the IGF outlet OIW prediction, additional IGF-specific features were included: *['IGF Fuel Gas Pressure (barg)', 'IGF Fuel Gas Temp (°C)', 'IGF Oil Level (%)', 'IGF Bulk Liquid Level (%)']*.

The descriptive statistics for the selected variables used in model training are presented in Table 7.1.

***Table 7.1**: Descriptive statistics of process variables used for training and testing the Machine Learning models*

Parameter	Count	Null	Mean	Std	Min	25%	50%	75%	Max
Operating Time (Day)	255	0	32.38	18.4	1.0	16.0	32.0	48.0	64.0
Operating Time (hrs)	255	0	768.0	441.67	6.0	384.0	768.0	1152.0	1530.0
Demulsifier Dosage (ppm)	255	0	20.13	5.43	12.8	16.5	17.5	25.6	36.5
Separator Oil BS&W (%)	253	2	8.29	6.22	0.27	3.7	7.0	12.0	45.0
Hydrocyclone Liners Online (#)	255	0	133.0	0.0	133.0	133.0	133.0	133.0	133.0
Hydrocyclone Liners Blanked (#)	255	0	22.0	0.0	22.0	22.0	22.0	22.0	22.0
PW Inlet Flow Rate (m^3/h)	255	0	226.5	214.06	78.62	184.59	201.74	223.48	2201.05
PW Inlet Temperature (°C)	255	0	26.6	1.27	21.62	25.89	26.65	27.39	30.03
Water Clarifier (ppm)	254	1	16.3	4.35	10.0	13.9	15.7	17.13	48.0
Hydrocyclone Pressure (barg)	254	1	18.7	1.1	4.1	18.2	18.89	19.1	19.98
HP Separator Interface Level (%)	255	0	48.8	7.46	19.0	45.3	49.0	54.6	65.0
PW Inlet to Hydrocyclone OIW (ppm)	255	0	76.4	26.81	19.0	60.0	73.0	89.0	256.0
DP Inlet/Underflow (bar)	255	0	3.5	1.17	0.27	2.8	3.4	4.1	7.8
DP Inlet/Overflow (bar)	255	0	6.2	2.35	0.3	5.0	6.15	7.4	13.0
PDCV Opening (%)	255	0	6.1	3.21	0.0	3.5	6.3	8.5	18.1
PW Hydrocyclone Outlet/IGF Inlet OIW (ppm)	255	0	39.7	13.97	10.0	32.0	37.0	43.0	134.0
IGF Fuel Gas Pressure (barg)	250	5	1.6	0.17	1.25	1.46	1.53	1.7	2.31
IGF Fuel Gas Temp (°C)	248	7	51.0	6.95	18.03	48.76	49.61	50.81	70.68

IGF Oil Level (%)	250	5	34.9	5.81	22.62	32.36	33.28	35.38	67.84
IGF Bulk Liquid Level (%)	250	5	71.1	12.36	1.87	70.9	74.51	76.54	83.5
IGF Outlet (ppm)	253	2	21.5	6.51	6.0	19.5	22.0	24.0	84.0

Three ensemble, decision tree-based machine learning algorithms, XGBoost, Gradient Boosting and Random Forest, were selected for their ability to capture complex, non-linear relationships between process variables and target outputs while preserving model interpretability, which is particularly valuable in offshore operations, where explainable predictions are essential for regulatory transparency and operational troubleshooting. These models are well-suited to industrial environments due to their robustness, accuracy, and explainability. XGBoost and Gradient Boosting are particularly efficient for large-scale datasets due to their boosting framework, which iteratively corrects prediction errors. In contrast, Random Forest offers inherent stability through bootstrap aggregation and facilitates interpretation through feature ranking (Shujaaddeen *et al.*, 2024; Muraina, 2022; Ray, 2019; Medar *et al.*, 2017).

Alternative models were explored but not selected due to limitations in performance or practicality. Linear regression lacked the flexibility to model the system's non-linear behavior. Support Vector Machines were excluded due to high computational demands and poor scalability with larger datasets. Deep Neural Networks were avoided due to their black-box nature and tendency to overfit, which is particularly problematic in small-to-medium sized industrial datasets where interpretability is critical. Simpler models such as K-Nearest Neighbors and Naive Bayes were also excluded, as they are not well-suited for high-dimensional, continuous regression tasks.

As the oil-in-water concentration is a continuous variable, this study exclusively employed regression modelling approaches rather than classification. Regression models were deemed more appropriate for capturing the quantitative variation in oil-in-water levels at different stages of treatment, rather than simply classifying samples as compliant or non-compliant. Accordingly, only regression-based predictions were developed to estimate oil-in-water concentrations at both the Hydrocyclone and IGF outlets.

7.5.3. Methodology and Model Implementation

The dataset underwent a systematic pre-processing workflow to ensure completeness, consistency, and relevance of the input features. In offshore operational environments, sensor readings often include missing values due to hardware or communication interruptions. To preserve short-term trend continuity essential for accurate time-series prediction, missing values were imputed using a forward-fill method, replacing each gap with the most recent valid observation. The dataset was subsequently reviewed for null entries, data type mismatches, and column integrity. A correlation matrix was used to inform feature selection, reducing dimensionality, eliminating redundancy, and mitigating multicollinearity.

Outliers were detected using three complementary methods: the interquartile range (IQR), extended IQR and the 5th/95th percentile thresholds. These data points were retained rather than excluded, to allow the model to learn from extreme process scenarios that could influence treatment efficiency. Additionally, two derived features were engineered to represent the oil removal efficiency across the Hydrocyclone and IGF units. These indicators served to quantify treatment performance and enhance model interpretability.

No feature scaling was applied; models were trained on raw values in native units (e.g., bar, °C, m^3/h, ppm). Tree-based algorithms such as XGBoost, Gradient Boosting and Random Forest are scale-invariant, so using unscaled data preserved both physical interpretability and feature-importance transparency (Schonlau & Zou, 2020; Chen & Guestrin, 2016).

XGBoost, known for its scalability and efficiency, constructs decision trees sequentially, correcting errors from prior trees to capture complex nonlinear relationships. Its core hyperparameters, including learning rate, maximum tree depth and number of estimators, were optimised using grid search cross-validation, producing a configuration tailored to the dataset. Gradient Boosting, which uses a similar boosting framework, was implemented with manually selected parameters: 700 estimators, a maximum depth of 4, a learning rate of 0.01, and a minimum of 5 samples required to split an internal node. These values were chosen to balance bias and variance while maintaining interpretability. The Random Forest model followed a bagging strategy, generating 1000 trees trained on bootstrapped subsets of the data. Remaining parameters were left at default settings to leverage the model's stability and reduce tuning complexity.

Feature importance was assessed using internal criteria specific to each model: frequency and gain for XGBoost and Gradient Boosting, and mean decrease in impurity for Random Forest. These metrics highlight each feature's relative contribution to predictive accuracy. In this study, only regression modelling was performed to estimate oil-in-water concentrations at various stages of treatment.

7.6. Results

To guide feature selection and improve model input quality, Pearson correlation methods was applied (Fig. 7.2). The resulting matrix aided in eliminating redundant features and understanding inter-variable dynamics and reinforced the choice of decision-tree-based models capable of handling multicollinearity and non-linear dependencies.

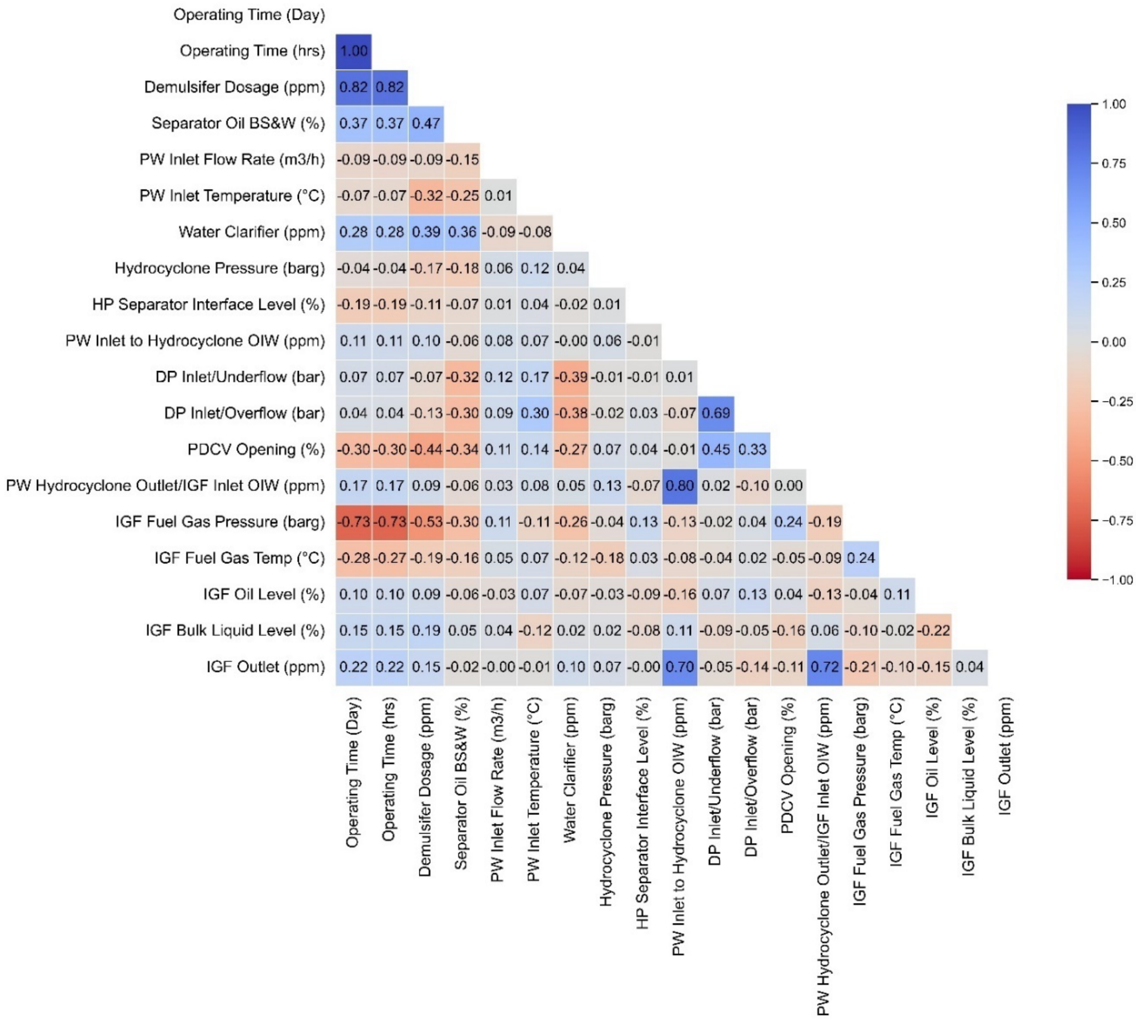


***Figure 7.2**: Pearson correlation matrix for key process variables.*

Pearson correlation revealed strong linear relationships among key process variables. Notably, the *PW Inlet to Hydrocyclone OIW (ppm)* exhibited high correlation with the *PW Hydrocyclone Outlet/IGF Inlet OIW (ppm)* (r = 0.80), which in turn strongly correlated with the *IGF Outlet* (r = 0.72). These results confirm the strong influence of upstream contamination levels on downstream treatment performance. Pressure differentials across the Hydrocyclone's inlet and reject lines also showed significant correlation (r = 0.69), reflecting internal process coupling, which may influence separation efficiency under fluctuating flow or pressure conditions.

7.6.1. Models Performance

Model performance was evaluated using three widely accepted regression metrics: Root Mean Squared Error (RMSE), which places greater emphasis on large prediction deviations; Mean Absolute Error (MAE), which captures the average magnitude of residuals; and the Coefficient of Determination (R^2), which quantifies the proportion of variance in the target variable explained by the model.

To assess model sensitivity to hyperparameters and tune performance, tuning curves were generated by systematically varying key parameters, namely, maximum tree depth, number of estimators, and minimum child weight. Each model configuration was trained on the full training set, and performance was assessed using RMSE, MAE, and R^2 across three data partitions: the training set, a validation score derived via 5-fold cross-validation, and the held-out test set. The cross-validation process involved splitting the training data into five folds (K = 5), iteratively training on four folds and validating on the fifth, offering a robust estimate of generalisation while preventing data leakage into the test set. This multi-metric, multi-split evaluation strategy provided a detailed understanding of how hyperparameter choices influence both prediction error magnitude and explanatory power across unseen data. It also supported the selection of optimal configurations that offered strong generalisation, serving as the foundation for comparative analysis of model performance across the Hydrocyclone and IGF treatment stages.

7.6.1.1. Hydrocyclone

A comprehensive comparison of the three machine learning models, XGBoost, Gradient Boosting and Random Forest for predicting oil-in-water concentrations at the Hydrocyclone outlet are presented in Fig. 7.3. The performance of the XGBoost model is illustrated in the top row, where it achieved an RMSE of 6.33 ppm, an MAE of 4.98 ppm, and an R^2 of 0.732, indicating that approximately 73.2% of the variance in the target variable was explained. The scatter and KDE plots reveal strong predictive accuracy within the mid-range concentration band (20–50 ppm), although the model tends to underpredict values above 50 ppm and slightly overpredict those below 20 ppm. The Gradient Boosting model, shown in the middle row, yielded comparable metrics (RMSE = 6.25 ppm, MAE = 4.92 ppm, R^2 = 0.738). Its predictions closely mirrored those of XGBoost, exhibiting similar patterns of underestimation at higher concentrations and a slightly wider spread at the lower end.

Random Forest model achieved an RMSE of 6.37 ppm, an MAE of 4.90 ppm, and an R^2 of 0.729. While performance was consistent in the 30–50 ppm range, the model showed greater dispersion and underprediction at higher concentrations, a characteristic limitation frequently associated with non-boosted ensemble methods when handling extreme observations.

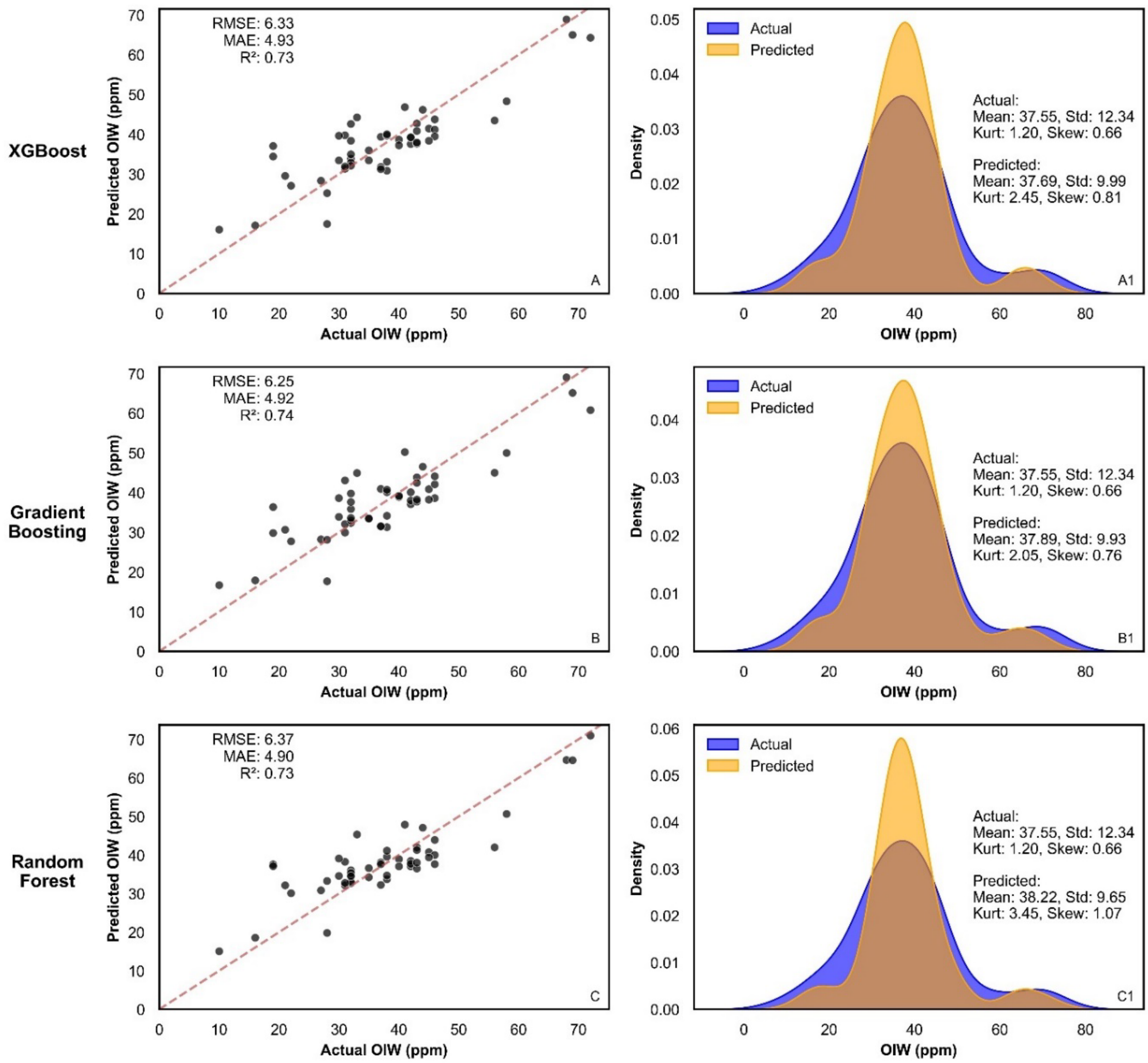


Figure 7.3: *Models performance for oil-in-water prediction at the Hydrocyclone outlet*

Fig. 7.4 presents a comparative residual analysis of the models for Hydrocyclone outlet predictions. Each row displays a time-series overlay (left) of actual versus predicted oil-in-water concentrations and a corresponding residual scatter plot (right) with LOWESS smoothing. The red dashed line denotes the zero-error reference. The XGBoost model shows strong temporal alignment between actual and predicted values, with residuals mostly confined within ±10 ppm and no clear systematic bias. Gradient Boosting also tracks the main trend effectively, though its residual plot reveals a slight underprediction at higher concentrations, as evidenced by the upward drift in the smoothing curve beyond 40 ppm. The Random Forest model, shown with moving-average smoothing for better trend visibility, captures broader process dynamics but exhibits greater variance and its residuals display more dispersion, particularly within the operationally critical 30–50 ppm range, and a tendency to overshoot at turning points.

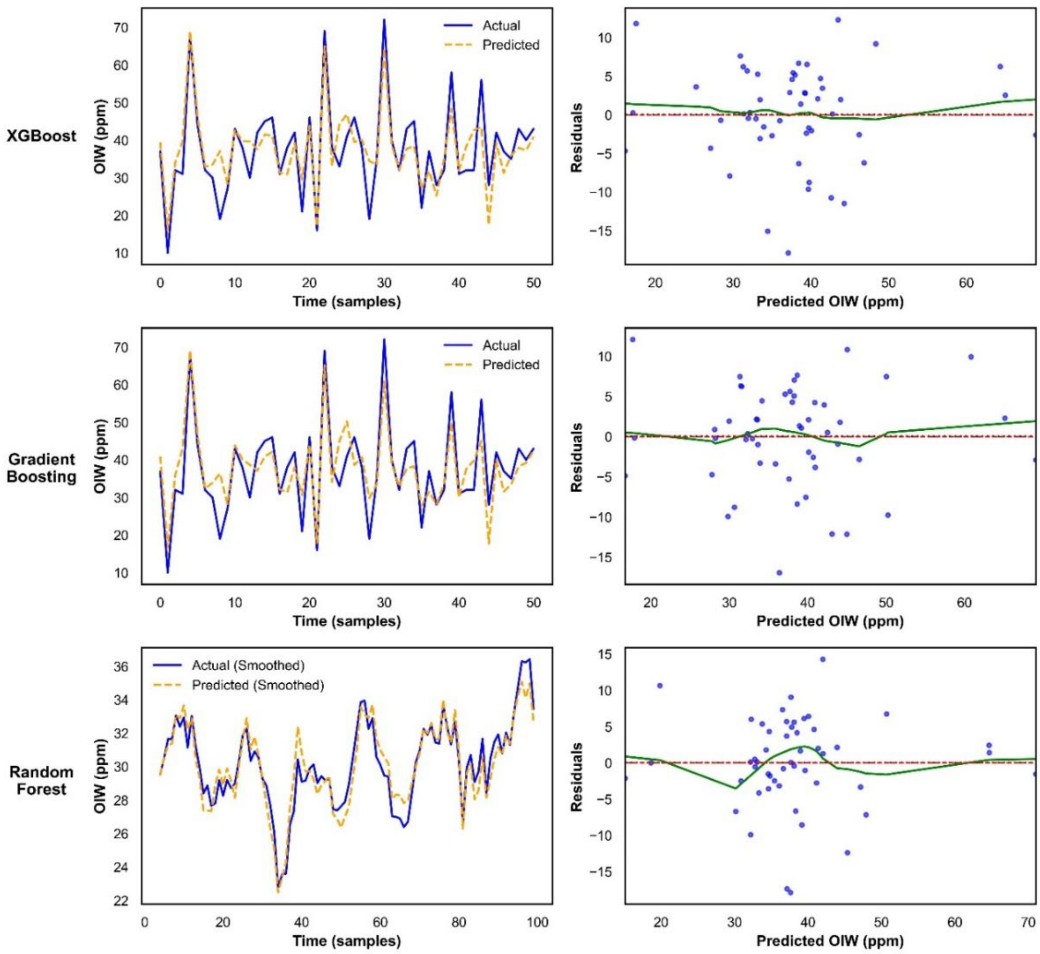


Figure 7.4*: Residual and time-series plots for Hydrocyclone oil-in-water predictions. Left: actual vs. predicted oil-in-water over time. Right: residuals with LOWESS smoothing and zero-error reference line.*

Feature importance is a key concept in predictive modelling, offering insights into how individual features influence model predictions. Different ensemble methods assess importance using various metrics. XGBoost commonly uses the F score, which counts how often a feature is used for data splits, along with alternatives like gain (improvement in accuracy) and cover (number of observations affected). Random Forest and Gradient Boosting typically rely on Mean Decrease in Impurity (MDI), which reflects the reduction in variance or Gini impurity, while permutation importance (Mean Decrease in Accuracy) provides a model-agnostic measure by evaluating the drop in performance when feature values are randomly shuffled. These metrics help identify influential predictors, guide feature selection, and enhance model interpretation, ultimately improving decision-making in applied contexts (Shujaaddeen *et al.*, 2024; Muraina, 2022; Ray, 2019; Medar *et al.*, 2017).

Feature importance rankings are shown in Fig. 7.5 to illustrate each model's performance in predicting the Hydrocyclone's outlet oil-in-water concentration. The bar plots quantify each model's internal assessment of feature contribution to predictive performance. The *PW Inlet to Hydrocyclone OIW (ppm)* consistently emerged as the dominant feature, with importance scores of 0.48 (XGBoost), 0.70 (Gradient Boosting), and 0.68 (Random Forest). This reflects the direct dependency of outlet quality on the incoming c*ontamination load. Additional influential features included Operating Time (in days or hours), Water Clarifier dosage and PW Inlet Temperature (°C)*, which likely represent long-term process drift, chemical control strategy and thermal behaviour, respectively.

