SIMULATION-BASED MODELING AND OPTIMIZATION OF DRILLING PARAMETERS INFLUENCING RATE OF PENETRATION IN NIGER DELTA FORMATIONS

HERALD KOLAWOLE TAIWO

ENG2002636

DEPARTMENT OF PETROLEUM ENGINEERING

FACULTY OF ENGINEERING

UNIVERSITY OF BENIN

BENIN CITY

JULY 2025

DECLARATION

I, **Herald Kolawole Taiwo**, with Matriculation number **ENG2002636**, do

hereby declare that:

1. This project work is based on a study undertaken by me in the Department of Petroleum Engineering, University of Benin, Benin City, under the supervision of Mr Isuk .
2. This research work has not been previously submitted for the award of a degree elsewhere.
3. All ideas and views are a product of my personal research; and where the views of others have been expressed, they were duly acknowledged.
4. All liabilities arising from the study are entirely mine and not of the supervisor.

_____ _____

TAIWO HERALD KOLAWOLE **DATE**

CERTIFICATION

This is to certify that this project work was carried out by **Herald Kolawole Taiwo** with Matriculation Number **ENG2002636** under my supervision. It is adequate and satisfactory, both in scope and content, for the award of Bachelor of Science (BENG) Degree in Petroleum Engineering of the University of Benin.

ENGR. NSISONG ISUK ` **DATE**
(PROJECT SUPERVISOR)

DR. O.A TAIWO ` **DATE**

(PROJECT COORDINATOR)

ENGR. DR. IKPONMWOSA OHENHEN **DATE**

(HEAD OF DEPARTMENT)

PROF. KEVIN CHINWUBA IGWILO ` **DATE**

(EXTERNAL SUPERVISOR)

APPROVAL

This project work is hereby approved in partial fulfilment of the requirements for the award of Bachelor of Engineering (BENG) Degree in Petroleum Engineering from the University of Benin,

DR. I.K OHEHEN	**DATE (HEAD OF DEPARTMENT)**

DEDICATION

I dedicate this work to God Almighty, the source of my strength, wisdom, and perseverance throughout this academic journey.

I also dedicate it to my family, especially my parents, for their unwavering support, prayers, and encouragement.

To everyone who believed in me — this is for you.

ACKNOWLEDGEMENT

I sincerely thank God Almighty for His grace, wisdom, and strength throughout my academic journey and the successful completion of this project.

I am deeply grateful to my project supervisor, Engr. Isuk, for his consistent guidance, encouragement, and constructive feedback. His support played a crucial role in shaping this research.

My sincere appreciation also goes to Prof. K.O Bello, whose availability and academic support during critical moments greatly contributed to my academic growth.

I am especially grateful to my senior colleague and friend, Onoakposeha Efekodha whose mentorship, guidance, and project format greatly helped me understand and structure my own work from start to finish.

I also extend my appreciation to Pastor and Dr. Johnson, whose prayers, spiritual and academic guidance, and advice on time management strengthened my focus and determination throughout this work.

I am also thankful to my friend, Udogwu Chika, for her continuous encouragement, timely advice, and motivation to complete this project on schedule.

Special thanks go to my father, Mr. Alamu Dele Taiwo, for his unwavering financial and moral support, which made this academic pursuit possible.

Finally, I appreciate my classmates and everyone who, in one way or another, offered support, shared ideas, or provided encouragement during the course of this research.