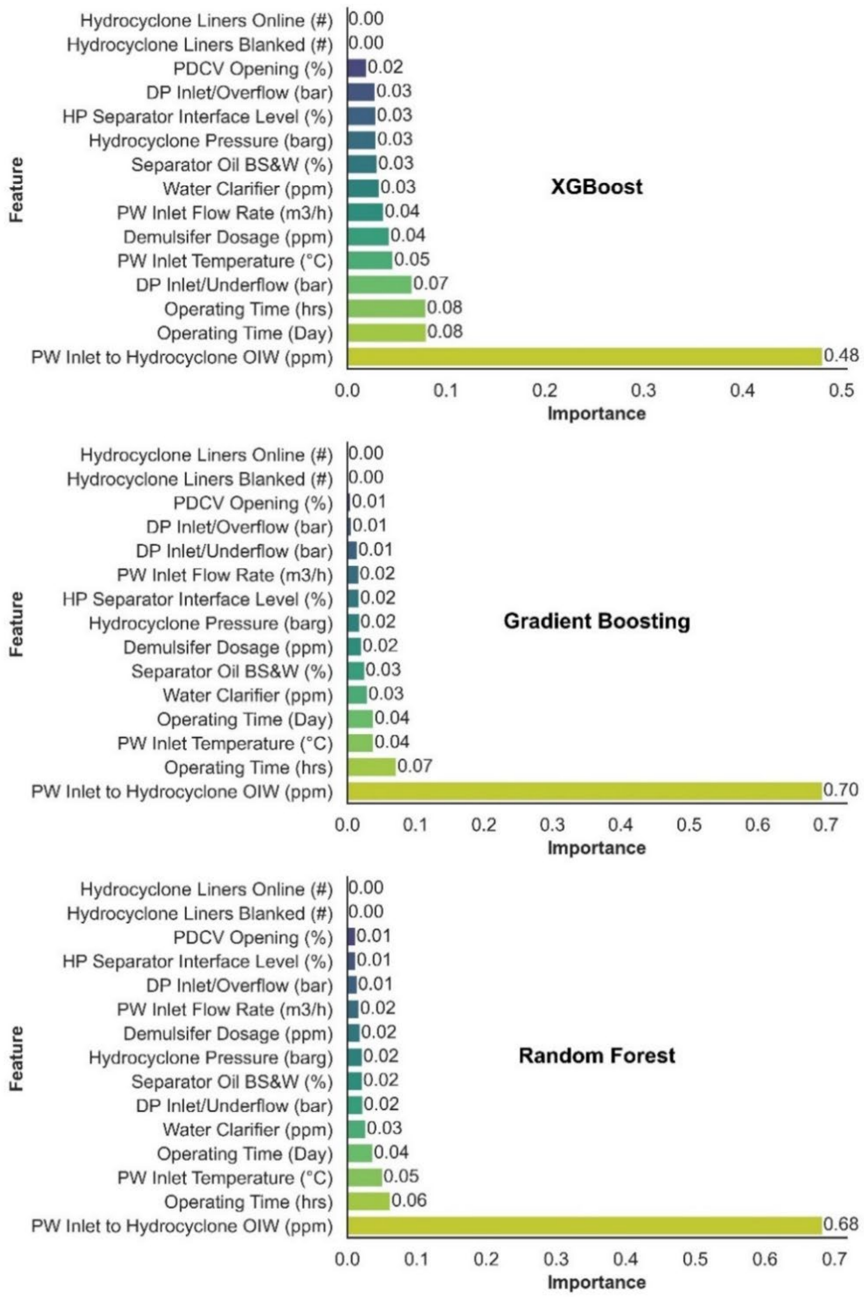


Figure 7.5: *Feature importance ranking for Hydrocyclone outlet oil-in-water prediction*

The partial dependence plots provide a two-dimensional representation of predicted oil-in-water concentrations based on the two most influential features: Inlet oil-in-water and operating time (Figure 7.6). The x-axis corresponds to the inlet concentration to the Hydrocyclone (ppm), while the y-axis reflects cumulative operating time—measured in days for XGBoost, and in hours for Gradient Boosting and Random Forest. Colour gradients illustrate the predicted outlet oil-in-water concentration, with warmer tones indicating higher concentrations.

In the XGBoost model, for instance, an inlet concentration of 60 ppm and an operating time of 40 days corresponds to a predicted outlet oil-in-water concentration of approximately 34–35 ppm. As inlet concentrations increase beyond 100 ppm, the model predicts progressively higher outlet values, with contours becoming denser and shifting toward warmer tones. Variation along the operating time axis remains relatively limited, as indicated by the mostly vertical orientation of the contour lines. This suggests that inlet quality is the dominant predictor of separation efficiency, while cumulative operating time plays a secondary role. Interestingly, regions of lower predicted oil-in-water concentration (<35 ppm) are concentrated in the lower-left quadrant, corresponding to both low inlet concentration and short operating time, suggesting this model captures degradation trends effectively.

The Gradient Boosting model, using operating time in hours, yields a more smoothed and conservative surface, with lower predicted outlet values overall. The predicted oil-in-water concentration remains near or below 10 ppm for the majority of the surface when inlet oil-in-water concentration is under 70 ppm, regardless of operating time. However, a distinct region of sudden increase appears around 100 ppm, reinforcing the model's recognition of threshold-dependent separation breakdown. Its behaviour suggests the model may prioritise short-term performance drivers over cumulative degradation.

The Random Forest model, while consistent in identifying the 60–70 ppm transition, exhibits stepped prediction zones due to its discrete split-based architecture. The predicted oil-in-water concentration values increase rapidly with both inlet concentration and prolonged operating hours. The upper-right quadrant reflects predicted outlet oil-in-water concentration values above 60 ppm, which may indicate performance deterioration or hydraulic short-circuiting under persistent load conditions.

Collectively, Fig. 7.5 and Fig. 7.6 provide interpretable confirmation that the Hydrocyclone's performance is highly contingent on inlet oil concentration and is further influenced by operating duration. Monitoring and controlling these two variables is essential for maintaining treatment efficiency and avoiding exceedance of discharge limits. Fig. 6 also reveals that while all models agree on general trends, they differ in how they represent the impact of cumulative operational time, which has implications for model selection depending on whether short-term control or long-term degradation is the dominant concern.

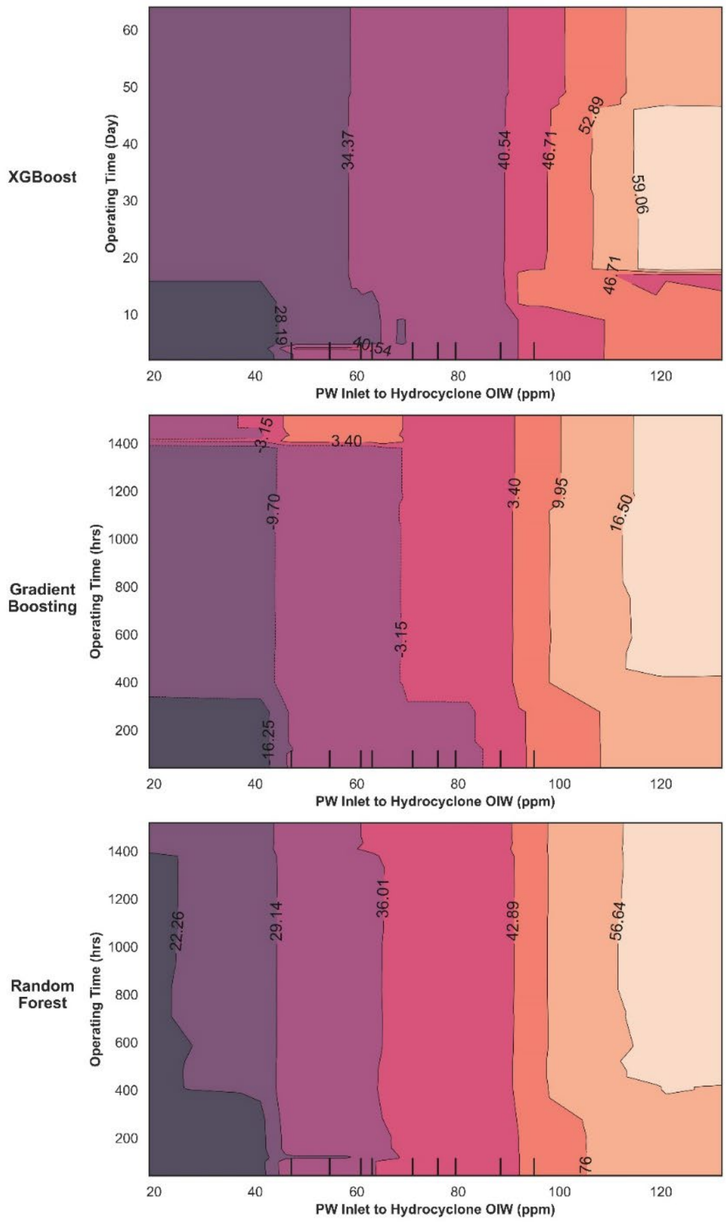


***Figure 7.6**: Joint influence of PW Inlet to Hydrocyclone OIW (ppm) and Operating Time on predicted outlet OIW concentrations from Hydrocyclone. Colour gradients indicate predicted OIW levels, with warmer tones representing higher concentrations. Contour lines highlight prediction intervals, illustrating how model outputs change across the feature space*

7.6.1.2. Induced Gas Flotation Predictions

The performance of XGBoost, Gradient Boosting, and Random Forest models in predicting oil-in-water concentrations at the IGF outlet is illustrated in Fig. 7.7. Each scatter plot includes summarised evaluation metrics, while the corresponding Kernel Density Estimate (KDE) distributions are displayed in the adjacent right-hand panels.

The XGBoost model demonstrated the strongest overall performance in predicting oil-in-water concentrations at the IGF outlet, achieving an RMSE of 4.14 ppm, an MAE of 2.86 ppm, and an R^2 of 0.54. These results indicate that the model yields relatively low prediction errors, with the MAE reflecting an average deviation of just under 3 ppm and the RMSE emphasizing that larger errors are not predominant. The R^2 value suggests that 54% of the variability in oil-in-water concentrations is explained by the model, capturing key trends despite some remaining unexplained variance. Visual inspection supports these findings: the scatter plot shows good agreement with the identity line, especially in the 15–30 ppm range, although underprediction becomes more pronounced at higher concentrations. The KDE plot further confirms this behavior, with a predicted distribution that is more peaked and left-skewed (Skew = –1.20), indicating a tendency to compress higher values while overrepresenting mid-range concentrations. Nonetheless, the closeness between the predicted and actual mean values (20.29 vs. 20.63 ppm) underscores the model's solid average performance across the dataset.

Gradient Boosting yielded the weakest performance among the three models, with an RMSE of 5.38 ppm, MAE of 3.32 ppm, and an R^2 of just 0.22. While the model captures the general trend, its predictions exhibit greater dispersion, especially at lower concentrations (<20 ppm), and several outliers can be observed. The KDE plot reveals pronounced overprediction in the mid-to-high range, with predicted values showing higher kurtosis (11.18) and strong right-skew (2.09), indicating systematic bias toward overestimation of oil-in-water concentration. This behaviour may reduce the model's usefulness in operational settings where precision near the discharge limit is critical.

The Random Forest model performed moderately well, achieving an RMSE of 4.32 ppm, MAE of 3.00 ppm, and R^2 of 0.50. Like XGBoost, it shows good agreement in the mid-range (15–30 ppm), though with slightly more dispersion. Its KDE curve closely tracks the actual distribution, with minor deviations and reduced skew (-0.46), indicating more symmetric predictions around the mean. The model underpredicts slightly at higher concentrations but avoids the extreme overprediction tendencies seen in Gradient Boosting.

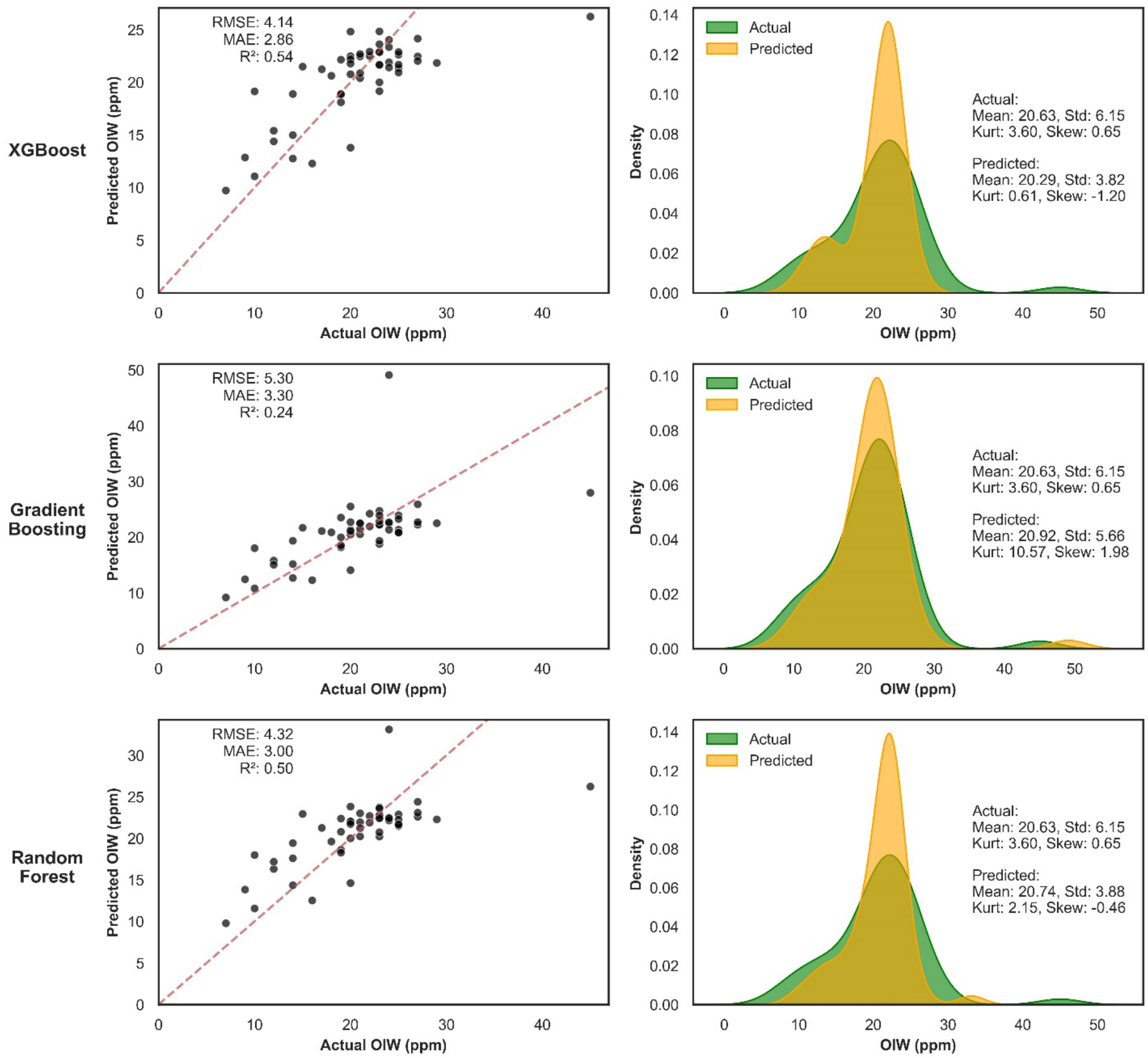


***Figure 7.7**: Models performance for* oil-in-water *concentration prediction at the Induced Gas Flotation outlet*

Fig. 7.8 presents the residual analysis for IGF outlet oil-in-water concentration predictions using XGBoost (A), Gradient Boosting (B), and Random Forest (C). Each subplot plots residuals (Actual − Predicted) against predicted OIW values, with a LOWESS smoothing curve (green) and a horizontal dashed red line indicating the zero-error reference. Insets show time-series overlays comparing actual and predicted OIW concentrations.

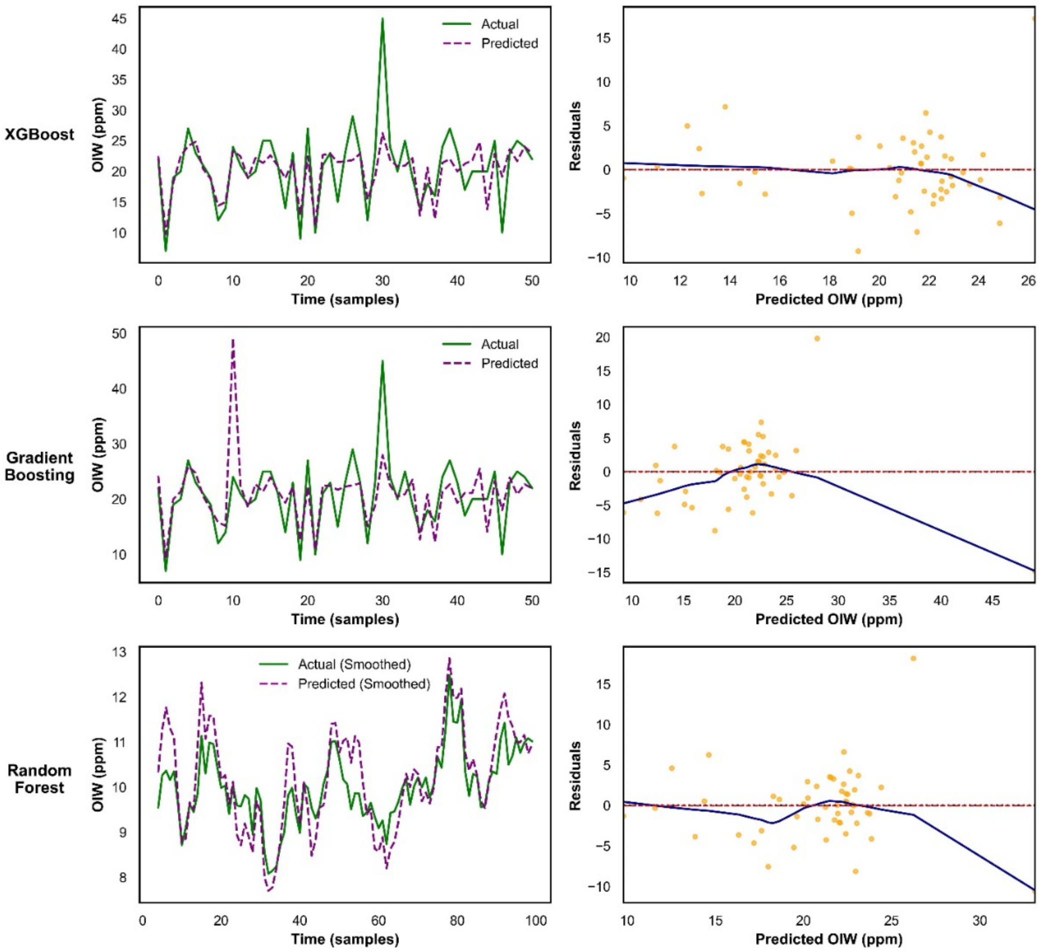


Figure 7.8: Residual and time-series plots for Induced Gas Flotation oil-in-water concentration predictions. Left: actual vs. predicted oil-in-water concentration over time. Right: residuals with LOWESS smoothing and zero-error reference line.

The XGBoost model demonstrates tight clustering of residuals around zero across most of the prediction range, with no pronounced curvature or heteroskedasticity. Most residuals fall within ±10 ppm, and the LOWESS line remains relatively flat, indicating stable performance with minimal bias. The time-series inset further confirms close tracking between actual and predicted values, especially in the mid-range.

In contrast, the Gradient Boosting model shows a more uneven residual distribution. The smoothing curve reveals an upward trend followed by a sharp decline beyond 24 ppm, indicating systematic overprediction in the 15–25 ppm range and subsequent underestimation at higher values. The time-series inset displays stronger fluctuations and a tendency to miss rapid changes in actual oil-in-water concentration, which contributes to its lower R^2 score and higher error metrics. The Random Forest model displays moderate residual spread with reasonably flat residuals across the predicted range. Some underprediction is evident at the upper end (>30 ppm), though the smoothing curve remains less volatile than that of Gradient

Boosting. The time-series inset shows smoothed predictions closely tracking the actual trend, with less overshooting compared to the other models.

Feature importance rankings in Fig. 7.9 show that the *PW Inlet to Hydrocyclone OIW (ppm)* is consistently the most influential predictor at the IGF outlet, with importance scores of 0.45, 0.48, and 0.53 for XGBoost, Gradient Boosting and Random Forest, respectively. This reflects the critical role of inlet quality in downstream performance. *IGF Oil Level (%)* also ranked highly, particularly in the Gradient Boosting and Random Forest models, underscoring its influence on retention time and separation efficiency.

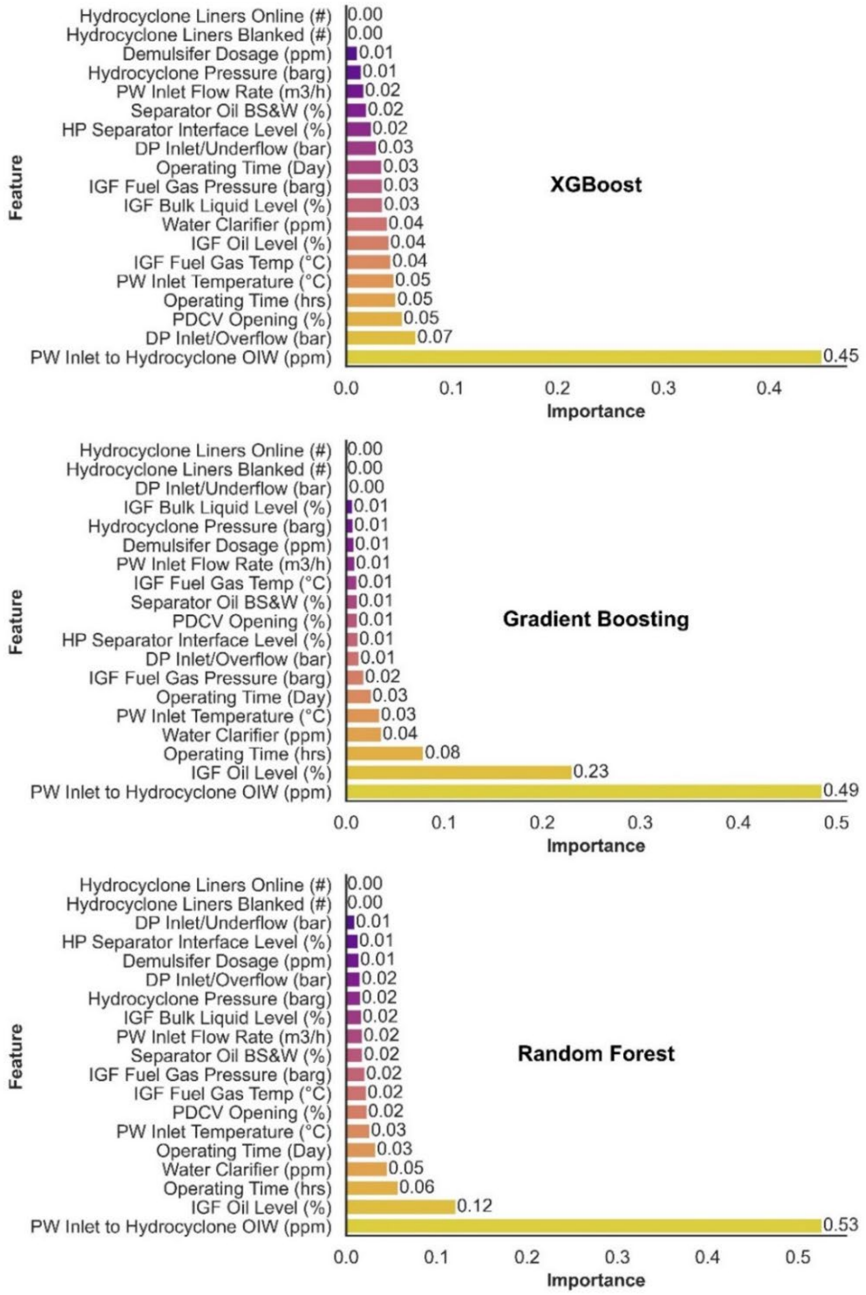


Figure 7.9*: Feature importance and partial dependence plots for IGF outlet oil-in-water concentration prediction using (A) XGBoost, (B) Gradient Boosting and (C) Random Forest.*

In all three models, the *PW Inlet to Hydrocyclone OIW (ppm)* consistently emerged as the most influential predictor at the IGF outlet, with importance scores of 0.45, 0.48, and 0.53 for XGBoost, Gradient Boosting, and Random Forest, respectively. This outcome is expected, as the IGF unit follows the Hydrocyclone in the deoiling sequence, making the quality of the inlet stream a key determinant of performance. IGF Oil Level (%) also ranked prominently, especially in the Gradient Boosting and Random Forest models, highlighting its role in influencing retention time and enhancing separation efficiency within the IGF unit.

Fig. 7.10 depicts two-dimensional Partial Dependence Plots (PDPs) that highlight the interaction effects of key operational variables on the predicted oil-in-water concentration at the IGF outlet across the three machine learning models. The XGBoost model, in particular, reveals a pronounced response surface, where predicted oil-in-water concentration rise sharply when the inlet concentration surpasses 60 ppm, especially under low differential pressure conditions (2–6 bar). This region indicates a critical transition zone, suggesting that reduced separation pressure gradients may hinder flotation performance under high contaminant loading. The nonlinear patterns captured by XGBoost underscore the compounded impact of upstream variability on the efficiency of downstream polishing stages.

In contrast, the Gradient Boosting model shows a smoother dependency, indicating that lower IGF oil levels (<30%) are associated with higher predicted outlet oil-in-water concentration. This suggests that maintaining a moderate oil blanket within the IGF may enhance coalescence and separation, likely due to improved phase stability or reduced turbulence. Notably, the lowest predicted oil-in-water concentration occur when both the inlet concentration and oil level remain within mid-range operating envelopes, reinforcing the importance of operational balance.

The Random Forest model exhibits more stepwise transitions and segmented decision boundaries, consistent with its tree-based structure. It identifies similar thresholds for PW Inlet OIW (~60 ppm) and IGF oil level (~30%) as critical inflection points. However, its piecewise nature implies that predictions may be more sensitive to slight variations near these thresholds, which could lead to abrupt shifts in predicted output.