ABSTRACT

This project investigates the effect of key drilling parameters on Rate of Penetration (ROP) using real-world field data from a selected well. The parameters analyzed include Weight on Bit (WOB), Rotational Speed (RPM), and mud properties such as Plastic Viscosity, Yield Point, and Gel Strength. The study aims to understand how variations in these parameters influence ROP and to identify combinations that could enhance drilling efficiency. Microsoft Excel was used for organizing, calculating, and analyzing the data, with additional tools such as Solver applied for basic optimization. By focusing on a practical, data-driven approach, this work contributes to ongoing efforts in optimizing drilling operations, especially in regions where advanced software and models may be inaccessible. The findings provide insight into the practical relationships between operational parameters and ROP, and highlight opportunities for performance improvement in similar field environments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF STUDY

In the field of petroleum engineering, drilling remains one of the most critical and cost-intensive operations in the upstream oil and gas sector. Among the various performance metrics, the Rate of Penetration (ROP) stands out as a key indicator of drilling efficiency. It represents the speed at which the drill bit advances through subsurface formations, typically measured in meters per hour or feet per hour. Optimizing ROP helps reduce drilling time, lower non-productive time (NPT), and minimize overall operational costs.

ROP is influenced by both controllable and uncontrollable factors, including Weight on Bit (WOB), Rotational Speed (RPM), mud properties such as Plastic Viscosity (PV), Yield Point (YP), and Gel Strength**,** as well as bit type and formation lithology. Understanding how these parameters interact and contribute to ROP is essential for improving drilling performance and avoiding equipment failure or delays.

Various researchers have applied advanced techniques such as Artificial Neural Networks (ANNs)**,** Response Surface Methodology (RSM)**,** and metaheuristic algorithms like **the** Bat Algorithm to predict or optimize ROP. For example, Moraveji and Naderi (2016) successfully combined RSM with the Bat Algorithm to develop predictive models for ROP. While these methods are powerful, they often require access to specialized software and computing resources, which may not be readily available to all students or field engineers.

This study analyzed simulated drilling data representing conditions typical of Niger Delta formations to examine how selected drilling parameters affect the rate of penetration (ROP). Python programming was the primary analytical tool used for data generation, statistical modeling, and visualization. Microsoft Excel was employed only for organizing and formatting the simulated datasets produced in Python. This combined workflow ensured computational accuracy, reproducibility, and efficiency within the available timeframe and resources.

1.2 PROBLEM STATEMENT

The drilling phase of oil and gas well development is one of the most expensive and technically demanding stages in upstream petroleum operations. A key performance indicator during drilling is the rate of penetration (ROP), which directly influences the time and cost required to reach the target depth. Achieving an optimal ROP can significantly improve drilling efficiency, reduce operational costs, and enhance overall well performance. However, ROP is a complex function of numerous interrelated factors, including weight on bit (WOB), rotary speed (RPM), bit type, mud rheology, and the lithological nature of the formation.

In many operational environments, particularly in developing regions or academic settings, there is limited access to sophisticated simulation software or real-time optimization tools. As a result, decisions affecting ROP are often made based on trial and error, general field experience, or outdated charts. This approach may lead to suboptimal drilling performance, increased non-productive time (NPT), and higher costs due to equipment wear, inefficient bit selection, or poor mud design.

Furthermore, although past research has introduced advanced mathematical models and machine learning techniques to predict ROP, such approaches typically require complex software, high computational resources, or large datasets. These requirements make it difficult for students, small operators, or engineers in low-resource settings to apply such models effectively. There is a clear gap between high-level ROP prediction studies and what is practically implementable in many real-world scenarios.

This study seeks to bridge that gap by generating and analyzing simulated data based on known ROP Models and parameter relationships. Using accessible tools such as Microsoft Excel, and Python programming, the project aims to evaluate the influence of selected drilling parameters on ROP. The study aims to demonstrate that accurate, data-driven optimization of drilling parameters can be achieved using open-source computational tools such as Python, thereby offering a cost-effective alternative to commercial drilling software.

1.3 AIM OF THE STUDY

The aim of this study is to analyze the effect of selected drilling parameters on the rate of penetration (ROP) using simulated data derived from literature-based models representative of Nigerian drilling conditions The objective is to evaluate how factors such as weig’’’ht on bit, rotary speed, mud properties, bit type, and formation characteristics influence ROP, with the goal of improving drilling efficiency through data-driven and computational methods.

1.4 OBJECTIVES OF THE STUDY

The specific objectives of this study are to:

1. Generate and organize simulated drilling data representing real field conditions within a Nigerian sedimentary basin, focusing on parameters that influence the rate of penetration (ROP).

2. Compute ROP values based on input variables such as weight on bit (WOB), rotary speed (RPM), and mud properties using Python-based analytical techniques.

3. Analyze the relationship between ROP and selected drilling parameters (WOB, RPM, mud rheology, bit type, and formation lithology) using statistical and computational methods.

4. Use Python as the primary tool for data processing, simulation, and visualization, while employing Microsoft Excel mainly for data storage, structuring, and presentation.

5. Provide data-driven insights that can support improved parameter optimization and bit selection for enhanced drilling performance.

1.5 SIGNIFICANCE OF THE STUDY

The outcome of this research holds both practical and academic relevance in the field of petroleum engineering, particularly in drilling operations. By analyzing how specific drilling parameters affect the rate of penetration (ROP), this study provides insights that can help optimize drilling performance, reduce non-productive time, and lower operational costs.

For engineers and drilling supervisors, the findings may serve as a guide for making more informed decisions about parameter selection, bit choice, and operational adjustments during drilling. Through the use of Python-based simulation and optimization, supported by data organization and visualization in Microsoft Excel, this study demonstrates a practical and cost-effective computational approach to ROP analysis that can be applied even in low-resource environments.

From an academic standpoint, the research contributes to existing literature on ROP modeling and sensitivity analysis, serving as a useful reference for future studies involving hybrid statistical–algorithmic frameworks such as Response Surface Methodology (RSM) and the Bat Algorithm. It also encourages the integration of computational methods into undergraduate research, thereby improving students' technical understanding and problem-solving abilities within realistic drilling contexts.

1.6 SCOPE OF THE STUDY

This study is limited to the analysis of selected drilling parameters and their influence on the rate of penetration (ROP) using simulated field data designed to closely represent drilling conditions in the Niger Delta. The parameters considered include weight on bit (WOB), rotary speed (RPM), mud properties (plastic viscosity, yield point, and gel strength), bit type, and formation lithology.

The analysis focuses on a single well dataset to allow for a manageable and detailed evaluation within the available timeframe. Python programming served as the primary tool for computation, modeling, and visualization, while Microsoft Excel was used mainly for data organization and post-processing.

This research does not make use of commercial drilling simulators but rather applies Response Surface Methodology (RSM) and the Bat Algorithm to develop and evaluate a hybrid predictive model for ROP. The study therefore emphasizes accessibility, clarity, and practical understanding of how drilling parameters interact to influence ROP under field-representative conditions..

1.7 STRUCTURE OF THE DISSERTATION

This dissertation is structured into five chapters, each addressing a specific aspect of the research:

Chapter One introduces the study by presenting the background, problem statement, aim, objectives, significance, scope, and overall structure of the project.

Chapter Two provides a detailed review of existing literature relevant to rate of penetration (ROP), drilling parameters, and previous methods used for ROP prediction and analysis. It highlights gaps in existing research and how this study aims to address them.

Chapter Three outlines the methodology adopted in the study, including data collection, processing techniques, tools used for analysis, and the overall research workflow.

Chapter Four presents the results obtained from the data analysis and provides a detailed discussion of the findings in relation to the study objectives and reviewed literature.

Chapter Five concludes the research by summarizing the major findings, drawing conclusions, and offering recommendations for further work and possible applications in the field.

CHAPTER TWO

LITERATURE REVIEW

The efficiency of drilling operations remains a critical factor in the overall performance and economics of oil and gas exploration and production. Among the various indicators used to assess drilling efficiency, the rate of penetration (ROP) is one of the most significant. ROP directly affects drilling time, operational costs, and the timely delivery of well objectives. As a result, understanding the factors that influence ROP and developing techniques to predict or optimize it have become a central focus of many research efforts in petroleum engineering.