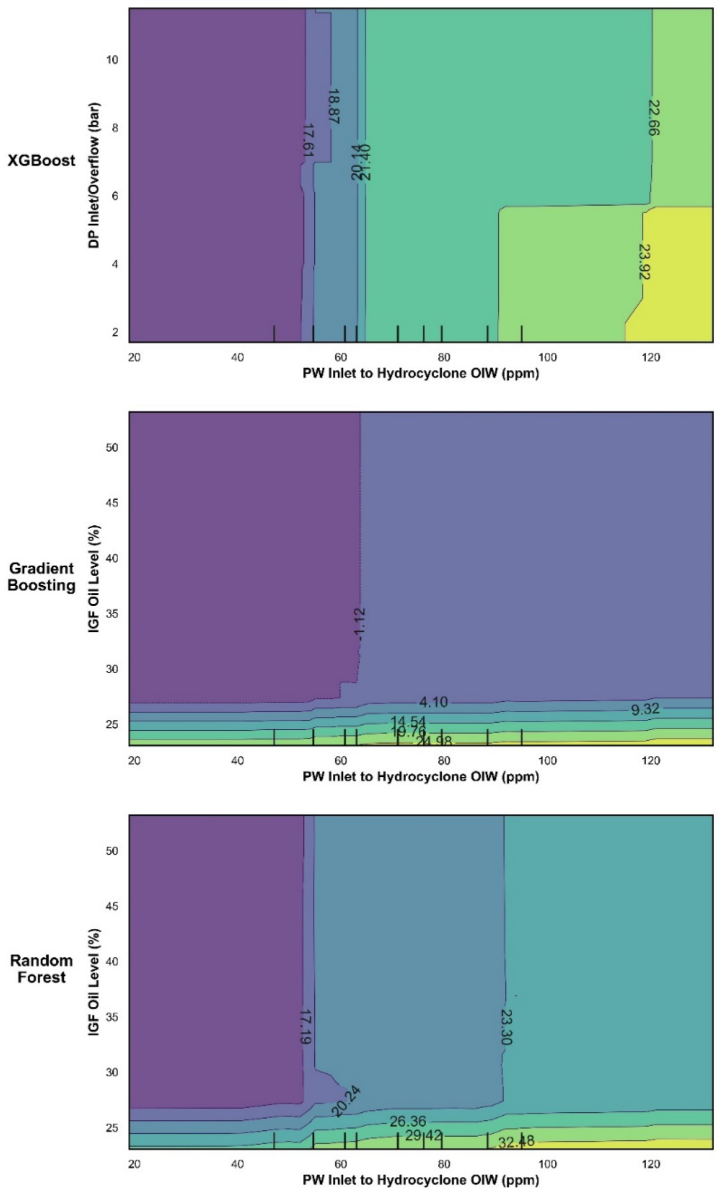


Figure 7.10*: Joint influence of PW Inlet to Hydrocyclone OIW (ppm) and Hydrocyclone's DP Inlet/Overflow and IGF Oil Level on predicted outlet OIW concentrations from IGF. Colour gradients indicate predicted OIW levels, with warmer tones representing higher concentrations. Contour lines highlight prediction intervals, illustrating how model outputs change across the feature space*

At the Hydrocyclone outlet, Gradient Boosting slightly outperformed the other models with an R^2 of 0.74 and RMSE of 6.26 ppm, while XGBoost and Random Forest performed similarly with R^2 values of 0.73 and RMSEs of 6.33 and 6.37 ppm, respectively. For the IGF outlet, XGBoost demonstrated the best performance (RMSE: 4.14 ppm, MAE: 2.86 ppm, R^2: 0.54), closely followed by Random Forest (RMSE: 4.32 ppm, R^2: 0.50). Gradient Boosting underperformed across all IGF metrics, particularly at higher oil-in-water concentrations, with an R^2 of 0.22 and RMSE of 5.38 ppm. The reduced accuracy observed at higher oil-in-water concentrations highlights the need for model refinement, enhanced sensing strategies, and potential integration with physical process models. Addressing these limitations is essential not only for improving prediction reliability under extreme conditions, but also for enabling real-time deployment and adaptive control in offshore produced water treatment systems.

7.7. Conclusion

This study developed a data-driven approach to predict oil-in-water concentrations in offshore produced water treatment systems using ensemble-based machine learning models. Comparing XGBoost, Gradient Boosting, and Random Forest, the analysis showed that XGBoost and Gradient Boosting achieved the best predictive performance at the IGF and Hydrocyclone outlets, respectively, effectively capturing nonlinear relationships in Hydrocyclone–Induced Gas Flotation (HIGF) systems. However, model accuracy declined at higher concentrations, indicating potential for refinement through enhanced feature engineering, physical constraints or hybrid modeling.

The performance outcomes align with established research showing that ensemble algorithms, particularly XGBoost, Gradient Boosting, and Random Forest, consistently outperform other models in effluent quality prediction (Liu *et al.*, 2022; Safeer *et al.*, 2022; Bentéjac *et al.*, 2021; Nourani *et al.*, 2021). This corroboration reinforces their reliability for offshore produced water modelling and confirms their applicability under full-scale operating conditions.

From an engineering perspective, the findings support more adaptive and informed control strategies for offshore water treatment. Using historical data for prediction enables better process optimisation, chemical dosing efficiency, and real-time compliance monitoring. The interpretability of decision tree-based models enhances transparency and accountability, critical for regulated offshore operations. Although based on data from a single facility, the methodology is scalable and suitable for integration into digital twin platforms. Future work should prioritise real-time implementation and validation under variable production conditions to advance intelligent water treatment aligned with sustainability and safety objectives.

CHAPTER 8

MACHINE LEARNING-AIDED MONTE CARLO SIMULATION FOR COMPLIANCE RISK AND PERFORMANCE ASSESSMENT IN OFFSHORE PRODUCED WATER DEOILING

"With man this is impossible, but with God all things are possible."

Matthew 19:26

CHAPTER 8: MACHINE LEARNING-AIDED MONTE CARLO SIMULATION FOR COMPLIANCE RISK AND PERFORMANCE ASSESSMENT IN OFFSHORE PRODUCED WATER DEOILING

8.1. Introduction

Produced water is the largest liquid waste stream from offshore oil and gas operations, often exceeding hydrocarbon output in mature fields (César *et al.*, 2024; Gamwo *et al.*, 2022; Al-Ghouti *et al.*, 2019). Its complex composition including dispersed and suspended oils, creates significant environmental and operational challenges (Ghafoori *et al.*, 2022; Liu *et al.*, 2021). To meet strict discharge regulations, offshore platforms rely on multistage systems like hydrocyclones and induced gas flotation (IGF) units. While effective, these systems are sensitive to operational variability such as flow, pressure and chemical dosing, often resulting in inconsistent oil removal and potential noncompliance (Bram *et al.*, 2020; Al-Ghouti *et al.*, 2019).

Traditional deterministic models and computational fluid dynamics (CFD) approaches offer useful insights but fall short in representing real-world variability and quantifying compliance risks (Hejabi *et al.*, 2021; Das e*t al.*, 2025). Recent data-driven models, particularly machine learning (ML), have improved predictive accuracy but generally focus on point predictions without explicitly accounting for uncertainty.

To address this gap, this chapter developed and validates a Machine Learning–aided Monte Carlo Simulation (ML-MCS) framework for offshore produced water deoiling systems. Specifically, it optimised the XGBoost model for predicting oil-in-water concentrations as depicted in Chapter 7, Section 7.6.1; Integrated statistically fitted distributions of key operational variables to quantify compliance probabilities and exceedance risks; and applied SHapley Additive exPlanations (SHAP) analysis to explain feature influences. Finally, it assesses the fidelity of simulated outputs against actual data to ensure both predictive accuracy and statistical representativeness, supporting improved performance monitoring and proactive compliance management under uncertain conditions.

8.2. Bridging Deterministic Limitations with Stochastic Simulation

Most modeling efforts rely on deterministic assumptions based on average conditions, limiting their ability to capture performance variability and the risk of exceedances (Hejabi *et al.*, 2021;

Belia *et al.*, 2009). While machine learning models, including neural networks and ensemble methods, have improved point prediction accuracy, these models often lack mechanisms to represent uncertainty. Similarly, mechanistic models like computational fluid dynamics provide detailed insight into separation behavior, but are computationally intensive and not suited for system-wide uncertainty analysis (Das *et al.*, 2025; Alardhi *et al.*, 2024; Ibrahim *et al.*, 2021; Jiang *et al.*, 2021; Wang *et al.*, 2021; de Oliveira-Junior *et al.*, 2020; Li *et al.*, 2020).

Stochastic simulation methods, particularly Monte Carlo simulation (MCS), offer a robust framework for evaluating system behavior under uncertainty. MCS has been widely applied in chemical risk analysis, environmental modelling, water and wastewater treatment as well as marine discharge forecasting (Balas, 2023; Chen *et al.*, 2011; Zhao, 2012; Chen *et al.*, 2010; Zhao *et al.*, 2008; Niu, 2008; Chen *et al.*, 2007; Zhao, 2007; Husain *et al.*, 2004).

Despite its utility, MCS has seen limited application in offshore produced water systems. In particular, few studies have used stochastic simulation to quantify the compliance risk associated with hydrocyclone and IGF performance under realistic operating variability. This represents a critical gap, as probabilistic modeling can provide a more realistic basis for performance evaluation and risk-aware operational strategies.

8.3. Methodology

This study employed a structured machine learning–aided Monte Carlo simulation framework to model oil-in-water (OIW) concentration dynamics in offshore produced water deoiling systems, focusing on the hydrocyclone and induced gas flotation (IGF) units. Operational data from an offshore platform (as detailed in Table 7.1, Chapter 7) were preprocessed by handling missing values through forward-fill imputation and identifying outliers using the interquartile range (IQR) and percentile-based methods. All analyses, including data preprocessing, model training, and Monte Carlo simulations, were conducted in Python 3.10 using libraries such as Scikit-learn, XGBoost, Optuna, NumPy, Pandas, and Matplotlib. Statistical distribution fitting and probabilistic sampling were implemented with SciPy and the random module.

Predictive models were developed using the XGBoost regression algorithm, identified in Chapter 7, Section 7.6.1 as one of the most suitable for produced water deoiling, with hyperparameter tuning performed via Optuna, an automated optimisation framework that efficiently searches for optimal parameter combinations using a sequential, trial-based approach, leveraging 5-fold cross-validation to maximise the coefficient of determination and identify ideal values for learning rate, number of estimators, maximum tree depth and regularisation terms.

To extend the framework into a probabilistic domain, empirical distributions of input features were statistically characterised to generate realistic data. The best-fitting distribution for each feature was selected based on goodness-of-fit scores, enabling the creation of 10,000 synthetic input samples for Monte Carlo simulations. These inputs were passed through the respective hydrocyclone and IGF models to predict oil-in-water concentrations under stochastic operating conditions, supporting comprehensive assessment of performance variability and compliance risk.

An overview of these methodological stages is summarised in Fig. 8.1, which illustrates the integrated workflow from data preparation to probabilistic risk analysis.

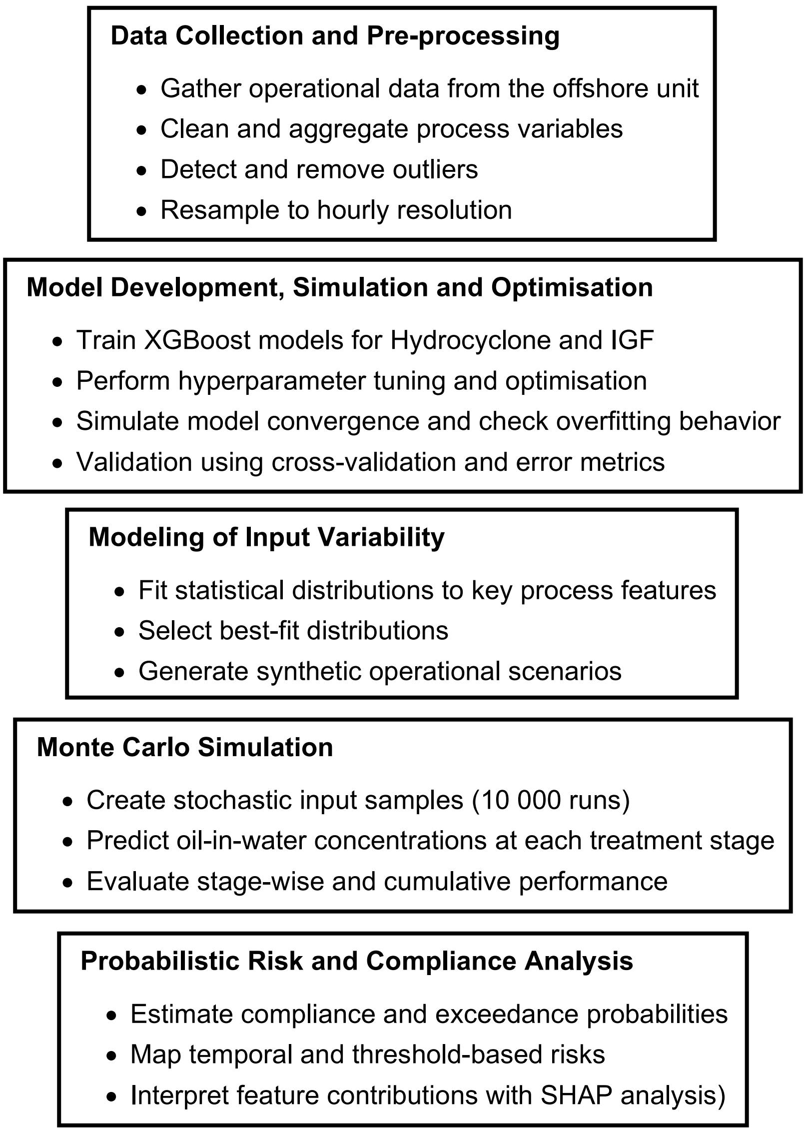


Figure 8.1: *Workflow of the ML-aided Monte Carlo simulation framework applied*

8.4. Results and Discussion

8.4.1. XGBoost Model Performance

Fig. 8.2 presents a comparison of predicted and actual oil-in-water concentrations for both XGBoost models of the Hydrocyclone (Fig. 8.2A) and IGF (Fig. 8.2B) treatment units. The Hydrocyclone model exhibits a strong linear relationship between predicted and observed values. The points are closely clustered around the ideal 1:1 line and the regression fit demonstrates a coefficient of determination (R^2 = 0.780), indicating that the model captures approximately 78% of the variance in the data. Additionally, the error metrics RMSE = 5.73 ppm and MAE = 4.90 ppm, reflect a high level of predictive accuracy and consistency across the validation set.

In contrast, the IGF model displays a more dispersed pattern. While the general trend is preserved, the predictions show greater deviation from the 1:1 line, particularly at higher concentrations. This is corroborated by an R^2 value of 0.588, suggesting moderate predictive power. The RMSE of 3.91 ppm and MAE of 2.71 ppm indicate that, although the absolute errors are smaller due to the narrower concentration range, the model's ability to explain variability is comparatively less robust..

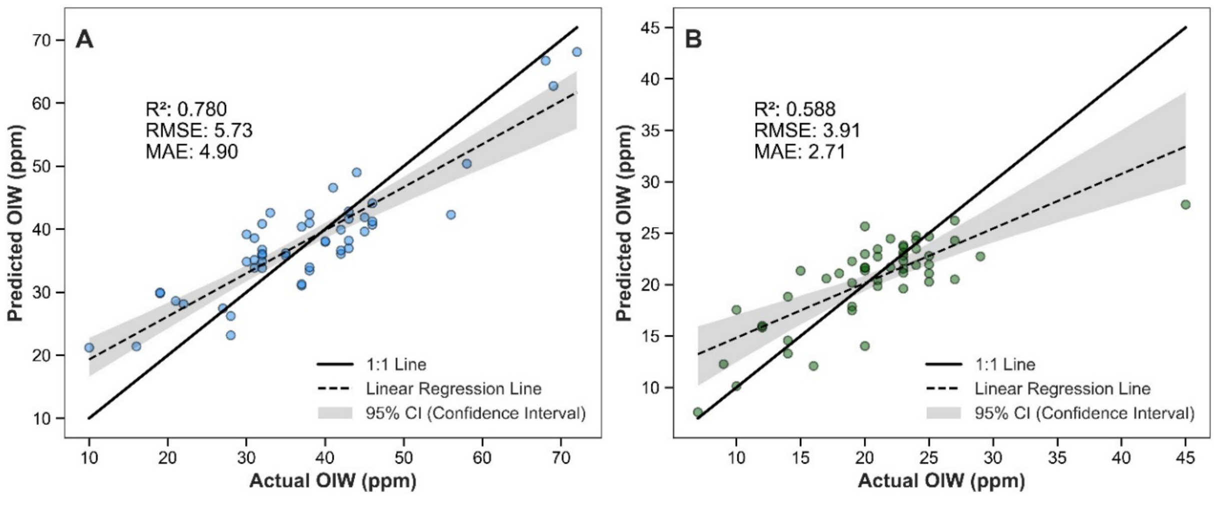


***Figure 8.2**: Optimised XGBoost model predicted vs actual oil-in-water outlet concentrations*

Following the pointwise performance assessment using scatter plots, Fig. 8.3 presents a Taylor diagram that offers a holistic evaluation of the statistical performance of the XGBoost models for both treatment units. The diagram simultaneously integrates three key statistical indicators: the correlation coefficient, standard deviation and centered root mean square error (RMSE).

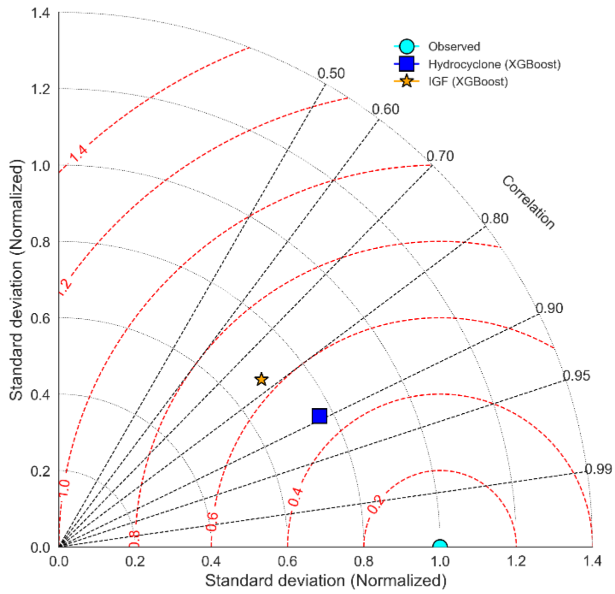


Figure 8.3: *Taylor diagram comparing Hydrocyclone and IGF optimised XGBoost model performances against observed oil-in-water concentrations*

In Fig. 8.3, the observed point is represented by the circle at a normalised standard deviation of 1 on the x-axis, serving as the benchmark for model evaluation. The Hydrocyclone model, demonstrates a higher correlation with the observed data (approximately 0.95) and a normalised standard deviation close to 1, indicating its superior ability to capture the variability present in the measurements. The IGF model, depicted by the star, displays a slightly lower correlation coefficient (approximately 0.90) and a marginally reduced normalised standard deviation, suggesting a minor underestimation of the observed variability.

The relative proximity of both models to the reference point underscores good predictive accuracy of the XGBoost approach for both separation units. However, the Hydrocyclone model shows slightly superior agreement, reflecting greater consistency and stronger predictive capacity in representing oil-in-water concentration dynamics.

8.5. Machine Learning Aided Monte Carlo Simulation (ML-MCS)

To accurately capture the variability and uncertainty inherent in key operational parameters of the produced water Hydrocyclone and IGF deoiling system, a detailed statistical distribution fitting exercise was performed. The features mentioned in Table 7.1 (Chapter 7) spanning process inputs and outputs from the Hydrocyclone and IGF units were individually assessed to identify the probability distribution that best characterises their empirical behavior. This process involved fitting each variable to a wide range of candidate distributions and selecting the optimal model based on goodness-of-fit statistics criteria as presented in Fig. 8.4.

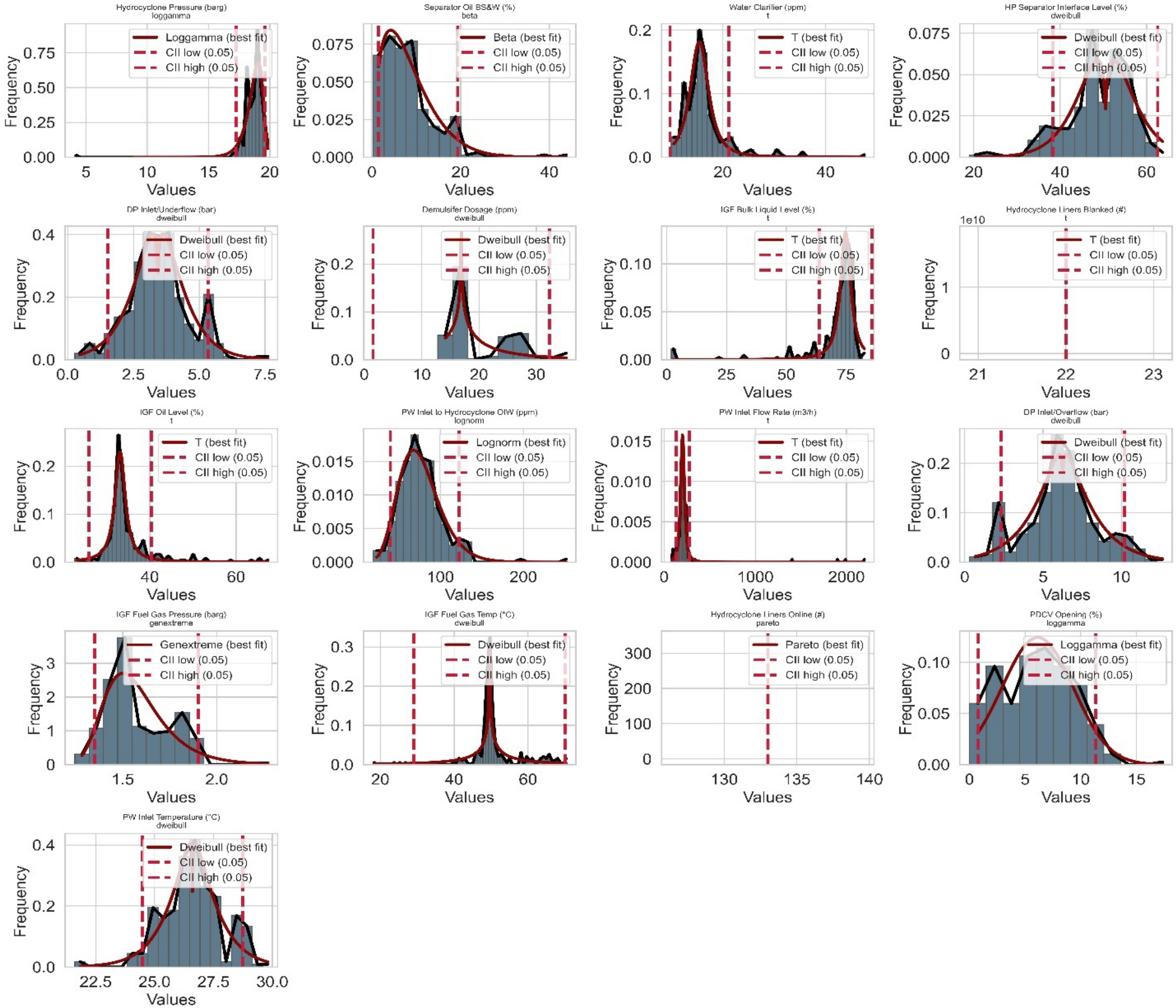


***Figure 8.4:** Best-fit probability distributions for the used operational features with Probability Density Functions and 5–95% confidence intervals used in the Monte Carlo Simulations*

Fig. 8.4 displays the empirical distributions of key operational variables used as input features in the XGBoost models for the Hydrocyclone and Induced Gas Flotation (IGF) units. Each subplot illustrates the histogram of observed data overlaid with the best-fitting probability distribution (solid black line) and the corresponding 5% confidence interval (CII) bounds (dashed red lines). The fitted distributions, ranging from lognormal, T, beta, gamma, to Pareto and Dewbil, capture the inherent variability and skewness present in each feature, providing a robust basis for generating realistic synthetic data during Monte Carlo simulations.

8.6. Distributional Fidelity of ML-MCS Predictions

To evaluate the representational accuracy of the ML-MCS, a comparative distributional analysis was performed between actual and simulated oil-in-water concentrations at three critical stages of the produced water deoiling process. Fig. 8.5 presents kernel density estimation (KDE) plots of the actual data (Fig. 8.5.A) and ML-simulated data (Fig. 8.5.B), along with annotated summary statistics for each distribution.

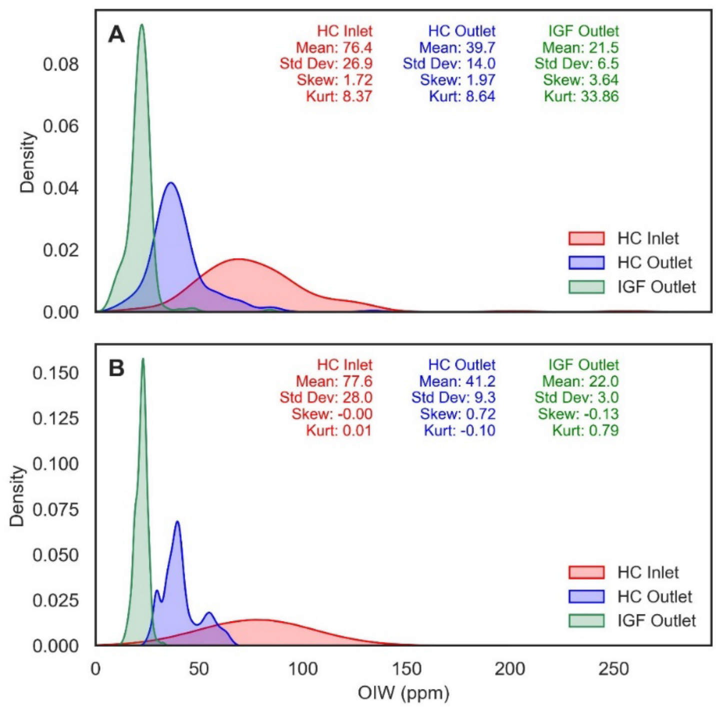


***Figure 8.5**: Comparative kernel density estimates of oil-in-water outlet concentrations between (A) actual and (B) ML-MCS*

The ML models accurately replicated average behavior across the three critical stages. At the Hydrocyclone inlet, the simulated oil-in-water mean (77.7 ppm) and standard deviation (28.0 ppm) closely match the actual values (76.4 ppm and 26.9 ppm respectively). Similar alignment is observed at the hydrocyclone outlet oil-in-water concentration mean (41.2 ppm for actual versus 39.7 ppm for simulated) and IGF outlet oil-in-water concentration mean (actual 22.0 ppm versus 21.5 ppm for simulated), confirming effective generalisation of the operational dynamics and stage-wise oil-in-water concentration reduction.