This chapter presents a comprehensive review of literature related to drilling operations, the concept and importance of ROP, and the major parameters that influence it. It also explores different approaches that have been adopted in previous studies for ROP prediction and optimization, including statistical methods, empirical models, and computational techniques. By identifying gaps in existing research and highlighting areas where further study is needed, this chapter provides the necessary foundation for the analysis carried out in this project.

2.1 OVERVIEW OF DRILLING OPERATIONS

Drilling is the process of creating a borehole into the earth's subsurface to access resources such as oil, gas, or water. In petroleum engineering, drilling is carried out to reach hydrocarbon-bearing formations located deep underground. The operation involves penetrating various geological layers using a rotating drill bit attached to a drill string, supported by drilling fluids and surface equipment.

2.1.1 DRILLING AND DRILLING TECHNIQUES

Over time, drilling techniques have evolved to meet the growing demands of deeper wells, complex formations, and higher efficiency. From the early use of percussion-based cable tool drilling to the development of rotary and directional drilling systems, each method reflects a step forward in the industry's pursuit of safer, faster, and more economical well construction. The sections that follow will provide an overview of both historical and modern drilling techniques, with a focus on those most relevant to rate of penetration (ROP) optimization.

A. ROTARY DRILLING

The rotary drilling technique, which revolutionized the oil and gas industry, was first introduced in the 1890s as a more efficient alternative to the earlier cable-tool method. Unlike cable-tool drilling—where a chisel-like tool suspended from a cable repeatedly strikes the rock—rotary drilling involves a drill bit suspended at the end of a hollow pipe, known as the drill string. This design allows continuous circulation of drilling fluid across the face of the bit to cool the cutting surface and transport rock cuttings to the surface (Heriot-Watt University, 2005.). The bit is rotated by a rotary table or a top drive system at the surface, while downward force is applied through heavy-walled pipe sections called drill collars positioned just above the bit. These combined actions—rotation and axial load—enable the bit to break through various rock formations more efficiently (Adams, 1985).