However, analysis of higher-order moments revealed a consistent smoothing effect in the simulated distributions. Actual data from the Hydrocyclone and IGF outlets exhibited pronounced right skewness (actual 1.97 versus simulated 3.64) and high kurtosis (actual 8.64 versus simulated 33.86), indicating heavy tails and sharp peaks associated with sporadic process excursions. In contrast, the ML-generated distributions are more symmetric and mesokurtic, with near-zero skewness (−0.11) and reduced kurtosis (1.05) at the IGF outlet, suggesting underrepresentation of extreme events.

The smoothing observed is characteristic of supervised learning, where models tend to regularise outliers, especially in sparse or noisy regions of the training data. While this enhances predictive stability, it can limit sensitivity to rare but operationally significant anomalies such as flow surges, chemical overdosing, or equipment failure.

These findings are reinforced by the quantile–quantile (QQ) plots with inset boxplots in Fig. 8.6. Across all stages, the QQ plots show strong alignment with the 1:1 reference line, indicating reliable shape reproduction across most quantiles. Notable deviations were observed at the hydrocyclone outlet tail in Fig. 8.6B, where extreme values are underpredicted. The inset boxplots in Fig. 8.6B corroborates this pattern, showing well-matched medians but narrower interquartile ranges in the simulated outputs, especially noticeable in the downstream stage as represented in Fig. 8.6C.

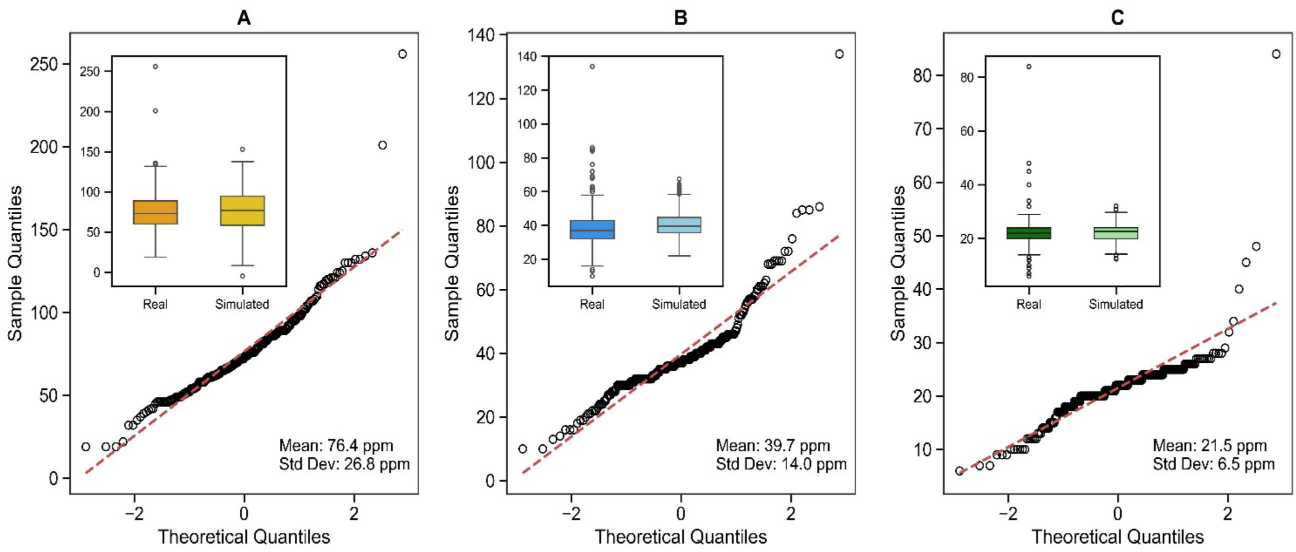


***Figure 8.6**: Quantile–Quantile (QQ) plots with inset boxplots comparing actual and simulated oil-in-water concentrations at the Hydrocyclone Inlet (A), Hydrocyclone Outlet (B) and IGF Outlet (C)*

This comparison highlights a central trade-off in data-driven simulation: while the ML models successfully generalise dominant system behaviors, such as the progressive oil-in-water reduction from the hydrocyclone inlet to the IGF outlet, the models also smooth distributional features, reducing skewness and suppressing extremes. This smoothing is characteristic of supervised learning, where model regularisation and limited data density in outlier regions lead to underrepresentation of rare but critical events.

8.7. Model Explainability and Feature Interaction Effects in ML-MCS Oil-In-Water Prediction

Fig. 8.7 and Fig. 8.8 present a SHAP-based interpretation of how different features influence the oil-in-water predictions at both the Hydrocyclone and IGF outlet stages. The summary plots (Fig. 8.7A and 8.7B) reveal the global features of importance across the entire dataset, while the waterfall plots (Fig. 8.8A and 8.8B) decompose the contributions of each individual feature for a single random data point in the simulation.

For the Hydrocyclone outlet (Fig. 8.7A), *PW Inlet to Hydrocyclone OIW (ppm)* was identified as the most impactful feature, with higher values consistently increasing predicted outlet concentrations, consistent with known performance degradation under elevated inlet loads as outlined in Chapter 7, Section 7.6.1.1. *Operating Time (Days)* and *PW Inlet Temperature (°C)* also increase oil-in-water concentration predictions, which could be attributed to equiment wear or thermal effects. Features such as *PW Inlet Flow Rate (m³/h)* exhibit broad SHAP distributions, indicating variable operational influence, whereas interface level and

Hydrocyclone liners show minimal impact. In the corresponding waterfall plot (Fig. 8.8A), the baseline prediction (~40 ppm) was raised significantly by *PW Inlet to Hydrocyclone OIW (ppm)* (+12.9 ppm) and reduced by *Operating Time (Days)*, *PW Inlet Flow Rate (m^3/h)* and *PW Inlet Temperature (°C)*, confirming localised compensatory effects.

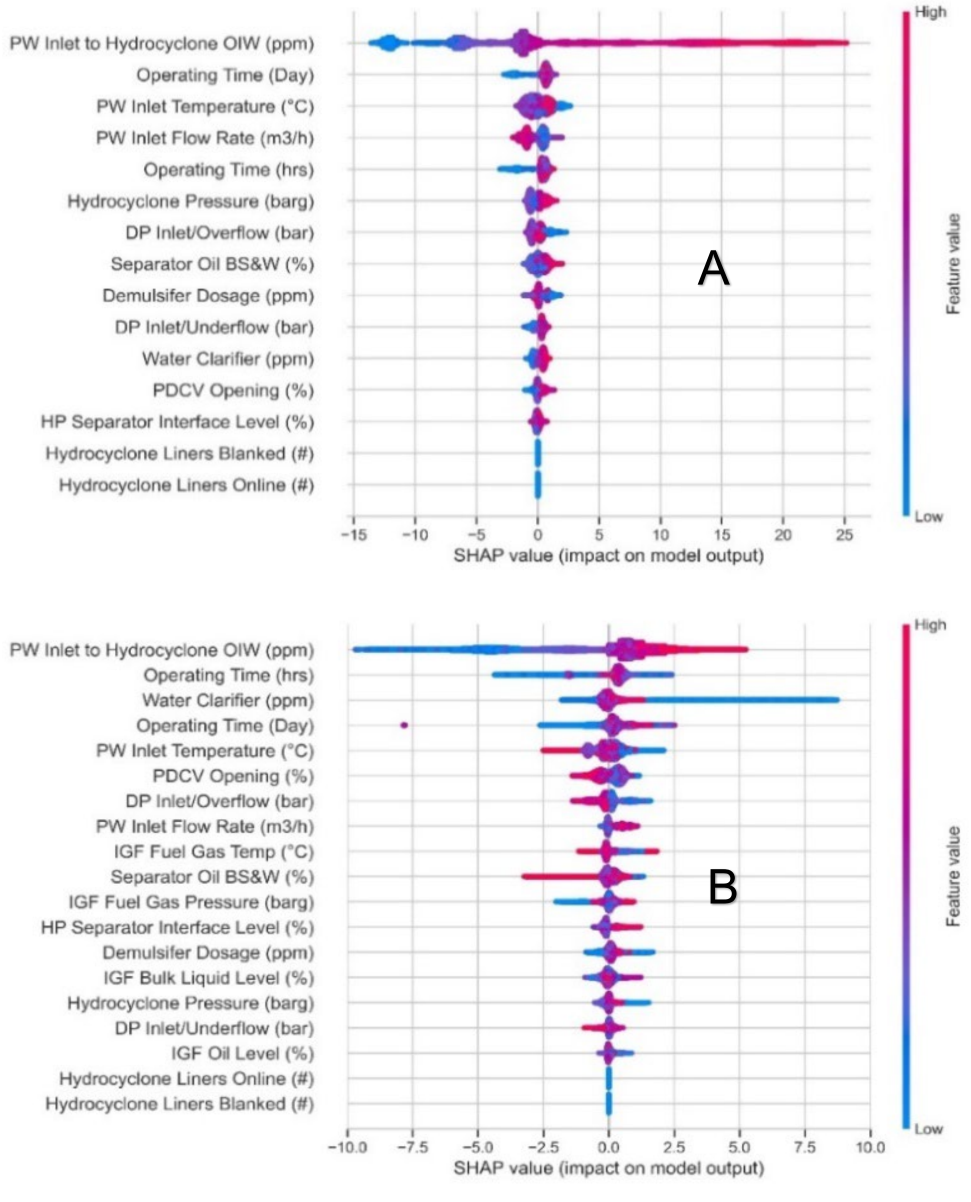


***Figure 8.7**: SHapley Additive exPlanations (SHAP) showing the impact of features on outlet oil-in-water concentration predictions across all data points for the Hydrocyclone (A) and IGF (B)*

At the IGF outlet (Fig. 8.7B), the SHAP summary plot identified *PW Inlet to Hydrocyclone OIW (ppm)* as the top driver as well, indicating upstream effects propagate through the deoiling units. *Operating Time (hrs)* and *Water Clarifier (ppm)* dosage follow as feature as important, where longer runtimes and lower dosing levels elevated the predicted oil-in-water concentrations, suggesting fouling or suboptimal chemical treatment. Other features, including

Dp Inlet/Overflow (bar) pressure, *PW Inlet Flow Rate (m³/h)*, and *IGF Oil Level (%)*, contributed moderately. The waterfall plot (Fig. 8.8A) shows how these factors influence the prediction, starting from a 21.69 ppm oil-in-water concentrations baseline, the forecast increased due to *PW Inlet to Hydrocyclone OIW (ppm), PW Inlet Flow Rate (m³/h)* and *IGF Oil Level (%)*,, but was offset by *DP Inlet/Overflow (bar)* pressure and *Water Clarifier (ppm)*.

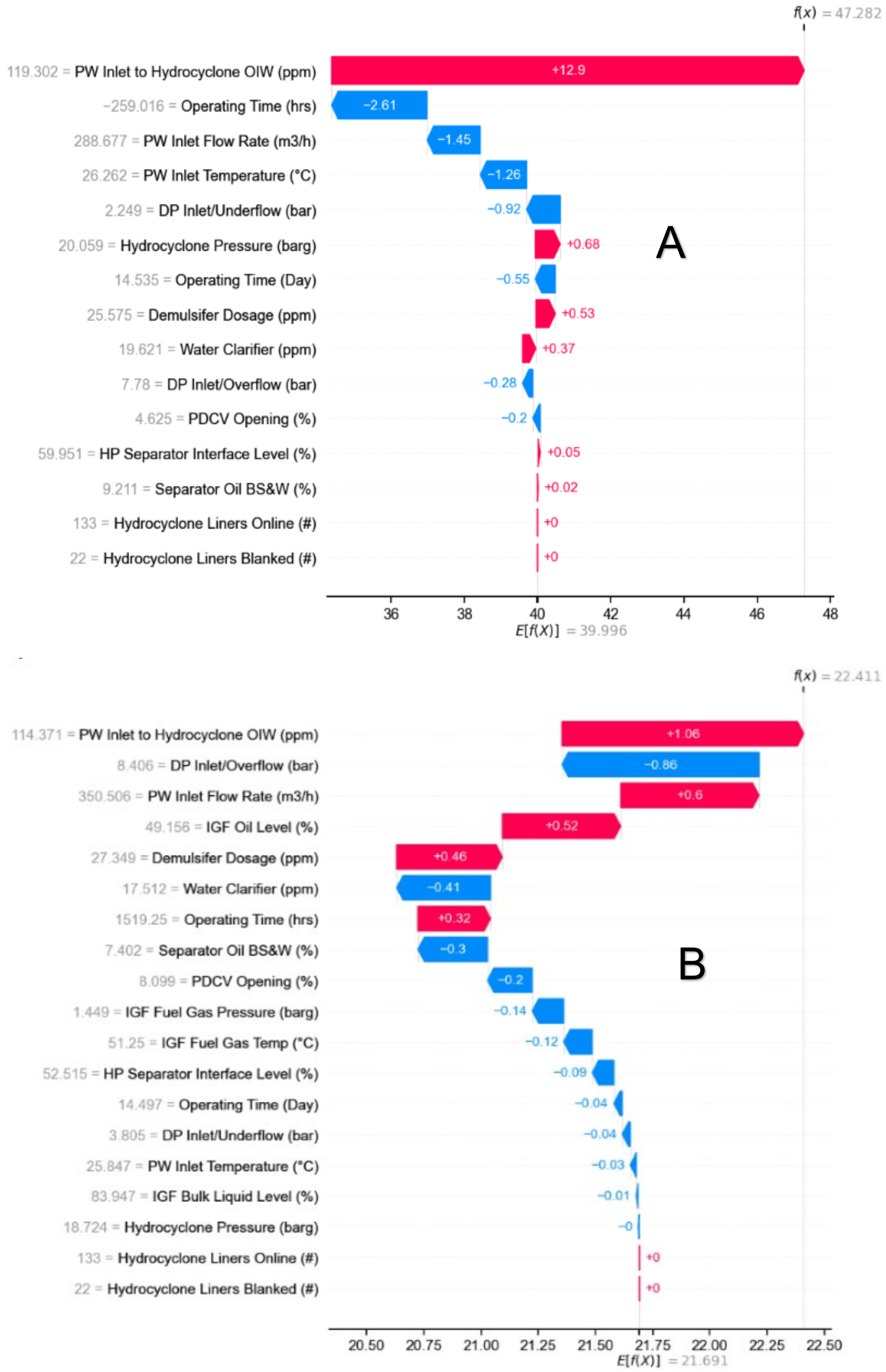


***Figure 8.8**: SHAP waterfall illustrating individual feature contributions for a random simulation of a single (A) Hydrocyclone outlet and (B) IGF outlet prediction*

8.8. Predictive Accuracy and Performance Mapping

Building upon the prior analysis of distributional fidelity (Fig. 8.5), a second layer of evaluation was conducted to assess the predictive representativeness of the ML-MCS outputs against actual inlet concentrations. Fig. 8.9 presents two-dimensional kernel density estimation (2D-KDE) plots comparing actual inlet oil-in-water concentration (x-axis) with corresponding simulated outlet oil-in-water concentration (y-axis) across both treatment stages. Actual measured outlet values were overlaid as scatter points, providing a concrete benchmark for visual and statistical interpretation. Marginal histograms further contextualised differences in distribution shape, including skewness, spread and modal behaviour.

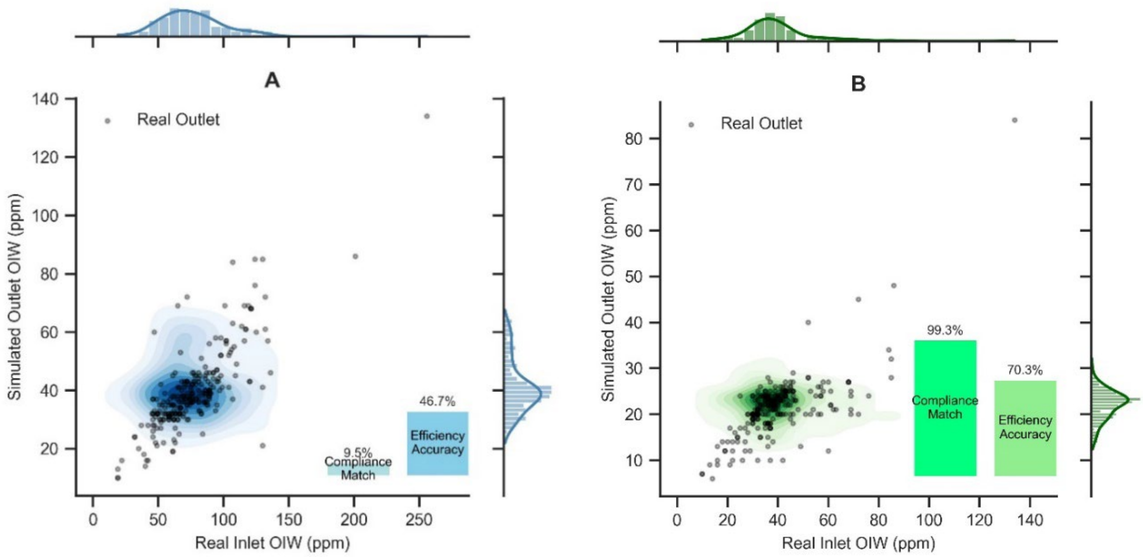


Figure 8.9: *Actual inlet versus simulated outlet oil-in-water for (A) Hydrocyclone and (B) IGF including actual outlet points and inset performance metrics*

At the Hydrocyclone stage (Fig. 8.9A), predictions are centered between 80–120 ppm (inlet) and 20–50 ppm (outlet). However, many actual outlet values linked to high inlet loads fall outside high-probability regions, indicating the model underpredicts during stress conditions. This is supported by a low Compliance Match (9.0%) and modest Efficiency Accuracy (46.7%), suggesting that the model struggles with upstream variability and extreme scenarios. These limitations point to the need for enhanced robustness.

In contrast, the IGF simulation (Fig. 8.9B) shows strong correlation with actual data. The predicted oil-in-water outlet concentration clustered between 20–30 ppm, with most actual points falling within high-probability contours. This stability is reflected in a Compliance Match of 93.3% and Efficiency Accuracy of 70.3%, confirming the model's effectiveness at the downstream stage, likely due to reduced inlet variability post-Hydrocyclone deoiling treatment.

Fig. 8.10 extends this analysis through Cumulative Distribution Function (CDF) and Reliability Function (1−CDF) curves. At the Hydrocyclone outlet (Fig. 8.10A and Fig. 8.10C), simulated and actual oil-in-water distributions show slight underestimation near 45–60 ppm. Exceedance probabilities above 30 ppm remain high (~65–70%) in both datasets, reaffirming the system's sensitivity and compliance risk under fluctuating loads. At the IGF outlet (Fig. 8.10B and Fig. 8.10D), the simulation closely matches the actual CDF across the entire range. The exceedance probability drops below 5%, with the simulated curve tightly tracking the empirical one, demonstrating strong predictive reliability for downstream compliance forecasting.

Overall, these results highlight a clear trend: the ML-MCS framework performs best under stable oil-in-water inlet conditions, as seen in the IGF stage (Fig. 8.10B and 8.10D). Predictive accuracy declined in high-variability regions, such as the Hydrocyclone treatment stage, where dynamic interactions and occasional outliers challenged the model.

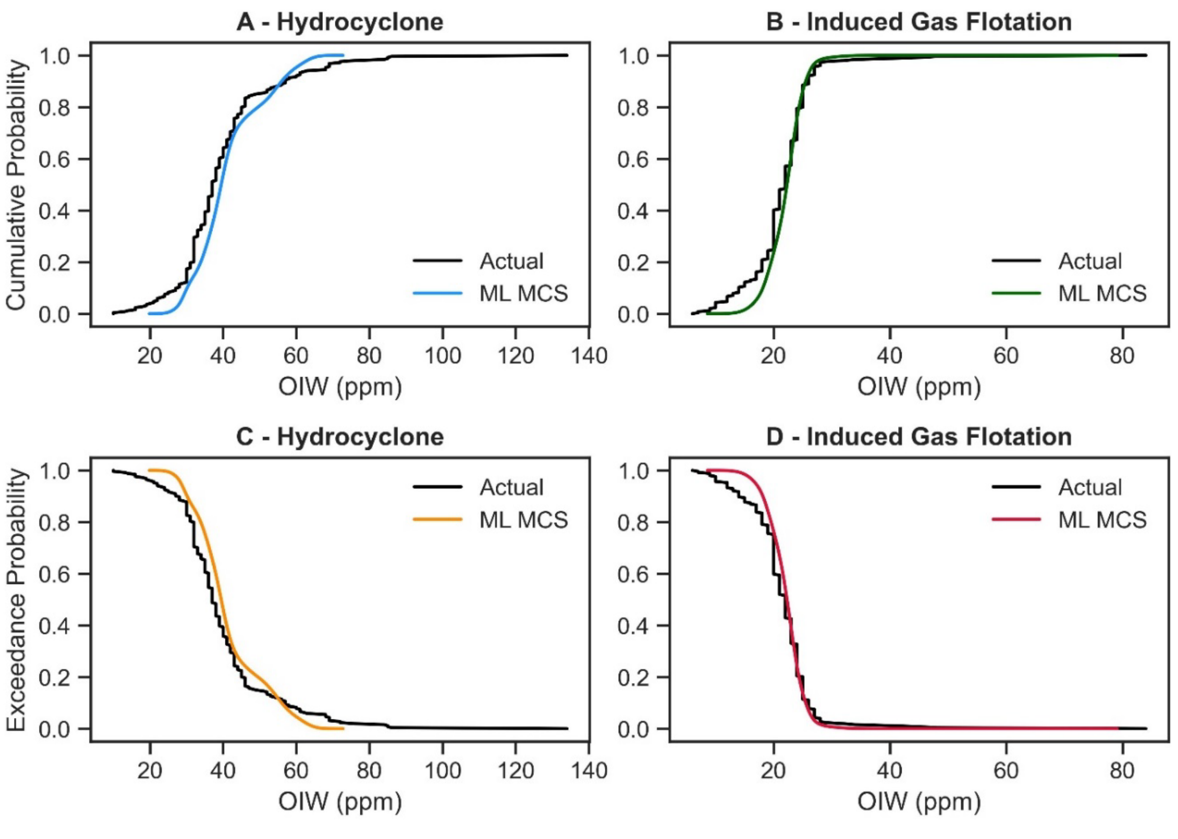


***Figure 8.10**: Actual and simulated oil-in-water distributions at the Hydrocyclone (A, C) and IGF (B, D) outlets using cumulative distribution (top) and reliability curves (bottom)*

Table 8.1 presents goodness-of-fit metrics evaluating how well the ML-MCS replicated actual oil-in-water concentrations at the Hydrocyclone and IGF outlets. The mean differences are 1.59 ppm oil-in-water concentration for the Hydrocyclone and 0.56 ppm for the IGF. Standard deviation differences (4.70 ppm and 3.50 ppm, respectively) suggest mild underestimation of variability, particularly upstream. The KS statistics, 0.16 for the Hydrocyclone and 0.17 for IGF, reflect moderate cumulative distribution alignment, while the Wasserstein distances (3.52 ppm and 1.64 ppm) further confirm a closer fit at the IGF outlet. Overall, the ML-MCS demonstrated good predictive fidelity, with a stronger performance observed downstream where inlet conditions are more stable.

Table 8.1: *Goodness-of-fit statistics for Monte Carlo-simulated oil-in-water outlet concentrations*

	Hydrocyclone Outlet	**IGF Outlet**
Mean Difference (ppm)	1.59	0.56
Std. Dev. Difference	4.70	3.50
KS Statistic	0.17	0.17
Wasserstein Distance	3.74	1.63

8.9. Spatiotemporal Compliance Landscape

A high-level depiction of how compliance probabilities evolve over time and across different discharge thresholds is illustrated in Fig. 8.11. This three-dimensional surface plot, based on 1,000 MCS over a 1,200-hour period, shows how often the simulated oil-in-water concentration fell below the 30 ppm threshold at each timestep across both treatment stages.

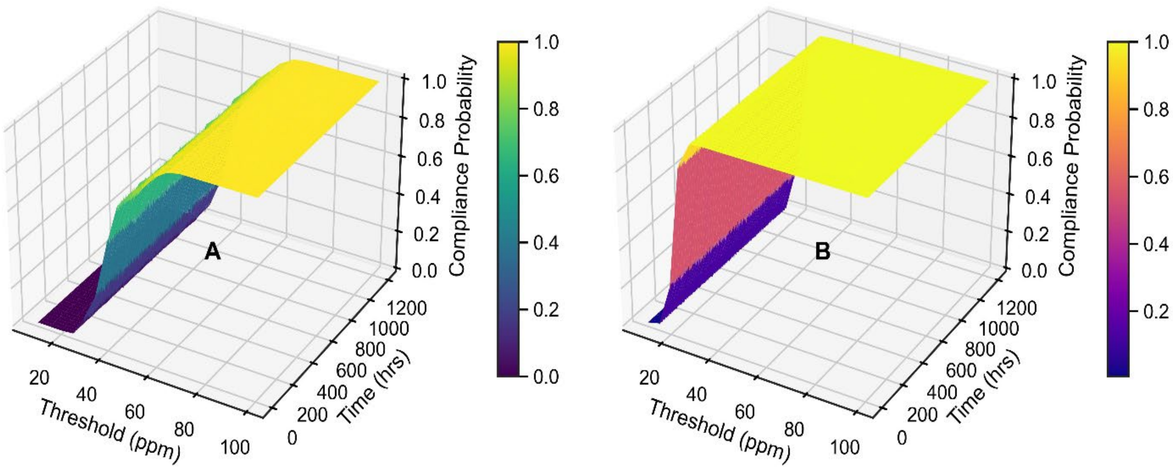


Figure 8.11*: Time- and threshold-dependent compliance probability surfaces for (A) Hydrocyclone and (B) IGF outlets*

For the Hydrocyclone outlet, the compliance landscape exhibited a gradual gradient: probabilities remained relatively low around stricter thresholds (e.g., 30-40 ppm), especially during early operation, but increased steadily with more lenient threshold standards (Fig. 2.8, Chapter 2). At thresholds above 50 ppm, compliance exceeded 90% across nearly all timesteps, while variability became more pronounced near the 40 ppm region, reflecting the Hydrocyclone's sensitivity to upstream fluctuations. In contrast, the IGF outlet demonstrated a steeper and more favorable compliance slope. Thresholds as low as 25 ppm yield over 95% compliance for the majority of the simulation period, indicating strong polishing efficiency. The sharp transition observed around the 30 ppm threshold, highlighted the IGF's critical role in securing regulatory performance, even when inlet conditions vary.

8.10. Temporal Evolution of Compliance and Exceedance Probabilities

Fig. 8.12 presents time series plots of compliance and exceedance probabilities at the Hydrocyclone and IGF outlets over the 1,200-hour simulated operational period, based on 1,000 runs of the ML-MCS framework. These plots capture how consistently each unit maintained discharge performance under stochastic conditions and highlighted the temporal dynamics of regulatory risk exposure.

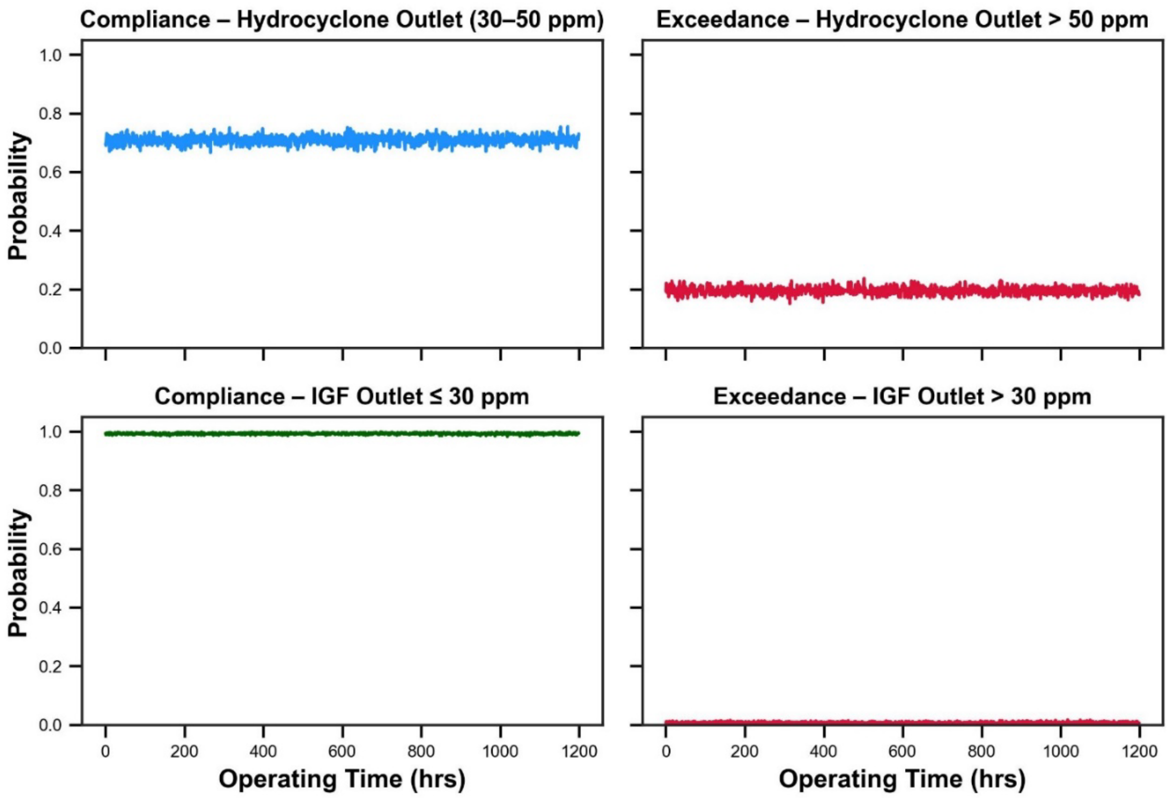


***Figure 8.12**: Compliance and exceedance probabilities over time for Hydrocyclone and IGF oil-in-water concentration outlets*

The Hydrocyclone oil-in-water outlet concentration maintained a relatively high and stable compliance probability within the 30–50 ppm band, generally oscillating between 70–80% across the simulation window. Meanwhile, exceedance probabilities above 50 ppm remained modest but persistent, hovering around 10–15%, likely reflecting sensitivity to short-term fluctuations in the inlet quality or process load. These excursions suggest that while overall reliability is strong, the Hydrocyclone remains vulnerable to high-load disturbances.

The IGF oil-in-water outlet concentration exhibited consistently high compliance with the ≤30 ppm threshold, exceeding 95% compliance across the entire timeline. Exceedance probabilities were negligible, reinforcing the unit’s role as a stable and effective polishing stage. This consistent behavior suggests that the ML-MCS model has accurately captured the IGF’s operational resilience and low sensitivity to upstream variability.

Importantly, the flat, time-invariant profiles observed across all the curves (Figure 8.12), indicate stable performance throughout the simulation horizon, with no signs of temporal degradation or drift. This stability supports the ML-MCS framework’s applicability for long-term compliance forecasting and underscores its value in predictive monitoring, regulatory reporting and proactive control optimisation under uncertainty.

8.11. Threshold-Based Compliance Assessment

A threshold-focused analysis, as shown in Figure 8.13, evaluated the performance of the Hydrocyclone and IGF oil-in-water outlet concentrations against varying regulatory discharge limits (as outlined in Fig. 2.8, Chapter 2). Derived from 1,000 ML-MCS, this analysis quantifies the probability of meeting or exceeding selected thresholds, providing insight into operational robustness and sensitivity to compliance standards.

For the Hydrocyclone outlet, compliance with a strict 30 ppm oil-in-water concentration limit was achieved in only 9.4% of cases, highlighting a high sensitivity to regulatory tightening and limited effectiveness under stringent discharge conditions. However, performance improved sharply for the more relaxed thresholds, rising to 54.5% at 40 ppm, 82.8% at 50 ppm and 95.9% at 60 ppm. This steep compliance gradient indicates that the Hydrocyclone’s effectiveness is highly dependent on inflow characteristics and the threshold limit. The corresponding exceedance probabilities mirror this trend: while exceedance at 30 ppm is nearly universal (~99.6%), it drops significantly to 17.2% at 50 ppm and just 4.1% at 60 ppm.

These results suggest that even small regulatory shifts can result in significant compliance changes, emphasizing the need for dynamic process control or integration with robust downstream treatment. In contrast, the IGF outlet displayed exceptional stability across all the oil-in-water threshold limits. At the 30 ppm benchmark, compliance reached 99.3%, with complete compliance (100%) observed at 40 ppm and above. Exceedance probabilities were correspondingly low, with 0.7% at 30 ppm and negligible beyond. This confirms the IGF unit's role as a highly reliable polishing stage that consistently buffers upstream variability and ensures regulatory adherence even under conservative discharge limits.

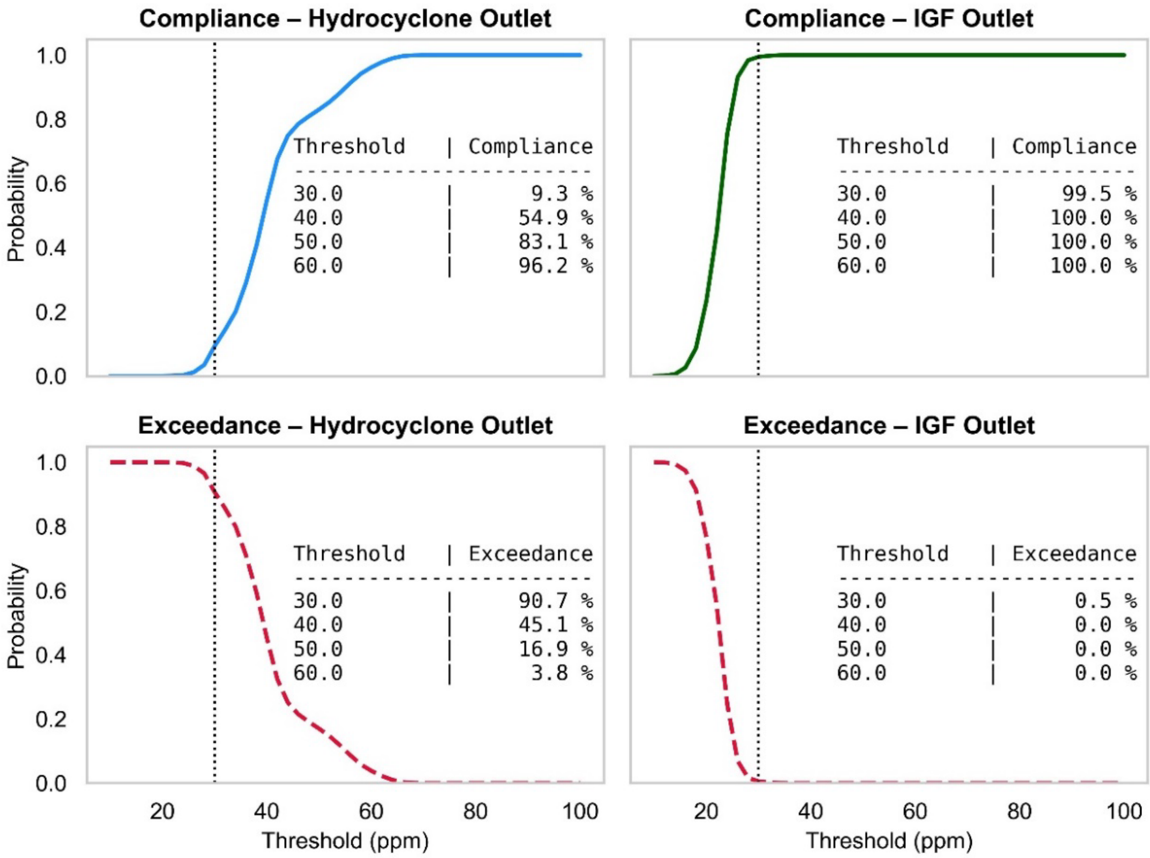


***Figure 8.13**: Compliance and exceedance probabilities for the different threshold limits*

8.12. Simulation-Based Exceedance Mapping

Fig. 8.14 visualises oil-in-water concentration exceedance patterns at the Hydrocyclone and IGF outlets over a 1,200-hour simulated window and 1,000 Monte Carlo runs. Each column represents a simulation and each row a timestep; red pixels mark threshold exceedances.

At the Hydrocyclone outlet (oil-in-water concentration >30 ppm), exceedances were frequent and randomly distributed across time and simulations, indicating persistent marginal non-compliance under a wide range of conditions (Fig. 8.14A). The lack of temporal or simulation clustering suggests broad sensitivity to inlet variability and internal dynamics, highlighting the need for fine-tuned, real-time control of parameters like differential pressure and flow (Bram *et al.*, 2020; Bram *et al.*, 2018). In contrast, the IGF outlet (oil-in-water concentration >30 ppm) showed rare and isolated exceedance events (Fig. 8.14B) as the sparse red pixels observed reflect effective mitigation of upstream fluctuations. Most simulations maintained full compliance, confirming the IGF's reliability as a downstream polishing stage.

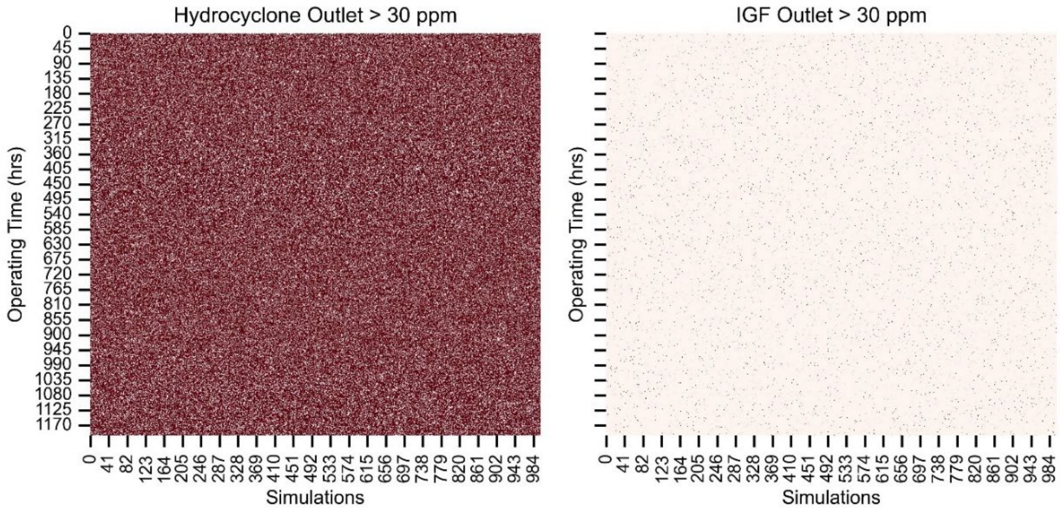


Figure 8.14: *Simulated exceedance heatmaps at Hydrocyclone and IGF (both >30 ppm oil-in-water outlet concentrations) over 1,000 Monte Carlo runs*

Fig. 8.14A and Fig. 8.14B reinforce the role of ML-MCS in exposing not only the magnitude but also the timing and recurrence of risk. This insight can inform dynamic mitigation strategies, such as targeted chemical dosing or adaptive upstream control based on when and where exceedances are most likely to occur.

8.13. Conclusion

This Chapter presented a stochastic modeling framework based on Monte Carlo simulation to assess compliance risk and performance variability in offshore produced water deoiling systems. By integrating actual operational data with probabilistically generated inputs, the framework enabled realistic simulation of oil-in-water discharge behavior under uncertain operating conditions. Focusing on Hydrocyclone and induced gas flotation units, the approach supported distributional, temporal and threshold-based analyses, offering detailed insights into treatment reliability and regulatory performance. Results demonstrate the model's ability to replicate the statistical characteristics of actual data including distribution shape, central tendency and tail behavior. Notably, the IGF unit exhibited high compliance reliability under variable inlet conditions, while the Hydrocyclone showed greater sensitivity to upstream fluctuations.

This probabilistic framework advances beyond deterministic models by enabling performance assessment under uncertainty and supporting proactive decision-making through the identification of compliance risk zones, quantification of exceedance likelihoods and diagnosis of stage-specific vulnerabilities. Its structure lends itself to integration into real-time monitoring systems and digital twins, enabling adaptive process control and continuous compliance tracking.

Beyond technical evaluation, the framework offers broader value for environmental governance by generating objective compliance metrics. Its statistically rich outputs such as exceedance heatmaps and compliance surfaces also provide robust documentation for regulatory audits, reinforcing transparency and operational accountability.

CHAPTER 9

CONCEPTUAL FOUNDATION FOR DIGITAL TWIN ARCHITECTURE, OPTIMISATION AND CONTROL INTEGRATION FOR OFFSHORE PRODUCED WATER DEOILING

"I am the way and the truth and the life."

John 14:6

CHAPTER 9: CONCEPTUAL FOUNDATION FOR DIGITAL TWIN ARCHITECTURE, OPTIMISATION AND CONTROL INTEGRATION FOR OFFSHORE PRODUCED WATER DEOILING

9.1. Introduction

In offshore oil and gas operations, produced water treatment systems must consistently achieve oil-in-water concentrations below regulatory thresholds, typically 30 ppm for Angolan waters, despite challenges such as fluctuating inlet conditions, aging infrastructure, and variable chemical performance. To address these demands, this chapter proposes a digital twin: a dynamic, data-driven replica of the produced water deoiling process designed to predict, assess, and optimise system behavior in real time.

Unlike traditional simulation tools, the digital twin integrates Machine Learning and Monte Carlo Simulation to deliver both deterministic forecasts and probabilistic risk assessments. By continuously ingesting operational data and emulating system dynamics, it acts as both a soft sensor and a control strategist, providing operators with foresight, compliance assurance, and actionable decision support.

Beyond monitoring, the framework incorporates real-time optimisation and actuation logic, allowing it not only to simulate what might happen but also to recommend what should be done, marking a shift from passive forecasting to adaptive, intelligent control.

This chapter proposes a digital twin for offshore produced water deoiling that integrates Machine Learning and Monte Carlo simulation to enable real-time prediction, compliance risk assessment and adaptive control. The objectives are to outline the architecture, demonstrate predictive and prescriptive capabilities and showcase integration with real-time optimisation and control strategies.

9.2. Rationale for Methodology

The architectural choices in this digital twin were driven by the need for accuracy, interpretability, adaptability, and computational efficiency. The core methodology integrates four synergistic components:

1. XGBoost Algorithm: Selected for its ability to handle nonlinear, multivariate tabular data, XGBoost delivers high predictive accuracy with built-in regularisation. Its

structure supports interpretable outputs via tree-based analysis and SHAP values, making it ideal for sensitive regulatory applications.

2. Monte Carlo Simulation: Rather than relying on fixed-point forecasts, MCS introduces variability by sampling thousands of plausible input scenarios from statistical distributions. This enables quantification of exceedance probabilities, offering probabilistic guarantees of compliance rather than binary predictions.
3. Limited-memory Broyden–Fletcher–Goldfarb–Shanno with Bound constraints **(**L-BFGS-B) Optimisation**:** A quasi-Newton method enabling the model to determine optimal operating conditions while strictly respecting historical and physical boundaries. Its low memory footprint makes it well-suited for embedded or edge applications, particularly in SCADA systems with limited computational resources (Qi, *et al.*, 2017).
4. SHapley Additive exPlanations (SHAP): Used to evaluate the impact of each feature on a given prediction. This ensures model transparency, supports root-cause diagnostics and enables identification of leverage points for optimisation (Mosca *et al.*, 2022).

9.2.1. Architecture of the Digital Twin

The digital twin framework is designed as a modular, multi-layered pipeline consisting of five interconnected modules:

1. Input Simulation Module: Actual and historical operational data are used to fit statistical distributions for key features as shown in Table 7.1. Simulated input vectors capture operational variability and noise.
2. Predictive Engine: Two XGBoost models independently trained and tuned for the Hydrocyclone and IGF units predict oil-in-water outlet concentrations, serving as fast, real-time surrogates for complex physical processes.
3. Risk and Compliance Layer: Monte Carlo-generated outputs are compared to regulatory thresholds to generate Cumulative Distribution Functions (CDFs) and reliability curves, quantifying compliance probabilities and risk of exceedance across different scenarios.
4. Sensitivity and Explainability Layer: SHAP values and Partial Dependence Plots explain model predictions, highlight dominant features, and support fault detection and

dosage optimisation, transparently building operator trust and facilitating regulatory auditing.

5. Deployment and Interface Layer: The twin can be deployed as a web-based interface (enabling real-time data exchange with SCADA or dashboard systems), an interactive dashboard, or as an edge computing module. It supports integration with industrial protocols such as OPC-UA, SCADA, and DCS.

9.3. Advanced Diagnostics and Optimisation Results

9.3.1. Uncertainty Visualisation with Confidence Bands

To assess prediction robustness under variable operational conditions, the digital twin framework incorporates both distributional and temporal uncertainty analyses. These visualisations provide critical insight into how predicted outlet oil-in-water concentrations fluctuate as probabilistic forecasts, rather than as static point estimates, due to upstream variability.

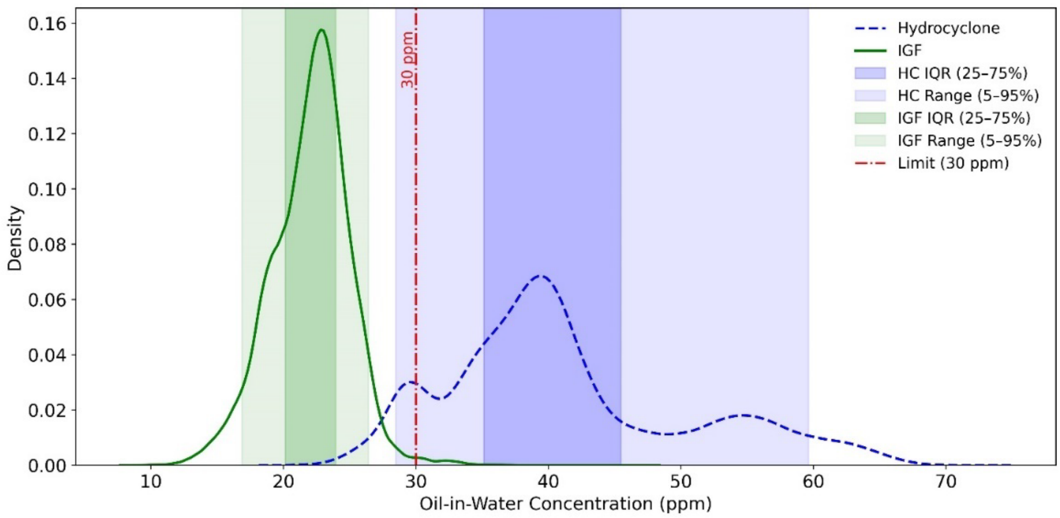


***Figure 9.1**: Predicted oil-in-water distributions for Hydrocyclone and IGF, with shaded ranges and 30 ppm threshold.*

Fig. 9.1 illustrates the predicted oil-in-water concentration distributions for the Hydrocyclone and IGF units, with confidence intervals reflecting simulation uncertainty. The Hydrocyclone distribution exhibits a right-skewed pattern, with a median concentration of approximately 39 ppm substantially above the regulatory threshold of 30 ppm. Its interquartile range (IQR) spans 35 to 44 ppm and the broader 5–95% confidence interval extends from 30 to 48 ppm, highlighting significant variability and frequent exceedance of compliance limits. By contrast, the IGF distribution is sharply peaked, with a median around 22 ppm and a narrow IQR of 20 to 24 ppm. Its 5–95% confidence interval (17–27 ppm) remains entirely below the discharge limit, indicating consistent and reliable performance.

These results underscore the importance of the two-stage treatment configuration (HIGF systems), where the IGF effectively mitigates upstream variability introduced by the Hydrocyclone.

Fig. 9.2 complement Fig. 9.1 by presenting rolling time-series simulations of predicted oil-in-water concentrations for the Hydrocyclone (Fig. 9.2A) and IGF unit (Fig. 9.2B), including interquartile range (IQR) bands to illustrate prediction variability.

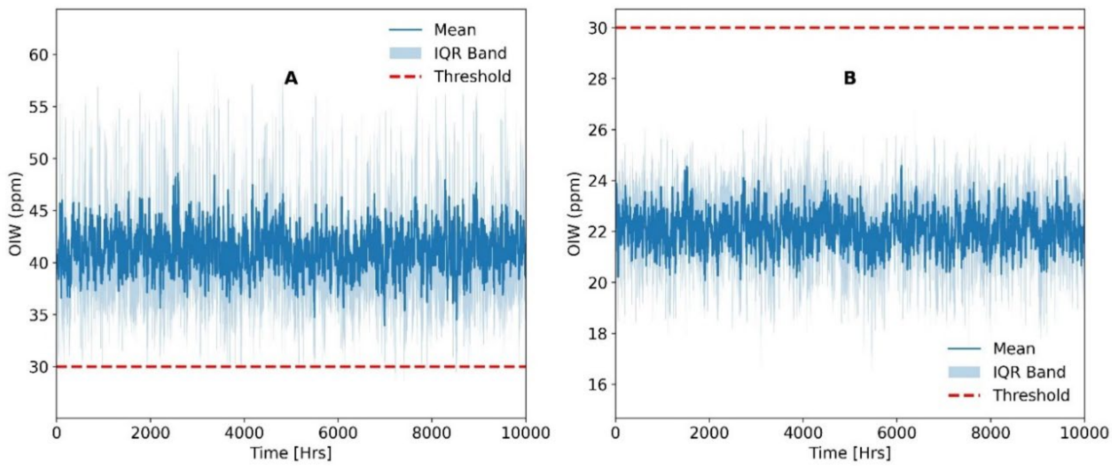


Figure 9.2: *Rolling oil-in-water concentration predictions: (A) Hydrocyclone with high variability; (B) IGF with stable, compliant performance.*

In Fig. 9.2A, the Hydrocyclone's mean predicted oil-in-water consistently exceeds the 30 ppm compliance threshold, with frequent peaks above 50 ppm. The wide IQR band underscores substantial variability and highlights the system's vulnerability to upstream disturbances, limiting its ability to maintain compliant discharge levels. Conversely, Fig. 9.2B shows the IGF with a narrow, stable oil-in-water distribution centered around a mean of ±22 ppm. The tight IQR band remains well below the regulatory limit throughout the simulation, reflecting the IGF's robust polishing capacity.