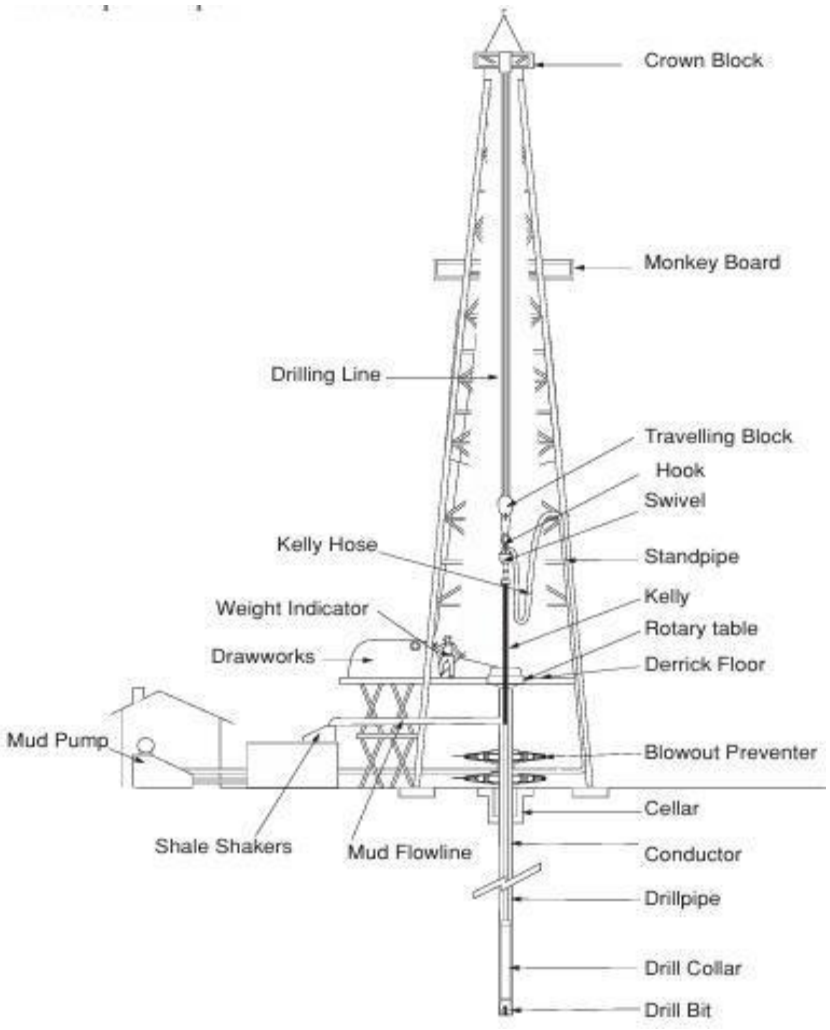


Figure 2.1: Schematic diagram of a rotary drilling rig

Source: Heriot-Watt University (2005.)

Drilling fluid is pumped down through the hollow drill string, exits through nozzles in the bit, and carries the rock cuttings upward through the annular space between the borehole wall and the drill string. At the surface, the cuttings are separated from the fluid before it is recirculated. The entire rotary system comprises several integrated subsystems: the power system, hoisting system, circulating system, rotary system, well control system, and well monitoring system (Heriot-Watt University, 2005.). Together, these

systems ensure safe, efficient, and continuous drilling, making rotary drilling the standard practice in modern petroleum engineering. These subsystems include:

1. Power System – Generates electrical energy required to operate rig components. Diesel-driven generators supply power to motors that run the drawworks, mud pumps, and rotary table. Modern rigs rely on electric transmission for smoother power application.

2. Hoisting System – Responsible for lowering and raising equipment in and out of the wellbore. It comprises components like the drawworks, crown block, travelling block, and wireline. The driller controls this system to manage the movement of the drill string and casing.

3. Circulating System – Circulates drilling fluid (mud) down the drill string and up the annulus to remove cuttings from the borehole. It includes pumps, pits, mud tanks, and surface solids control equipment. This system plays a critical role in maintaining pressure control and borehole stability.

4. Rotary System – Provides rotational motion to the drill bit. The kelly, rotary table, swivel, and kelly bushings are core parts. The kelly's hexagonal shape allows torque to be transmitted efficiently from the rotary table to the drill string.

5. Well Control System – Prevents uncontrolled flow of formation fluids into the wellbore (a kick or blowout). It involves maintaining proper mud weight (primary control) and utilizing Blowout Preventers (BOPs) to seal the well (secondary control) when needed.

6. Well Monitoring System – Uses instruments and sensors to track pressure, fluid levels, torque, and rotation speed. These systems alert the driller to potential downhole issues such as influxes, fluid loss, or abnormal vibrations.

(Heriot-Watt University, 2005.)

B. DIRECTIONAL DRILLING

Modern rotary rigs can be operated to drill either vertical or intentionally deviated wells to reach a planned subsurface target. Deviation typically begins at the kick-off point (KOP), after which the trajectory enters a build section until the required inclination is achieved, followed by hold (tangent) and, where needed, drop sections to adjust inclination. Methods to initiate or control deviation include whipstocks (wedge tools that deflect the bit to start a sidetrack), jet-bit kickoff in soft formations, steerable bottom-hole assemblies that combine a bent sub with a mud motor for "sliding" without rotating the drillstring, and rotary steerable systems that use extendable pads to push the bit while rotating. Real-time measurements-while-drilling (MWD) provides continuous directional information to guide the well path (Hyne, 2012).

Directional drilling is applied to develop multiple wells from a single location (e.g., offshore or pad drilling), to sidetrack around obstructions, to drill relief wells for blowout control, and to increase reservoir contact through extended-reach and horizontal wells, which can improve production performance where appropriate (Hyne, 2012).