9.3.2. Optimisation for Minimum Oil-in-Water Concentrations Under Constraints

A prescriptive optimisation layer was introduced to transform the digital twin from a passive forecasting tool into an active decision-support system capable of recommending operational settings that minimise oil-in-water concentrations at each unit's outlet. The optimisation problem was formulated as a bounded minimisation task using the L-BFGS-B algorithm, with the objective function defined as the predicted oil-in-water output from each trained and optimised XGBoost model as outlined in Chapter 8, Section 8.4.1. Feature values were strictly constrained within their historical operational ranges to guarantee operational feasibility, and separate optimisation routines were performed for the Hydrocyclone and IGF units, each initialised using mean input values. To visually interpret these results, Fig. 9.3 presents the optimised input settings across key operational parameters for both units, which collectively define a calibrated operational envelope designed to achieve the lowest predicted oil-in-water concentrations within realistic constraints.

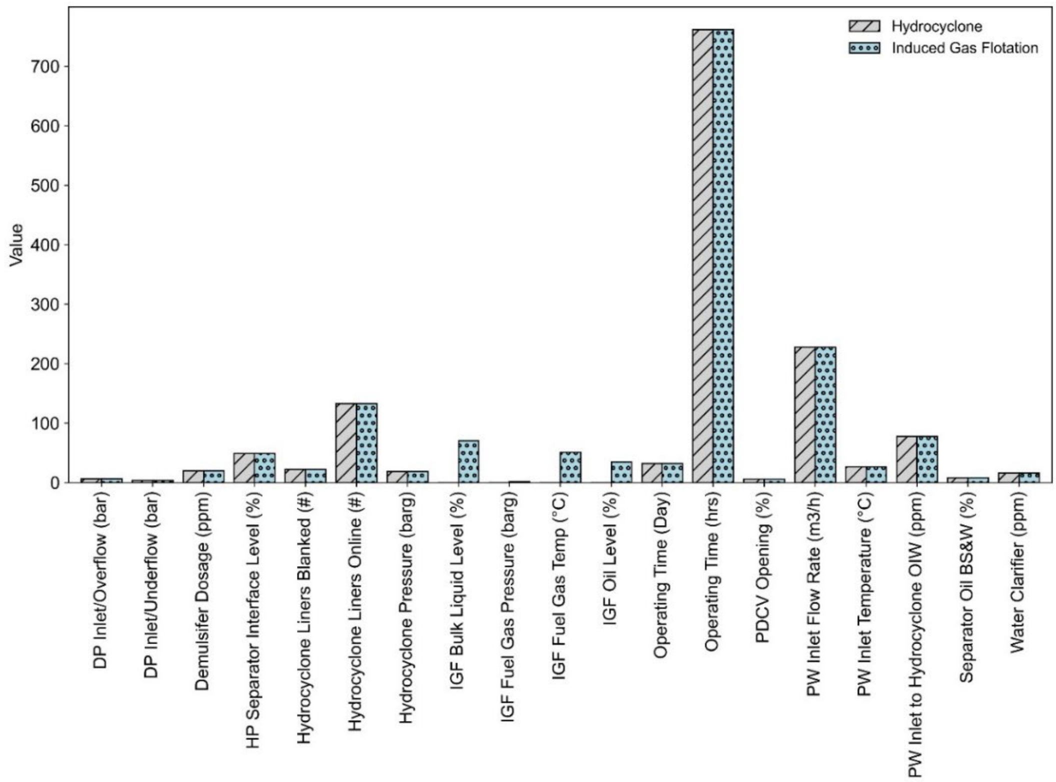


Figure 9.3: *Optimised control input settings for Hydrocyclone and IGF units across key operational parameters*

The optimised control input settings define a calibrated operational envelope aimed at achieving the lowest predicted oil-in-water concentrations while adhering to practical field constraints. The configuration prioritises extended operating time approximately 32 days (or 762 hours), to support high throughput and stable system performance. A moderate demulsifier dosage of around 20 ppm and an HP separator interface level of nearly 49% enhance upstream phase separation, delivering cleaner water to downstream polishing stages. The system operates at a produced water (PW) inlet flow rate of about 228 m³/h and a temperature of 27°C, balancing residence time and promoting effective droplet coalescence.

A high hydrocyclone pressure of 18.7 barg ensures strong centrifugal forces for initial oil removal. The use of 133 liners online, with 22 blanked, maximises capacity while maintaining operational flexibility. Moderate differential pressures across the inlet/overflow (6.2 bar) and inlet/underflow (3.5 bar) stabilise flow distribution and minimise upset risks. The PDCV opening, set at 6%, ensures controlled throttling and downstream pressure stability.

In the polishing stage, IGF parameters are fine-tuned with a fuel gas pressure of 1.6 barg, fuel gas temperature of 51°C, an oil layer level of 35%, and a bulk liquid level of nearly 71%, optimizing gas-liquid interactions for final oil removal. A water clarifier dosage of 16 ppm further enhances solids and oil droplet settling before discharge.

To interpret the implications of these optimised settings, Fig. 9.4 visualises the predicted minimum oil-in-water concentrations at the outlets of both treatment units relative to the 30 ppm regulatory threshold of Angola. The Hydrocyclone achieves a minimum predicted oil-in-water concentration of approximately 36.8 ppm, remaining above the compliance limit, whereas the IGF reaches around 22.4 ppm, well within the acceptable discharge range, reinforcing the Hydrocyclone's role as a preparatory separator and underscores the IGF's critical function as the final compliance gatekeeper.

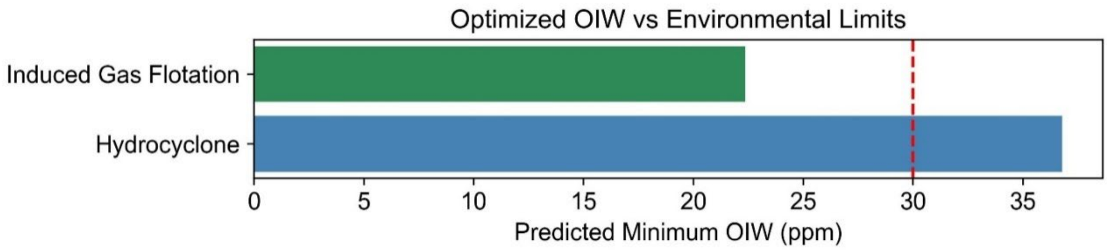


***Figure 9.4**: Predicted minimum oil-in-water concentrations after optimisation for Hydrocyclone and IGF units*

9.3.3. Validation via Precision Metrics and Correlation Coefficients

To evaluate the predictive performance of the digital twin models, both linear and rank-based correlation metrics, as well as a tight-error tolerance measure, were computed. The Pearson correlation coefficient assesses the strength and direction of a linear relationship between predicted and actual oil-in-water concentrations, with values close to 1 indicating strong linear agreement (Shaqiri *et al.*, 2023; Shiekh and El-Hashash, 2022). The Spearman rank correlation coefficient, on the other hand, evaluates whether the predicted and actual values maintain a consistent relative ordering, regardless of exact numerical distances, thus capturing monotonic relationships (Shaqiri *et al.*, 2023; Shiekh and El-Hashash, 2022).

Additionally, the percentage of predictions within ±10% of actual values provides a direct measure of local prediction precision, reflecting how closely the model estimates match real values at each point (Fig. 9.5).

For the Hydrocyclone, a high Pearson correlation of 0.89 and a Spearman coefficient of 0.81 indicate that the model effectively captures overall trends and maintains a generally correct ranking of OIW concentrations across varying operational states. However, only 37.25% of predictions fall within ±10% of actual values, suggesting significant local variability and highlighting challenges in capturing rapid or small-scale fluctuations in this more dynamic unit.

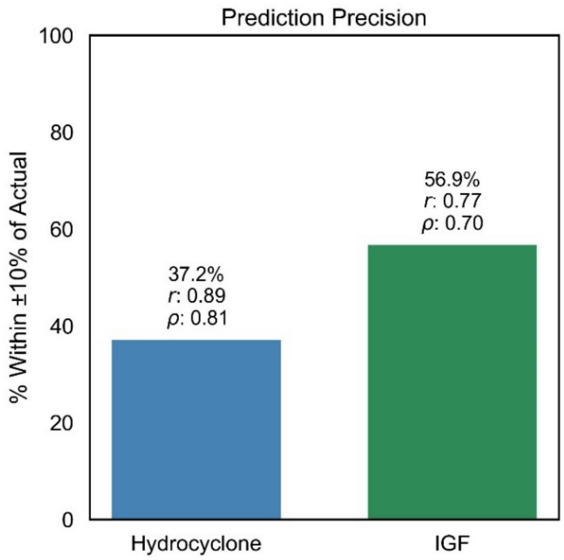


***Figure 9.5**: Prediction precision (% within ±10%), with correlation (r) and rank (ρ) for hydrocyclone and IGF models.*

The IGF model exhibits a moderate Pearson correlation of 0.77 and a Spearman coefficient of 0.70, indicating a slightly weaker ability to reproduce overall trends and ranks compared to the Hydrocyclone model. Nevertheless, 56.86% of IGF predictions lie within ±10% of the true values, demonstrating higher local accuracy and reflecting the IGF unit's more stable and predictable behavior.

Together, these results illustrate the complementary strengths of each model: the Hydrocyclone model is more robust in identifying general trends across wide operating conditions, while the IGF model excels at fine-scale prediction accuracy essential for meeting discharge compliance. These insights emphasise the importance of using multiple evaluation metrics to comprehensively understand model behavior and guide operational decision-making in offshore produced water treatment systems.

9.3.4. Stress Testing via Inlet Degradation Simulation

To evaluate the digital twin's adaptability under realistic operating disturbances, a series of inlet degradation scenarios were simulated. These scenarios emulate a common offshore challenge: increasing oil-in-water concentration at the inlet of the Hydrocyclone, caused by events such as reservoir breakthrough, separator malfunction or production ramp-up. In this test, the inlet oil-in-water concentration was incrementally increased from the mean of 76 ppm to 150 ppm, while all other input variables were re-optimised at each oil-in-water concentration increment using the L-BFGS-B algorithm. The goal was to assess whether the digital twin could maintain compliance under stress by dynamically identifying new optimal settings for each increment.

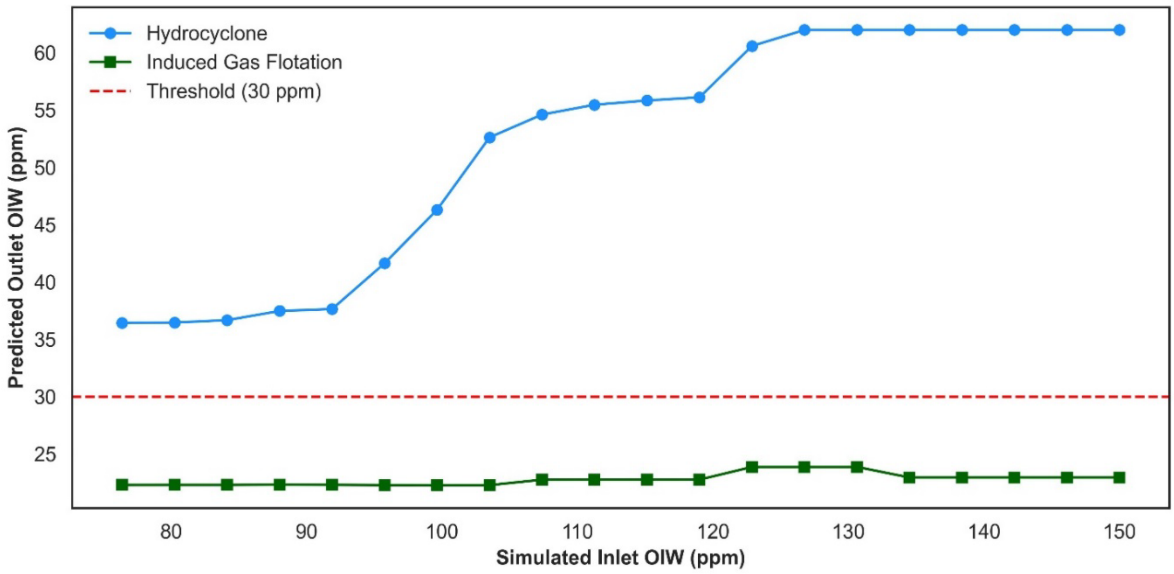


***Figure 9.6**: Predicted outlet oil-in-water concentrations from Hydrocyclone and IGF units under increasing simulated inlet oil-in-water loads*

Fig. 9.6 illustrates the dynamic response of the digital twin when subjected to a systematic increase in inlet oil-in-water concentrations, ranging from approximately 76 to 150 ppm. Two key trends emerge: the predicted outlet oil-in-water from the Hydrocyclone rises nonlinearly with increasing inlet values, showing a steep incline between ±95 and ±125 ppm before plateauing beyond 130 ppm, where the simulation indicates saturation. The IGF unit, however, maintains a relatively stable outlet oil-in-water prediction across the entire inlet range, consistently remaining below 25 ppm.

This differential behavior suggests that while the Hydrocyclone becomes less effective or reaches saturation limitations at higher inlet loads, the IGF unit demonstrates strong robustness. The inclusion of the Angolan regulatory compliance threshold at 30 ppm further emphasises these findings: the Hydrocyclone begins to exceed allowable discharge limits at inlet concentrations above ±85 ppm, whereas the IGF unit consistently stays within compliant bounds.

This analysis highlights the importance of inlet condition forecasting and adaptive control logic, as the upstream Hydrocyclone becomes increasingly vulnerable under heavy loading, necessitating downstream compensation or optimisation to ensure overall compliance. Furthermore, it demonstrates the digital twin's prescriptive agility, its capacity to track disturbances, recalculate optimal settings and maintain system performance within regulatory constraints. Notably, the simulation reveals a limitation for the Hydrocyclone above 130 ppm inlet oil-in-water concentration, a behavior not observed in the IGF unit within the simulated range presented in Fig. 9.5

9.3.5. Model Predictive Control (MPC)-Inspired Noise-Robust Optimisation Loop

While single-pass optimisation demonstrates the digital twin's prescriptive power under deterministic inputs, real-world process control requires adaptive, step-wise decision-making under uncertainty. To simulate this control-oriented behavior, a multi-step optimisation loop was implemented, mimicking the core logic of Model Predictive Control (MPC).

This design simulated a real-time closed-loop control scenario in which control recommendations were re-optimised at every decision step based on updated predictions. Each loop iteration mimics a rolling horizon of a Model Predictive Control scheme:

- Prediction and optimisation occurred at fixed intervals, simulating control cycles typical in offshore SCADA systems.
- Gaussian noise (2%, 5%, 10%) was injected into control inputs after each optimisation step, representing actuator errors, sensor drift or unmodeled process disturbances.
- After executing the perturbed control action, the digital twin predicted the new outlet concentrations, updated the system state and triggered the next optimisation step.

This iterative feedback loop enabled the twin to correct disturbances in near-real-time, continuously steering the process toward regulatory compliance, even in the presence of uncertainty. The loop emulated adaptive supervisory control rather than hard actuation, aligning with offshore operational safety protocols.

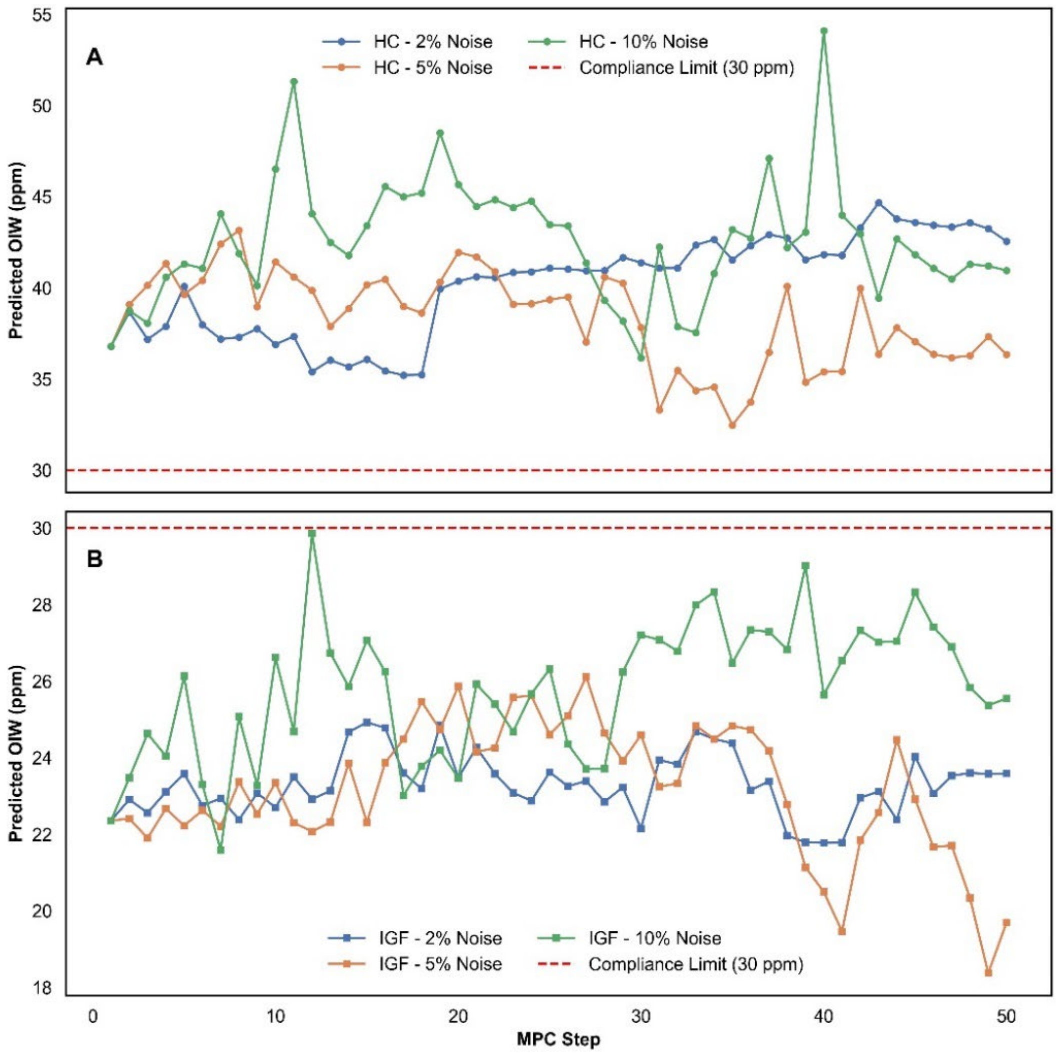


***Figure 9.7**: Predicted outlet oil-in-water concentrations over 50 MPC steps under different process noise levels: (A) Hydrocyclone and (B) IGF*

Fig. 9.7 presents a two-tiered analysis of predicted outlet oil-in-water concentrations under varying process noise levels for both the Hydrocyclone and IGF units across 50 Model Predictive Control (MPC) steps under 2%, 5%, and 10% process noise. In Fig. 9.7A for all scenarios, the predicted oil-in-water consistently exceeded the regulatory compliance limit of 30 ppm. At 2% noise, oil-in-water values fluctuate within a relatively narrow band of 35 to 44 ppm, while 5% noise introduces greater variability, with oscillations between approximately 34 and 45 ppm. Under 10% noise, the system exhibits marked instability, with frequent spikes reaching up to 54 ppm, underscoring the Hydrocyclone's sensitivity to process uncertainty and its inability to maintain compliance independently.

Fig. 9.7B illustrates the IGF unit's response under identical noise conditions. The IGF demonstrates strong resilience, maintaining all predicted oil-in-water concentrations below the

30 ppm limit. At 2% noise, outlet concentrations remain tightly controlled between 21 and 24 ppm. Even under 5% and 10% noise, the system stays within safe margins, with maximum values approaching 26 and 29 ppm, respectively. Notably, even at 10% noise, no compliance breaches occur, highlighting the IGF's robust performance.

This comparative analysis emphasises that while the Hydrocyclone is effective for bulk oil removal, the IGF unit plays a critical role in ensuring regulatory compliance for Angolan waters, particularly under uncertain operating conditions.These results further validate the necessity of an integrated digital twin control strategy that leverages the IGF's adaptive capacity to compensate for upstream variability and maintain outlet quality within acceptable thresholds.

To further analyse system behavior and control dynamics, Fig. 9.8 presents the evolution of all controllable variables across the 50 MPC steps. At each simulated inlet value, all other controllable features were re-optimised using the L-BFGS-B algorithm, and predicted outlet concentrations for both the Hydrocyclone and IGF were recorded. The bounded nature of the optimisation ensured realistic recommendations within known plant operating limits. This multi-panel layout highlights how the optimisation engine dynamically modulated operational features to steer the process in response to each new noisy condition, demonstrating the system's ability to adaptively maintain performance.

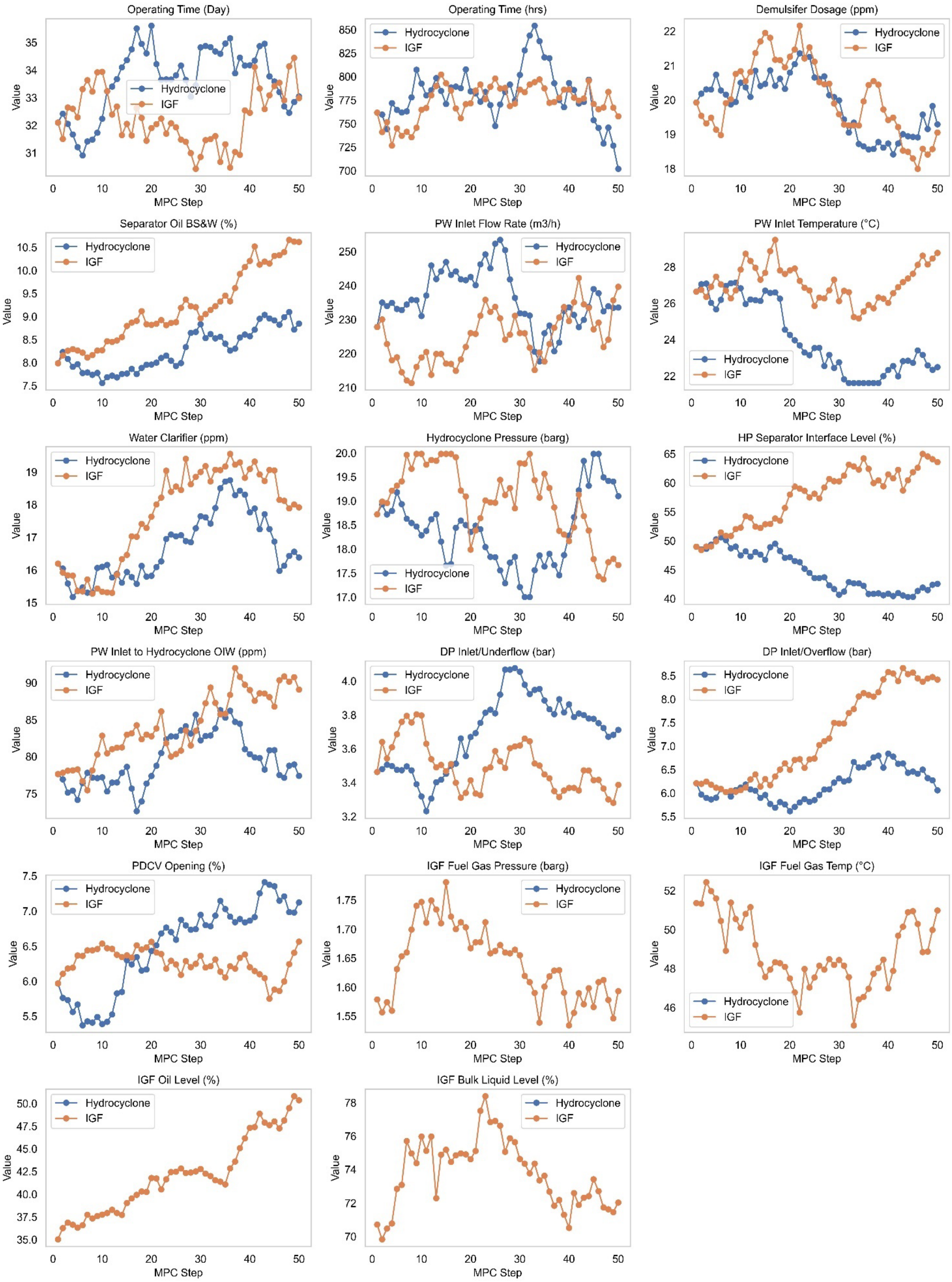


***Figure 9.8**: Evolution of controllable operational variables across 50 MPC steps for Hydrocyclone and IGF units, illustrating dynamic adjustments made by the optimisation engine to maintain performance under varying noise conditions.*

9.4. Conclusion

This chapter developed and validated a comprehensive digital twin framework for offshore produced water deoiling, integrating machine learning, Monte Carlo simulation, optimisation, and interpretability layers. The findings demonstrate that while the Hydrocyclone serves effectively as a bulk separator, it struggles to meet regulatory thresholds on its own, particularly under variable or degraded inlet conditions. In contrast, the IGF unit is critical in ensuring final compliance, consistently maintaining oil-in-water concentrations below 30 ppm even under stress and noise scenarios.

The incorporation of L-BFGS-B-based optimisation enables operational adjustments within realistic constraints, providing a powerful prescriptive tool for informed decision-making. Furthermore, the MPC-inspired control loop showcases the digital twin's potential as an adaptive, supervisory system capable of making real-time adjustments in response to uncertainty. Overall, the framework not only supports regulatory compliance but also enhances operational efficiency and environmental stewardship, laying the groundwork for future integration into autonomous control architectures.

CHAPTER 10

CONCLUSIONS AND RECOMMENDATIONS

"My grace is sufficient for you, for my power is made perfect in weakness."

2 Corinthians 12:9

CHAPTER 10: CONCLUSIONS AND RECOMMENDATIONS

10.1. Conclusions

This thesis has advanced the field of offshore produced water management by integrating deep process understanding, advanced physicochemical characterisation, anomaly detection, Machine Learning modeling, probabilistic simulation and digital twin technologies into a unified, practical framework. Through detailed physicochemical characterisation, it was shown that variations in produced water properties significantly influence the efficiency of HIGF systems, confirming that comprehensive produced water characterisation can guide more adaptive and resilient treatment strategies.

The integration of data-driven anomaly detection algorithms enabled real-time identification of operational deviations, providing early warning capabilities and reducing potential downtime while supporting continuous compliance and proactive operational control. By developing and comparing ensemble-based Machine Learning models, it was found that Extreme Gradient Boosting (XGBoost) and Gradient Boosting offer superior predictive accuracy and robustness for oil-in-water outlet concentration prediction in produced water deoiling systems. Critical operational features such as produced water inlet oil-in-water concentration, clarifier dosage and Hydrocyclone differential pressures were identified as the most influential, providing clear insights for operational optimisation.

The Machine Learning–aided Monte Carlo simulation framework allowed for probabilistic compliance risk assessment, forecasting exceedance probabilities under varying conditions and explicitly quantifying operational uncertainties. While this probabilistic approach improved scenario understanding and risk mapping, limitations such as smoothing of extreme events were noted.

The digital twin architecture designed and validated in this study successfully integrated predictive analytics, real-time optimisation and control recommendations. The incorporation of Model Predictive Control (MPC)-inspired logic demonstrated the capability to dynamically recalibrate operational setpoints, maintain compliance under fluctuating inlet conditions and reduce manual intervention.

Overall, the research confirmed that Machine Learning models, combined with stochastic simulation and digital twin strategies, can theoretically be validated for offshore produced water HIGF deoiling system's deployment, enhancing predictive monitoring, control optimisation and compliance assurance.

10.2. Recommendations

Based on the findings and insights of this thesis, the following recommendations are proposed to guide future research and industrial application:

- Field validation and pilot testing: Validate the models in real offshore conditions to confirm accuracy and operational readiness.
- Gradual deployment: Begin with monitoring and advisory use, then move to semi-autonomous and fully autonomous control.
- Improve data and interfaces: Enhance data integration and develop user-friendly, intuitive dashboards to support operators.
- Adaptive learning: Include online learning and drift detection to keep models accurate and effective over time.
- Advanced probabilistic modeling: Adopt multivariate or generative methods to better capture complex operational scenarios and risks.
- System-wide integration: Extend the digital twin to include upstream and downstream processes for holistic optimisation.
- Cross-disciplinary collaboration: Partner with technology providers and engineers to standardise deployment and integration.

10.3. Final Reflections

This thesis demonstrated that offshore produced water deoiling can evolve from rigid, reactive systems to intelligent, predictive and adaptive operations. By harnessing the combined power of Machine Learning, probabilistic simulation and digital twin technology, the work provides a roadmap for achieving higher operational efficiency, improved compliance and enhanced environmental stewardship.

Beyond technical contributions, this approach embodies a strategic shift toward safer, cleaner and smarter offshore operations, reinforcing the industry's commitment to responsible resource management and sustainability. Future efforts should focus on real-time deployment, continuous learning and expansion to integrated, plant-wide treatment systems, driving the next generation of intelligent water management solutions in offshore oil and gas production.

REFERENCES

REFERENCES

Abbas, A., Gzar, H. and Rahi, M., 2021, February. Oilfield-produced water characteristics and treatment technologies: a mini review. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1058, No. 1, p. 012063). IOP Publishing.

Abuhasel, K., Kchaou, M., Alquraish, M., Munusamy, Y., Jeng, Y.T., 2021. Oily wastewater treatment: Overview of conventional and modern methods, challenges, and future opportunities. *Water* 13(7), 980.

Abdalla, K. and Hammam, G., 2014. Correlation between biochemical oxygen demand and chemical oxygen demand for various wastewater treatment plants in Egypt to obtain the biodegradability indices. *International Journal of Sciences: Basic and Applied Research*, 13(1), pp.42-48.

Abdi, H. and Williams, L.J., 2012. Partial least squares methods: partial least squares correlation and partial least square regression. *In Computational Toxicology*: Volume II (pp. 549-579). Totowa, NJ: Humana Press.

Adewoye, A.J., 2020. Improvement to hydrocyclone used in separating particles from produced water in the oil and gas industry (Doctoral dissertation).

Aghbashlo, M., Peng, W., Tabatabaei, M., Kalogirou, S.A., Soltanian, S., Hosseinzadeh-Bandbafha, H., Mahian, O. and Lam, S.S., 2021. Machine learning technology in biodiesel research: a review. *Progress in Energy and Combustion Science*, 85, p.100904.

Ahan, J.A., 2014. *Characterisation of produced water from two offshore oil fields in Qatar*. Master's thesis. Qatar University.

Alardhi, S.M., Jabbar, N.M., Breig, S.J.M., Hadi, A.A., Salman, A.D., Al Saedi, L.M., Khadium, M.K., Showeel, H.A., Malak, H.M., Mohammed, M.M. and Le, P.C., 2024. Artificial neural network and response surface methodology for modeling oil content in produced water from an Iraqi oil field. *Water Practice & Technology*, 19(8), pp.3330-3349.

Alberta Innovates, 2017. New steam generation technology could eliminate GHG emissions and water treatment requirements in SAGD. Alberta Innovates. Edmonton, Alberta, Canada. Accessed online at: https://albertainnovates.ca/project/new-steam-generation-technology-could-eliminate-ghg-emissions-and-water-treatment-requirements-in-sagd/

Ali, H.N., 2017. Fundamentals of petroleum geology. *Springer Handbook of Petroleum Technology*, pp.321-357.

Ali, Z.A., Abduljabbar, Z.H., Taher, H.A., Sallow, A.B. and Almufti, S.M., 2023. Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review. *Academic Journal of Nawroz University*, *12*(2), pp.320-334.

Al-Ghouti, M.A., Al-Kaabi, M.A., Ashfaq, M.Y. and Da'na, D.A., 2019. Produced water characteristics, treatment and reuse: A review. *Journal of Water Process Engineering*, *28*, pp.222-239.

Al-Kaabi, M., Al-Ghouti, M., Ashfaq, M.Y., Ahmed, T. and Zouari, N., 2019. An integrated approach for produced water treatment using microemulsions modified activated carbon. *Journal of Water Process Engineering*, *31*, p.100830.

Al-Kayiem, H.H., Osei, H., Yin, K.Y., Hashim, F.M., 2014. A comparative study on the hydrodynamics of liquid–liquid hydrocyclonic separation. WIT Trans. Eng. Sci. 82, 361-370.

Amakiri, K., Canon, A., Molinari, M. and Angelis-Dimakis, A., 2022. Review of oilfield produced water treatment technologies. *Chemosphere*, *298*, p.134064.

Amakiri, K., Ogolo, N., Angelis-Dimakis, A. and Albert, O., 2023. Physicochemical assessment and treatment of produced water: A case study in Niger delta Nigeria. *Petroleum Research*, *8*(1), pp.87-95.

Ambrosio, J. 2014. *Handbook on oil production research. Energy Science, Engineering and Technology Series.* New York: Nova Science Publishers.

Amer, M., Goldstein, M. and Abdennadher, S., 2013, August. Enhancing one-class support vector machines for unsupervised anomaly detection. In *Proceedings of the ACM SIGKDD workshop on outlier detection and description* (pp. 8-15).

Anjum, A., Agbaje, P., Hounsinou, S. and Olufowobi, H., 2022, June. In-vehicle network anomaly detection using extreme gradient boosting machine. In *2022 11th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1-6). IEEE.

Arthur, J., Langhus, B. and Patel, C., 2005. Technical summary of oil & gas produced water treatment technologies. *All Consulting, LLC, Tulsa, OK.*

ASTM International, 2022. *ASTM D7678-17(2022) – standard test method for total oil and grease (TOG) and total petroleum hydrocarbons (TPH) in water and wastewater with solvent extraction using Mid-IR laser spectroscopy*. West Conshohocken, PA: ASTM International.

Aziz, N., Akhir, E.A.P., Aziz, I.A., Jaafar, J., Hasan, M.H. and Abas, A.N.C., 2020, October. A study on gradient boosting algorithms for development of AI monitoring and prediction systems. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 11-16). IEEE.

Balas, E.A., 2023. A hybrid Monte Carlo simulation risk model for oil exploration projects. *Marine Pollution Bulletin*, *194*, p.115270.

Baarimah, S.O., Baarimah, A.O., Alaloul, W.S., Bazel, M.A., Mohammed, F. and Bawahab, M., 2023, October. A bibliometric analysis on the applications of artificial intelligence in petroleum engineering. *In 2023 4th International Conference on Data Analytics for Business and Industry* (ICDABI) (pp. 152-159). IEEE.

Baarimah, A.O., Bazel, M.A., Alaloul, W.S., Alazaiza, M.Y., Al-Zghoul, T.M., Almuhaya, B., Khan, A. and Mushtaha, A.W., 2024. Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis. *Case Studies in Chemical and Environmental Engineering*, p.100926.

Bagheri, M., Roshandel, R. and Shayegan, J., 2018. Optimal selection of an integrated produced water treatment system in the upstream of oil industry. *Process Safety and Environmental Protection*, *117*, pp.67-81.

Bakke, T., Klungsøyr, J. and Sanni, S., 2013. Environmental impacts of produced water and drilling waste discharges from the Norwegian offshore petroleum industry. *Marine environmental research*, *92*, pp.154-169.

Bazel, M.A., Ahmad, M. and Mohammed, F., 2023. blockchain technology in Malaysian context: bibliometric analysis and systematic review. International Journal on Advanced Science, Engineering & Information Technology, 13(5).

Belia, E., Amerlinck, Y., Benedetti, L., Johnson, B., Sin, G., Vanrolleghem, P.A., Gernaey, K.V., Gillot, S., Neumann, M.B., Rieger, L. and Shaw, A., 2009. Wastewater treatment modelling: dealing with uncertainties. Water Science and Technology, 60(8), pp.1929-1941.

Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, pp.1937-1967.

Bram, M.V., Hansen, L., Hansen, D.S. and Yang, Z., 2018. Hydrocyclone separation efficiency modeled by flow resistances and droplet trajectories. *IFAC-papersonline*, 51(8), pp.132-137.

Bram, M.V., Jespersen, S., Hansen, D.S., Yang, Z., 2020. Control-oriented modeling and experimental validation of a deoiling hydrocyclone system. *Processes* 8(9), 1010.

Brar, L.S., Sharma, R.P., Elsayed, K., 2015. The effect of the cyclone length on the performance of Stairmand high-efficiency cyclone. *Powder Technology* 286, 668-677.

Bukar, U.A., Sayeed, M.S., Razak, S.F.A., Yogarayan, S., Amodu, O.A. and Mahmood, R.A.R., 2023. A method for analyzing text using VOSviewer. MethodsX, 11, p.102339.

Cauxeiro, C., Lopez, M., Miguel, A., Caetano, Á. and Cauxeiro, G., 2022. The oligocene to miocene turbidite deposits of the cabo ledo, inner kwanza basin, Angola: stratigraphic architecture and sedimentary processes. *Angolan Mineral, Oil & Gas Journal*, *3*(3), pp.42-47.

César, S.D., 2022. Topside facilities' ability to handle a new blend of crude oil (Doctoral dissertation, Cape Peninsula University of Technology).

César, S.D., De Jager, D. and Njoya, M., 2024. Environmental trade-offs in energy production: A review of the produced water life cycle and environmental footprint. *Marine Pollution Bulletin*, *203*, p.116480.

César, S.D., De Jager, D. and Njoya, M., 2024. The role of Hydrocyclone and Induced Gas Flotation technologies in offshore produced water deoiling advancements. *Petroleum Research*.

Chakraborty, S., Patra, G., Meikap, B.C., 2018. Role of vortex finder depth on pressure drop and performance efficiency in a ribbed hydrocyclone*. South African Journal of Chemical Engineering* 25(1), 103-109.

Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), pp.1-58.

Chen, Z., Zhao, L., Lee, K. and Hannath, C., 2007. Modeling and assessment of the produced water discharges from offshore petroleum platforms. *Water Quality Research Journal*, *42*(4), pp.303-310.

Chen, Z., Zhao, L. and Lee, K., 2010. Environmental risk assessment of offshore produced water discharges using a hybrid fuzzy-stochastic modeling approach. *Environmental Modelling & Software*, *25*(6), pp.782-792.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Cheng, B., Zhao, J., Yang, C., Tian, Y. and Liao, Z., 2017. Geochemical evolution of occluded hydrocarbons inside geomacromolecules: a review. *Energy & Fuels*, *31*(9), pp.8823-8832.

Choi, Y., Baek, S., Kim, J., Choi, J., Hur, J., Lee, T., Park, C. and Lee, B., 2017. Characteristics and biodegradability of wastewater organic matter in municipal wastewater treatment plants collecting domestic wastewater and industrial discharge. *Water*, 9(6), p.409.

Cilliers, J.J., 2000. Particle Size Separation| Hydrocyclones for particle size separation.

Cordes, E.E., Jones, D.O., Schlacher, T.A., Amon, D.J., Bernardino, A.F., Brooke, S., Carney, R., DeLeo, D.M., Dunlop, K.M., Escobar-Briones, E.G., Gates, A.R., 2016. Environmental impacts of the deep-water oil and gas industry: a review to guide management strategies. *Frontiers in Environmental Science* 4, 58.

Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U. and Poch, M., 2018. Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. *Environmental modelling & software*, *106*, pp.89-103.

Costa-Campi, M.T., del Rio, P. and Trujillo-Baute, E., 2017. Trade-offs in energy and environmental policy. *Energy Policy*, 104, pp.415-418.

Cutler, A., Cutler, D.R. and Stevens, J.R., 2012. Random forests. *Ensemble machine learning: Methods and applications*, pp.157-175.

Das, T. and Jäschke, J., 2018. Modeling and control of an inline deoiling hydrocyclone. *IFAC-PapersOnLine*, 51(8), pp.138-143.

Das, A., Sudhanshu, R., Shukla, G. and Chaturvedi, N.D., 2025. Artificial intelligence-based approach for water-cut prediction in crude oil processing. *Engineering Applications of Artificial Intelligence*, 142, p.109892.

Dawoud, H., Saleem, H., Alnuaimi, N. and Zaidi, S., 2021. Characterisation and treatment technologies applied for produced water in Qatar. *Water*, *13*(24), p.3573.

de Araújo, C.A.O., Scheid, C.M., Loureiro, J.B.R., Klein, T.S. and Medronho, R.A., 2020. Hydrocylone for oil-water separations with high oil content: Comparison between CFD simulations and experimental data. *Journal of Petroleum Science and Engineering*, 187.

de Campos Souza, P.V., 2020. Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature. *Applied soft computing*, *92*, p.106275.

de Canete, J.F., del Saz-Orozco, P., Gómez-de-Gabriel, J., Baratti, R., Ruano, A. and Rivas-Blanco, I., 2021. Control and soft sensing strategies for a wastewater treatment plant using a neuro-genetic approach. *Computers & Chemical Engineering*, *144*, p.107146.

de Ville, B., 2013. Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*(6), pp.448-455.

Ding, S., Li, H., Su, C., Yu, J. and Jin, F., 2013. Evolutionary artificial neural networks: a review. *Artificial Intelligence Review*, 39, pp.251-260.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W.M., 2021. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*, 133, pp.285-296.

Dórea, H., Bispo, J., Aragão, K., Cunha, B., Navickiene, S., Alves, J., Romão, L. and Garcia, C., 2007. Analysis of BTEX, PAHs and metals in the oilfield produced water in the State of Sergipe, Brazil. *Microchemical journal*, *85*(2), pp.234-238.

Durand, B., 2019. *Petroleum, natural gas, and coal: Nature, formation mechanisms, future prospects in the energy transition*. EDP sciences. ISBN: 9782759822317.

Durdevic, P., Pedersen, S., Yang, Z., 2017. Challenges in modelling and control of offshore deoiling hydrocyclone systems. *Journal of Physics: Conference Series*, 783(1), 012048.

Durdevic, P. and Yang, Z., 2018. Dynamic efficiency analysis of an off-shore hydrocyclone system, subjected to a conventional PID-and robust-control-solution. *Energies*, 11(9), p.2379.

Dwyer, B.P., 2016. Treatment of Oil & Gas Produced Water (No. SAND2016-1153). Sandia National Lab. (SNL-NM), Albuquerque, NM (United States).

Effendi, H., Mursalin, M. and Hariyadi, S., 2022. Rapid Water Quality Assessment as a Quick Response of Oil Spill Incident in Coastal Area of Karawang, Indonesia. *Frontiers in Environmental Science*, 10, p.757412.

Ekechukwu, O.M., Asim, T., Hawez, H.K., 2024. Recent Developments in Hydrocyclone Technology for Oil-in-Water Separation from Produced Water. *Energies* 17(13), 3181.

Eldos, H., Khan, M., Zouari, N., Saeed, S. and Al-Ghouti, M., 2022. Characterisation and assessment of process water from oil and gas production: A case study of process wastewater in Qatar. *Case Studies in Chemical and Environmental Engineering*, 6, p.100210.

El-Rawy, M., Abd-Ellah, M.K., Fathi, H. and Ahmed, A.K.A., 2021. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *Journal of Water Process Engineering*, 44, p.102380.

Energy Institute., 2023. Statistical Review of World Energy 2023. 72nd Edition. [Online]. Available from: file:///C:/Users/nre/Downloads/Statistical%20Review%20of%20World%20Energy.pdf

Endres, E., Dueck, J., Neesse, T., 2012. Hydrocyclone classification of particles in the micron range. *Minerals Engineering* 31, 42-45.

Epelle, E. and Gerogiorgis, D., 2020. A review of technological advances and open challenges for oil and gas drilling systems engineering. *AIChE Journal*, *66*(4), p.e16842.

Erikli, Ş., Okay, A.B., 2015, July. Inlet diameter and flow volume effects on separation and energy efficiency of hydrocyclones. In: *IOP Conference Series: Materials Science and Engineering* 90(1), 012074.

Eyitayo, S.I., Watson, M.C., Kolawole, O., Xu, P., Bruant, R., Henthorne, L., 2023. Produced water treatment: Review of technological advancement in hydrocarbon recovery processes, well stimulation, and permanent disposal wells. *SPE Production & Operations* 38(01), 51-62.

Fadul, A.M.A., 2023. Anomaly Detection based on Isolation Forest and Local Outlier Factor. *Africa University*.

Fakhru'l-Razi, A., Pendashteh, A., Abdullah, L., Biak, D., Madaeni, S. and Abidin, Z., 2009. Review of technologies for oil and gas produced water treatment. *Journal of hazardous materials*, *170*(2-3), pp.530-551.

Faksness, L., Grini, P. and Daling, P., 2004. Partitioning of semi-soluble organic compounds between the water phase and oil droplets in produced water. *Marine pollution bulletin*, 48(7-8), pp.731-742.

Feng, D.C., Liu, Z.T., Wang, X.D., Chen, Y., Chang, J.Q., Wei, D.F. and Jiang, Z.M., 2020. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials, 230, p.117000.

Fillo, J., Koraido, S. and Evans, J.M., 1992. Sources, characteristics, and management of produced waters from natural gas production and storage operations. In *Produced water: Technological/environmental issues and solutions* (pp. 151-161). Boston, MA: Springer US.

Gabardo, I., Platte, E., Araujo, A. and Pulgatti, F., 2011. Evaluation of produced water from Brazilian offshore platforms. *Produced Water: Environmental Risks and Advances in Mitigation Technologies*, pp.89-113.

Gamwo, I.K., Azam, H. and Baled, H., 2022. Produced water treatment technologies: An overview. *Solid–Liquid Separation Technologies*, pp.1-24.

Gangwar, A., Rawat, S., Rautela, A., Yadav, I., Singh, A., Kumar, S., 2024. Current advances in produced water treatment technologies: A perspective of techno-economic analysis and life cycle assessment. *Environment, Development and Sustainability*, 1-35.

Gazali, A., Alkali, A., Mohammed, Y., Djauro, Y., Muhammed, D., Kodomi, M., 2017. Environmental impact of produced water and drilling waste discharges from the Niger Delta petroleum industry. *IOSR Journal of Engineering* (IOSRJEN) 7(06), 2250-3021.

Geekiyanage, S.C., Tunkiel, A. and Sui, D., 2021. Drilling data quality improvement and information extraction with case studies. *journal of petroleum Exploration and Production*, *11*, pp.819-837.

Ghafoori, S., Omar, M., Koutahzadeh, N., Zendehboudi, S., Malhas, R.N., Mohamed, M., Al-Zubaidi, S., Redha, K., Baraki, F., Mehrvar, M., 2022. New advancements, challenges, and future needs on treatment of oilfield produced water: A state-of-the-art review. *Separation and Purification Technology* 289, 120652.

Ghodrat, M., Qi, Z., Kuang, S.B., Ji, L., Yu, A.B., 2016. Computational investigation of the effect of particle density on the multiphase flows and performance of hydrocyclone. *Minerals Engineering* 90, 55-69.

Grand View Research. 2023. Produced water treatment market size, share, trends analysis and segment forecasts 2023 – 2030.

Gray, M., 2020, July. Reuse of produced water in the oil and gas industry. In *SPE International Conference and Exhibition on Health, Safety, Environment, and Sustainability?* (p. D021S004R001). SPE.

Guerra, K., Dahm, K., Dundorf, S., 2011. *Oil and gas produced water management and beneficial use in the Western United States* (pp. 1-113). US Department of the Interior, Bureau of Reclamation, Washington, DC.

Gul Zaman, H., Baloo, L., Pendyala, R., Singa, P.K., Ilyas, S.U. and Kutty, S.R.M., 2021. Produced water treatment with conventional adsorbents and MOF as an alternative: A review. *Materials*, *14*(24), p.7607.

Guo, B., Lyons, W. and Ghalambor, A. 2011. *Petroleum Production Engineering, A Computer-Assisted Approach*. Oxford: Golf Professional Publishing. ISBN: 9780080479958.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y.M., Park, J.P., Kim, J.H. and Cho, K.H., 2015. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences*, *32*, pp.90-101.

Hansen, L., Durdevic, P., Jepsen, K.L. and Yang, Z., 2018. Plant-wide optimal control of an offshore de-oiling process using MPC technique. IFAC-papersonline, 51(8), pp.144-150.

Hedar, Y., 2018. Pollution impact and alternative treatment for produced water. I*n E3S Web of Conferences* (Vol. 31, p. 03004). EDP Sciences.

Hejabi, N., Saghebian, S.M., Aalami, M.T. and Nourani, V., 2021. Evaluation of the effluent quality parameters of wastewater treatment plant based on uncertainty analysis and post-processing approaches (case study). Water Science and Technology, 83(7), pp.1633-1648.

Henze, M., Harremoes, P., Jansen, J. and Arvin, E. *Wastewater Treatment, Biological and Chemical Processes*. 3rd Edition. Berlin: Springer-Verlag, 2002.

Herawati, N., Dahlan, M., Yusuf, M., Iqbal, M., Roni, K. and Nasir, S., 2023. Removal of total dissolved solids from oil-field-produced water using ceramic adsorbents integrated with reverse osmosis. *Materials Today: Proceedings*, *87*, pp.360-365.

Hu, L., Jiang, W., Xu, X., Wang, H., Carroll, K., Xu, P. and Zhang, Y., 2022. Toxicological characterisation of produced water from the Permian Basin. *Science of the Total Environment*, 815, p.152943.

Husain, T., Veitch, B. and Bose, N., 2004. An ecological risk assessment methodology for screening discharge alternatives of produced water. *Human and Ecological Risk Assessment*, *10*(3), pp.505-524.

Husveg, T., Rambeau, O., Drengstig, T., Bilstad, T., 2007. Performance of a deoiling hydrocyclone during variable flow rates. *Minerals Engineering* 20(4), 368-379.