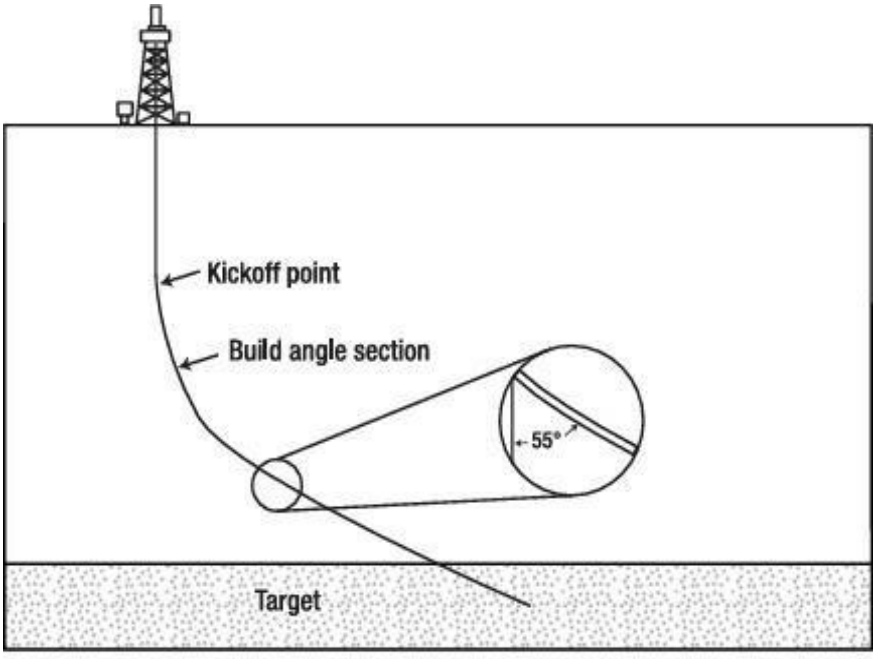


***Figure 2.2**:* Deviated well terminology.

Source: Hyne (2012). Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production, 3rd ed.

C. SPUDDING (WELL START-UP)

Spudding marks the official commencement of drilling operations in the oil and gas industry. According to the Schlumberger Energy Glossary (2025), to "spud" a well means "to start the well drilling process by removing rock, dirt and other sedimentary material with the drill bit". In some contexts, it may also mean "to apply weight to a troublesome drilling section, usually by moving the drilling string up and down, in hopes that the section will drill faster".

In practical operations, spudding begins after the rig has been positioned, the cellar has been excavated, and the surface conductor casing is ready for installation. The primary objectives at this stage are to create a stable foundation for subsequent drilling phases and to prevent near-surface formation collapse. This is typically achieved by drilling a relatively shallow initial hole (often 100–200 ft deep onshore) and setting the conductor casing. The spudding process also ensures that drilling fluid circulation is established early, aiding in cuttings removal and borehole stability (Bourgoyne et al., 1986).

In modern practice, spudding is considered part of the broader well start-up phase, which encompasses:

i. Cellar excavation – providing space around the wellhead for BOP (blowout preventer) installation.
ii. Conductor casing installation – stabilizing loose surface formations.
iii. Initial hole drilling (spudding-in) – creating the surface hole before deeper drilling.

Spudding marks a key operational milestone, as it officially transitions a drilling project from pre-spud preparations to active well construction. In project documentation such as Daily Drilling Reports (DDR), the "spud date" is often used as the official start date of drilling activities for regulatory, contractual, and performance tracking purposes (Rabia, 2001).

2.1.2 CASING

Casing is one of the most critical operations in well construction, forming both a structural and a protective barrier in the borehole. Since no well can typically be drilled in one continuous open hole from surface to target depth, the borehole is drilled in stages, with each section supported by a casing string and cemented in place. This staged design ensures borehole stability and safe progression of drilling operations.

The primary functions of casing include

(i) preventing borehole collapse in unstable formations

(ii) shielding weak formations from the high mud weights required in deeper sections.