Ibrahim, A.F., Al-Dhaif, R., Elkatatny, S. and Shehri, D.A., 2021. Applications of artificial intelligence to predict oil rate for high gas–oil ratio and water-cut wells. *ACS omega*, 6(30), pp.19484-19493.

Icke, O., van Es, D.M., de Koning, M.F., Wuister, J.J.G., Ng, J., Phua, K.M., Koh, Y.K.K., Chan, W.J. and Tao, G., 2020. Performance improvement of wastewater treatment processes by application of machine learning. *Water Science and Technology*, 82(12), pp.2671-2680.

Ige, A.B., Adepoju, P.A., Akinade, A.O. and Afolabi, A.I., 2025. Machine learning in industrial applications: An In-depth review and future directions.

Igunnu, E. and Chen, G., 2014. Produced water treatment technologies*. International journal of low-carbon technologies*, 9(3), pp.157-177.

Jahn, F., Cook, M. and Graham, M., 2008. *Hydrocarbon exploration and production*, 2nd Ed. Vol. 55, Elsevier. ISBN: 9780080568836.

Janson, A., Santos, A., Katebah, M., Hussain, A., Minier-Matar, J., Judd, S. and Adham, S., 2015. Assessing the biotreatability of produced water from a Qatari gas field. *SPE journal*, *20*(05), pp.1113-1119.

Jesson J, Matheson L, Lacey FM. 2011. *Doing your systematic review - traditional and systematic techniques*. SAGE Publications Ltd, London

Ji, G., Sun, T., Zhou, Q., Sui, X., Chang, S. and Li, P., 2002. Constructed subsurface flow wetland for treating heavy oil-produced water of the Liaohe Oilfield in China. *Ecological Engineering*, *18*(4), pp.459-465.

Jiang, M., Liu, D., Zhao, L., Jia, G., 2011. Study on the effect of cone angle values on the pressure and velocity field of hydrocyclones. In: *5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1-4.

Jiang, W., Pokharel, B., Lin, L., Cao, H., Carroll, K.C., Zhang, Y., Galdeano, C., Musale, D.A., Ghurye, G.L. and Xu, P., 2021. Analysis and prediction of produced water quantity and quality in the Permian Basin using machine learning techniques. *Science of The Total Environment*, 801, p.149693.

Jiang, W., Xu, X., Hall, R., Zhang, Y., Carroll, K.C., Ramos, F., Engle, M., Lin, L., Wang, H., Sayer, M. and Xu, P., 2022. Characterisation of produced water and surrounding surface water in the Permian Basin, the United States. *Journal of Hazardous Materials*, 430, p.128409.

Jesus, G., Casimiro, A. and Oliveira, A., 2021. Using machine learning for dependable outlier detection in environmental monitoring systems. *ACM Transactions on Cyber-Physical Systems*, *5*(3), pp.1-30.

Jiménez, S., Micó, M. M., Arnaldos, M., Medina, F., Contreras, S., 2018. State of the art of produced water treatment. *Chemosphere* 192, 186-208.

Johnsen, S., Røe Utvik, T., Garland, E., de Vals, B. and Campbell, J., 2004, March. Environmental fate and effect of contaminants in produced water. *In SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production*. OnePetro.

Kadiyala, A. and Kumar, A., 2018. Applications of python to evaluate the performance of bagging methods. *Environmental Progress & Sustainable Energy*, 37(5), pp.1555-1559.

Katragadda, S., Kehinde, O. and Kezron, I.E., 2020. Anomaly detection: Detecting unusual behavior using machine learning algorithms to identify potential security threats or system failures. *International Research Journal of Modernization in Engineering Technology and Science*, *2*(5), pp.1342-1350.

Kaparthi, S. and Bumblauskas, D., 2020. Designing predictive maintenance systems using decision tree-based machine learning techniques. *International Journal of Quality & Reliability Management*, *37*(4), pp.659-686.

Kharoua, N., Khezzar, L., Nemouchi, Z., 2010. Hydrocyclones for de-oiling applications—A review. *Petroleum Science and Technology* 28(7), 738-755.

Knutsen, T., 2011. Aerobic biological treatment of produced water from oil production (Master's thesis, University of Stavanger, Norway).

Kramer, O. and Kramer, O., 2013. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors*, pp.13-23.

Kuhn, M., Weston, S., Keefer, C. and Coulter, N., 2012. Cubist models for regression. *R package Vignette R package version 0.0*, *18*, p.480.

Kumar, L., Chugh, M., Kumar, S., Kumar, K., Sharma, J. and Bharadvaja, N., 2022. Remediation of petrorefinery wastewater contaminants: A review on physicochemical and bioremediation strategies. *Process Safety and Environmental Protection*, 159, pp.362-375.

Kvenvolden, K., 2006. Organic geochemistry–A retrospective of its first 70 years. *Organic Geochemistry*, 37(1), pp.1-11.

Lacalamita, D., Mongioví, C. and Crini, G., 2024. Chemical oxygen demand and biochemical oxygen demand analysis of discharge waters from laundry industry: monitoring, temporal variability, and biodegradability. *Frontiers in Environmental Science*, 12, p.1387041.

Lahlou, F., Mackey, H., McKay, G., Onwusogh, U. and Al-Ansari, T., 2020. Water planning framework for alfalfa fields using treated wastewater fertigation in Qatar: An energy-water-food nexus approach. *Computers & Chemical Engineering*, 141, p.106999.

Lampropoulos, G., 2023. Artificial intelligence, big data, and machine learning in industry 4.0. In *Encyclopedia of Data Science and Machine Learning* (pp. 2101-2109). IGI Global.

Larson, A., 2018. Produced Water: Oil and gas terminology glossary. *Water Environment Federation*, Factsheet. WSEC-2017-FS-013 IWWC.

Lee, J.H., Shin, J. and Realff, M.J., 2018. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering*, 114, pp.111-121.

Lee, J.W., Jung, K.J. and Kim, Y.J., 2020. Numerical analysis on performance of induced gas flotation machine using MUSIG model. *Engineering Applications of Computational Fluid Mechanics*, 14(1), pp.778-789

Lee, S., Park, J., Kim, N., Lee, T. and Quagliato, L., 2023. Extreme gradient boosting-inspired process optimisation algorithm for manufacturing engineering applications. *Materials & Design*, *226*, p.111625.

Leffler, W., Pattarozzi, R. & Sterling, G. 2011. *Deepwater petroleum exploration & production: A nontechnical guide*. Oklahoma: PennWell Books.

Li, Y., Wang, J., Ji, H., Li, O. and Nie, S., 2020. Numerical simulation analysis of main structural parameters of hydrocyclones on oil-gas separation effect. *Processes*, 8(12), p.1624.

Liu, Y., Lu, H., Li, Y., Xu, H., Pan, Z., Dai, P., Wang, H., Yang, Q., 2021. A review of treatment technologies for produced water in offshore oil and gas fields. *Science of the Total Environment* 775, 145485.

Liu, S., Lo, C.K. and Kan, C.W., 2022. Application of artificial intelligence techniques in textile wastewater decolorisation fields: A systematic and citation network analysis review. *Coloration Technology*, *138*(2), pp.117-136.

Lowe, M., Qin, R. and Mao, X., 2022. A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. *Water*, *14*(9), p.1384.

Lu, J., Wang, X., Shan, B., Li, X. and Wang, W., 2006. Analysis of chemical compositions contributable to chemical oxygen demand (COD) of oilfield produced water. *Chemosphere*, 62(2), pp.322-331.

Lusinier, N., Seyssiecq, I., Sambusiti, C., Jacob, M., Lesage, N., 2019. Biological treatments of oilfield produced water: A comprehensive review. *SPE Journal*, 2135-2147.

Ly, Q., Truong, V., Ji, B., Nguyen, X., Cho, K., Ngo, H. and Zhang, Z., 2022. Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. *Science of The Total Environment*, 832, p.154930.

Lyons, W., Plisga, G. & Lorenz, M., 2015. *Standard Handbook of Petroleum and Natural Gas Engineering Book*. 3rd Edition. Texas: Gulf Professional Publishing.

Mallants, D., Šimůnek, J. and Torkzaban, S., 2017. Determining water quality requirements of coal seam gas produced water for sustainable irrigation. *Agricultural water management*, *189*, pp.52-69.

Malviya, A. and Jaspal, D., 2021. Artificial intelligence as an upcoming technology in wastewater treatment: a comprehensive review. Environmental Technology Reviews, 10(1), pp.177-187.

Marcot, B.G. and Penman, T.D., 2019. Advances in Bayesian network modelling: Integration of modelling technologies. *Environmental modelling & software*, *111*, pp.386-393.

Mastouri, R., Borghei, S.M., Nadim, F., Roayaei, E., 2010. The effect of temperature and impeller speed on mechanically Induced Gas Flotation (IGF) performance in separation of Oil from oilfield-produced water. *Petroleum Science and Technology 28*(14), 1415-1426.

Matheri, A., Ntuli, F., Ngila, J., Seodigeng, T. and Zvinowanda, C., 2021. Performance prediction of trace metals and cod in wastewater treatment using artificial neural network. Computers & Chemical Engineering, 149, p.107308.

Maulud, D. and Abdulazeez, A.M., 2020. A review on linear regression comprehensive in machine learning. *Journal of applied science and technology trends*, 1(2), pp.140-147.

McCabe, P., 2013. Oil and Natural gas: global resources. In *Fossil Energy* (pp. 7-23). Springer, New York, NY.

Medar, R., Rajpurohit, V.S. and Rashmi, B., 2017, August. Impact of training and testing data splits on accuracy of time series forecasting in machine learning. *In 2017 International Conference on Computing, Communication, Control and Automation* (ICCUBEA) (pp. 1-6). IEEE.

Miller, T., Durlik, I., Adrianna, K., Kisiel, A., Cembrowska-Lech, D., Spychalski, I. and Tuński, T., 2023. Predictive Modeling of urban lake water quality using machine learning: A 20-year study. *Applied Sciences*, *13*(20), p.11217.

Mousa, K. and Al-Hasan, A., 2017, May. Oilfield produced water treatment by coagulation/flocculation processes. In *Proceedings of the Second Conference of Post Graduate Researches (CPGR'2017) College of Engineering*.

Mosca, E., Szigeti, F., Tragianni, S., Gallagher, D. and Groh, G., 2022, October. SHAP-based explanation methods: a review for NLP interpretability. In *Proceedings of the 29th international conference on computational linguistics* (pp. 4593-4603).

Moya-Anegón, S.G.F.D., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F.J. and Herrero-Solana, V., 2007. Visualizing the marrow of science. Journal of the American Society for Information Science and Technology, 58(14), pp.2167-2179.

Muraina, I., 2022, May. Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. In *7th international Mardin Artuklu scientific research conference* (pp. 496-504).

Nasiri, M. and Jafari, I., 2017. Produced water from oil-gas plants: A short review on challenges and opportunities. *Periodica Polytechnica Chemical Engineering*, *61*(2), pp.73-81

Nasiri, M., Jafari, I. and Parniankhoy, B., 2017. Oil and gas produced water management: a review of treatment technologies, challenges, and opportunities. *Chemical engineering communications*, 204(8), pp.990-1005.

Neff, J., Lee, K. and DeBlois, E., 2011. Produced water: overview of composition, fates, and effects. Produced water: *Environmental risks and advances in mitigation technologies*, pp.3-54.

Newhart, K.B., Holloway, R.W., Hering, A.S. and Cath, T.Y., 2019. Data-driven performance analyses of wastewater treatment plants: A review. *Water research*, *157*, pp.498-513.

Ni, L., Tian, J., Song, T., Jong, Y., Zhao, J., 2019. Optimizing geometric parameters in hydrocyclones for enhanced separations: a review and perspective. *Separation & Purification Reviews* 48(1), 30-51.

Niu, H., 2008. *Dispersion of offshore discharged produced water in the marine environment: hydrodynamic modeling and experimental study* (Doctoral dissertation, Memorial University of Newfoundland).

Nourani, V., Asghari, P. and Sharghi, E., 2021. Artificial intelligence-based ensemble modeling of wastewater treatment plant using jittered data. *Journal of Cleaner Production*, 291, p.125772.

Olajire, A., 2020. Recent advances on the treatment technology of oil and gas produced water for sustainable energy industry-mechanistic aspects and process chemistry perspectives. *Chemical Engineering Journal Advances*, *4*, p.100049.

Oliveira-Junior, J. and de Pereira, M., 2020. Forecasting Total Oil and Grease in produced water using Machine Learning methods in an oil extraction plant. *Marine Systems & Ocean Technology*, *15*(2), pp.124-134.

OPEC, 2023. OPEC Annual Statistical Bulletin. 58th Edition. [Online]. Available from: https://asb.opec.org/ASB_Chapters.html

Osei, H., Al-Kayiem, H.H., Hashim, F.M., 2015, November. Numerical studies on the separation performance of liquid-liquid Hydrocyclone for higher water-cut wells. In: *IOP Conference Series: Materials Science and Engineering* 100(1), 012014.

Overton, E., Wetzel, D., Wickliffe, J. and Adhikari, P., 2020. Spilled oil composition and the natural carbon cycle: the true drivers of environmental fate and effects of oil spills. *In Scenarios and Responses to Future Deep Oil Spills* (pp. 33-56). Springer, Cham.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*.

Pampanin, D.M. and Sydnes, M.O., 2013. Polycyclic aromatic hydrocarbons a constituent of petroleum: presence and influence in the aquatic environment. *Hydrocarbon*, *5*, pp.83-118.

Piccioli, M., Aanesen, S.V., Zhao, H., Dudek, M., Øye, G., 2020. Gas flotation of petroleum produced water: a review on status, fundamental aspects, and perspectives. *Energy & Fuels 34*(12), 15579-15592.

Qambar, A. and Al Khalidy, M., 2022. Prediction of municipal wastewater biochemical oxygen demand using machine learning techniques: A sustainable approach. *Process Safety and Environmental Protection*, 168, pp.833-845.

Qi, P., Zhou, W. and Han, J., 2017, April. A method for stochastic L-BFGS optimisation. In *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 156-160). IEEE.

Rakesh, A., Kumar Reddy, V.T.R., Narasimha, M., 2014. Air-Core Size Measurement of operating Hydrocyclone by electrical resistance tomography. *Chemical Engineering & Technology* 37(5), 795-805.

Rana, A., Rawat, A.S., Bijalwan, A. and Bahuguna, H., 2018, August. Application of multi layer (perceptron) artificial neural network in the diagnosis system: a systematic review. In *2018 International conference on research in intelligent and computing in engineering (RICE)* (pp. 1-6). IEEE.

Ray, S., 2019, February. A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.

Ray, S.S., Verma, R.K., Singh, A., Ganesapillai, M. and Kwon, Y.N., 2023. A holistic review on how artificial intelligence has redefined water treatment and seawater desalination processes. Desalination, 546, p.116221.

Razak, N.A., Mukhtar, H. and Mohshim, D.F., 2022. Characterisation of produced water from petroleum hydrocarbons terminal. *Journal of Positive School Psychology*, 6(3), pp.4415-4419.

Richerand, F., 2012. An historical perspective on produced water treatment. Society of Petroleum Engineering.

Rymarczyk, T., Kozłowski, E., Kłosowski, G. and Niderla, K., 2019. Logistic regression for machine learning in process tomography. *Sensors*, 19(15), p.3400.

Saboe, D., Ghasemi, H., Gao, M., Samardzic, M., Hristovski, K., Boscovic, D., Burge, S., Burge, R. and Hoffman, D., 2021. Real-time monitoring and prediction of water quality parameters and algae concentrations using microbial potentiometric sensor signals and machine learning tools. *Science of the Total Environment*, 764, p.142876.

Safeer, S., Pandey, R., Rehman, B., Safdar, T., Ahmad, I., Hasan, S. and Ullah, A., 2022. A review of artificial intelligence in water purification and wastewater treatment: Recent advancements. *Journal of Water Process Engineering*, 49, p.102974.

Saidi, M., Maddahian, R., Farhanieh, B., 2013. Numerical investigation of cone angle effect on the flow field and separation efficiency of deoiling hydrocyclones. *Heat and Mass Transfer* 49, 247-260.

Salles, R., Mendes, J., Araújo, R., Melo, C. and Moura, P., 2022, July. Prediction of key variables in wastewater treatment plants using machine learning models. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 1-9). IEEE.

Saththasivam, J., Loganathan, K., Sarp, S., 2016. An overview of oil–water separation using gas flotation systems. *Chemosphere* 144, 671-680.

Schobert, H., 2013. *Chemistry of fossil fuels and biofuels*. Cambridge University Press.

Schneider, P. and Xhafa, F., 2022. Chapter 3—Anomaly detection: Concepts and methods. *Anomaly detection and complex event processing over IoT data streams*, pp.49-66.

Schonlau, M. and Zou, R.Y., 2020. The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), pp.3-29.

Schulz, E., Speekenbrink, M. and Krause, A., 2018. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of mathematical psychology*, *85*, pp.1-16.

Schweidtmann, A.M., Esche, E., Fischer, A., Kloft, M., Repke, J.U., Sager, S. and Mitsos, A., 2021. Machine learning in chemical engineering: A perspective. Chemie Ingenieur Technik, 93(12), pp.2029-2039.

Selley, R., and Sonnenberg, S., 2014. *Elements of Petroleum Geology.* 3rd Ed. Academic Press. ISBN: 9780123860323

Shah, H., Majumder, A.K., Barnwal, J.P., 2006. Development of water split model for a 76 mm hydrocyclone. *Minerals Engineering* 19(1), 102-104.

Shaqiri, M., Iljazi, T., Kamberi, L. and Ramani-halili, R., 2023. Differences Between the correlation coefficients pearson, kendall and spearman. *Journal of Natural Sciences and Mathematics of UT*, *8*(15-16), pp.392-397.

Shekhar Jha, H., Mustafa Ud Din Sheikh, H. and Lee, W.J., 2021, November. Outlier detection techniques help us identify and remove outliers in production data to improve production forecasting. In *SPE/AAPG/SEG Asia Pacific Unconventional Resources Technology Conference* (p. D012S001R040). URTEC.

Sheth, V., Tripathi, U. and Sharma, A., 2022. A comparative analysis of machine learning algorithms for classification purpose. *Procedia Computer Science*, *215*, pp.422-431.

Shiekh, R.H.A. and El-Hashash, E.F., 2022. A comparison of the pearson, spearman rank and kendall tau correlation coefficients using quantitative variables. *Asian Journal of Probability and Statistics*, pp.36-48.

Shujaaddeen, A.A., Ba-Alwi, F.M., Zahary, A.T. and Alhegami, A.S., 2024, August. A model for measuring the effect of splitting data method on the efficiency of machine learning models: A comparative study. In *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)* (pp. 1-13). IEEE.

Si, M. and Du, K., 2020. Development of a predictive emissions model using a gradient boosting machine learning method. *Environmental Technology & Innovation*, *20*, p.101028.

Siagian, U.W., Widodo, S., Wardani, A.K., Wenten, I.G., 2018. Oilfield produced water reuse and reinjection with membrane. In: *MATEC Web of Conferences* 156, 08005.

Speiser, J.L., Miller, M.E., Tooze, J. and Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*, *134*, pp.93-101.

Spherical Insights., 2023. Global produced water treatment market size, share, analysis and forecast 2022-2032

Stewart, M and Arnold, K., 2008. *Gas-Liquid and Liquid-Liquid Separators*. Texas: Gulf Professional Publishing.

Suthaharan, S., 2016. Support vector machine. In *Machine learning models and algorithms for big data classification: thinking with examples for effective learning* (pp. 207-235). Boston, MA: Springer US.

Suzuki, R. and Alhadeff, E., 2022. Improving the quality of produced water discharged in open waters by incorporating a microbiological treatment on operating offshore platforms: Melhorar a qualidade da água produzida descartada em águas abertas, incorporando um tratamento microbiológico nas plataformas offshore operacionais. *Brazilian Journal of Development*, *8*(7), pp.54680-54699.

Tang, Z., Yu, L., Wang, F., Li, N., Chang, L., Cui, N., 2018. Effect of particle size and shape on separation in a hydrocyclone. *Water* 11(1), 16.

Tian, J., Ni, L., Song, T., Zhao, J., 2020. CFD simulation of hydrocyclone-separation performance influenced by reflux device and different vortex-finder lengths. *Separation and Purification Technology* 233, 116013.

Tissot, B. and Welte, D., 1984. *Petroleum Formation and Occurrence: A New Approach to Oil and Gas Exploration*. 2nd Ed. Berlin: Springer Verlag.

Tudorache, V., Avram, L. and Antonescu, N., 2020. Aspects on offshore drilling process in deep and very deep waters. *Technical Sciences*, 5(2), pp.157-172.

USEIA, 2016. *Trends in U.S. oil and natural gas upstream costs*. U.S. Energy Information Administration. Washington DC.

Vakamalla, T.R., Mangadoddy, N., 2019. The dynamic behaviour of a large-scale 250-mm hydrocyclone: A CFD study. *Asia-Pacific Journal of Chemical Engineering* 14(2), e2287.

Vallabhan, K., Holden, C. and Skogestad, S., 2020. A First-Principles Approach for Control-Oriented Modeling of De-oiling Hydrocyclones. *Industrial & Engineering Chemistry Research*, 59(42), pp.18937-18950.

Van Eck, N. and Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), pp.523-538.

Van Campen, L.J.A.M., 2014. Bulk dynamics of droplets in liquid-liquid axial cyclones (PhD Thesis). Delft University of Technology.

Van Houdt, G., Mosquera, C. and Nápoles, G., 2020. A review on the long short-term memory model. Artificial Intelligence Review, 53(8), pp.5929-5955.

Velthoen, J., Dombry, C., Cai, J.J. and Engelke, S., 2023. Gradient boosting for extreme quantile regression. *Extremes*, *26*(4), pp.639-667.

Vieira, L.G., Barrozo, M.A., 2014. Effect of vortex finder diameter on the performance of a novel hydrocyclone separator. *Minerals Engineering* 57, 50-56.

Vinutha, H.P., Poornima, B. and Sagar, B.M., 2018. Detection of outliers using interquartile range technique from intrusion dataset. In *Information and decision sciences: Proceedings of the 6th international conference on ficta* (pp. 511-518). Springer Singapore.

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M. and Souihi, N., 2021. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Science of the total environment*, 784, p.147138.

Wang, Y., Cheng, Y., Liu, H., Guo, Q., Dai, C., Zhao, M. and Liu, D., 2023. A review on applications of artificial intelligence in wastewater treatment. *Sustainability*, 15(18), p.13557.

Wang, R., Yu, Y., Chen, Y., Pan, Z., Li, X., Tan, Z. and Zhang, J., 2022. Model construction and application for effluent prediction in wastewater treatment plant: Data processing method optimisation and process parameters integration. *Journal of Environmental Management*, 302, p.114020.

Wu, X., Zheng, Z., Wang, L., Li, X., Yang, X. and He, J., 2023. Coupling process-based modeling with machine learning for long-term simulation of wastewater treatment plant operations. *Journal of Environmental Management*, *341*, p.118116.

Yang, X., Peng, K., An, L., Huang, P. and Feng, P., 2022, August. GMBLOF: A machine learning algorithm of novelty detection based on local outlier factor. In *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)* (pp. 20-25). IEEE.

Yang, J., Yang, F., Zhou, Y., Wang, D., Li, R., Wang, G. and Chen, W., 2021. A data-driven structural damage detection framework based on parallel convolutional neural network and bidirectional gated recurrent unit. *Information Sciences*, *566*, pp.103-117.

Ye, Z., Yang, J., Zhong, N., Tu, X., Jia, J. and Wang, J., 2020. Tackling environmental challenges in pollution controls using artificial intelligence: A review. *Science of the Total Environment*, *699*, p.134279.

Zaibel, I., Arnon, S. and Zilberg, D., 2022. Treated municipal wastewater as a water source for sustainable aquaculture: A review. *Reviews in Aquaculture*, 14(1), pp.362-377.

Zhang, Y., Cai, P., Jiang, F., Dong, K., Jiang, Y., Wang, B., 2017. Understanding the separation of particles in a hydrocyclone by force analysis. Powder Technology 322, 471-489.

Zhao, L., 2007. *Modeling and assessment of produced water discharges emitted from offshore petroleum platforms in the East Coast of Canada* (Doctoral dissertation, Concordia University).

Zhao, L., Chen, Z. and Lee, K., 2008. A risk assessment model for produced water discharge from offshore petroleum platforms-development and validation. *Marine pollution bulletin*, *56*(11), pp.1890-1897.

Zhao, L., 2012. *Development of an Integrated System for the Simulation and Assessment of Produced Water Discharges from Offshore Platforms* (Doctoral dissertation, Concordia University).

Zhao, Q., Cui, B., Wei, D., Song, T., Feng, Y., 2019. Numerical analysis of the flow field and separation performance in hydrocyclones with different vortex finder wall thickness. Powder Technology 345, 478-491.

Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J. and Yang, Y., 2020. Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. *Process Safety and Environmental Protection*, *133*, pp.169-182.

Zheng, J., Chen, B., Thanyamanta, W., Hawboldt, K., Zhang, B. and Liu, B., 2016. Offshore produced water management: A review of current practice and challenges in harsh/Arctic environments. *Marine pollution bulletin*, *104*(1-2), pp.7-19.

Zhou, Y., Duan, N., Wu, X. and Fang, H., 2018. COD discharge limits for urban wastewater treatment plants in China based on statistical methods. *Water*, 10(6), p.777.