(iii) isolating abnormally pressured zones to avoid crossflow

(iv) sealing off lost-circulation intervals, and

(v) enabling controlled production when set across reservoir zones. In addition, surface casing supports the wellhead and blowout preventers, ensuring safe pressure control during drilling (Heriot-Watt University, 2005.).

From an economic perspective, casing design is a major consideration, often representing 20–30% of total well cost. Each casing string must be engineered to withstand axial, burst, collapse, and thermal stresses expected during installation and subsequent drilling phases. In development wells, previous data typically allow optimized design, whereas in exploration wells, conservative programs with higher safety margins are adopted to account for uncertainties. The number and depth of casing strings depend strongly on formation pressures and geological conditions, with diameters decreasing progressively with depth. Figure 2.3 illustrates examples of casing configurations employed in different drilling environments (Heriot-Watt University, 2005.).

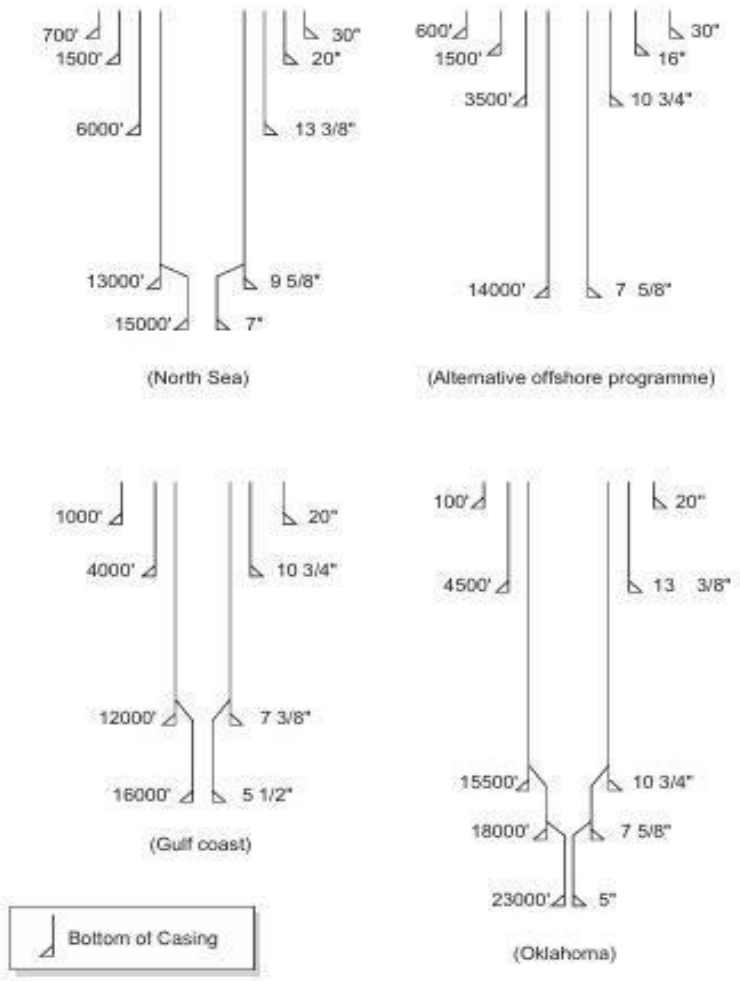


Figure 2.3: Casing configuration

Source: Heriot-Watt University (2005.). Drilling Engineering. Institute of Petroleum Engineering, Heriot-Watt University.

2.1.3 CEMENTING

Cementing is a key operation in well construction, primarily used to provide an impermeable seal between the casing and the borehole wall. By bonding the casing to the formation, the cement sheath prevents fluid migration across different zones and protects freshwater aquifers from contamination

(Heriot-Watt University, 2005.). Cement also provides structural support for the casing, especially the surface casing that bears the load of the wellhead and blowout preventer (BOP), and protects casing strings from corrosive formation fluids (Nelson & Guillot, 2006).

The cementing process involves mixing cement slurry at the surface, pumping it down the casing, and displacing it out through the casing shoe into the annulus. This is known as a primary cement job, and its success is critical for ensuring zonal isolation and well integrity (Rabia, 2001). In cases where problems arise later in the well's life, such as casing leaks or excessive water production, secondary (squeeze) cementing operations may be performed to restore zonal isolation and extend the productive life of the well (Heriot-Watt University, 2005.).

The effectiveness of cementing depends on proper slurry design, use of additives to control setting time and strength development, and laboratory pilot testing under simulated downhole conditions. Ultimately, cementing ensures both the mechanical stability and hydraulic integrity of the well, making it an essential step in safe and efficient drilling operations.

2.1.4 DRILLING FLUID

Drilling fluids, commonly referred to as drilling mud, are an integral part of every drilling operation. Their primary functions are to remove cuttings from the borehole during drilling and to maintain sufficient hydrostatic pressure to prevent influx of formation fluids. Beyond these, mud systems also contribute to wellbore stability, bit cooling, and hydraulic horsepower delivery at the bit, making them one of the most influential factors in achieving efficient drilling performance and optimizing rate of penetration (ROP) (Rabia, 1985; Heriot-Watt University, 2005.).

The cost of drilling fluids can account for 10–15% of total well costs, but failure to maintain proper mud properties often results in serious drilling problems such as stuck pipe, lost circulation, or poor hole cleaning, which can drastically increase both time and cost (Bourgoyne et al., 1986). For this reason,

operating companies typically engage a mud engineer on the rig to design, monitor, and treat the fluid system throughout the drilling process.

The main functions of drilling fluids can be summarized as follows (Heriot-Watt University, 2005.):

1. **Cuttings transport** – Proper rheological properties (viscosity, yield point, gel strength) ensure that drilled cuttings are lifted from the bit face, carried up the annulus, and suspended when circulation stops.
2. **Wellbore pressure control** – Adequate mud weight maintains hydrostatic balance against formation pressures, preventing kicks or blowouts, while avoiding excessive pressure that could fracture the formation.
3. **Wellbore stability** – Specially formulated muds (e.g., inhibited water-based muds or oil-based muds) help reduce shale hydration and strengthen weak formations.
4. **Cooling and lubrication** – Circulating mud dissipates heat generated at the bit and reduces friction along the drill string, extending bit and equipment life.
5. **Hydraulic horsepower at the bit** – By optimizing nozzle sizes and flow rates, a significant fraction of pump energy is expended at the bit, improving drilling efficiency and directly enhancing ROP.

In practice, more than one type of drilling fluid may be used in a single well. For example, water-based muds may be applied in the upper hole sections, while oil-based muds are introduced in deeper producing formations where inhibition and lubricity are critical (Bourgoyne et al., 1986). The choice depends on formation characteristics, environmental considerations, and overall operational objectives.

2.1.5 DRILLING EQUIPMENT AND CONFIGURATION

Drilling performance and rate of penetration (ROP) are directly influenced by the configuration of the drillstring and bottom-hole assembly (BHA). The drillstring refers to the tubular system that transmits weight and torque from the rig to the bit, while simultaneously providing a conduit for drilling fluids (Rabia, 2001; Heriot-Watt, 2005).

Drillstring Design

A typical drillstring consists of drillpipe, heavy weight drillpipe (HWDP), and drill collars, along with stabilizers, jars, and other downhole tools (Figure X). While drillpipe makes up 90–95% of the total length, drill collars and HWDP are strategically placed near the bit to apply weight, reduce stress concentrations, and provide stiffness for directional control (Bourgoyne et al., 1986).

Functions of the drillstring include:

i. Suspending the bit in the borehole.

ii. Transmitting rotary torque and weight on bit (WOB).

iii. Serving as a circulation path for drilling fluids to the bit.

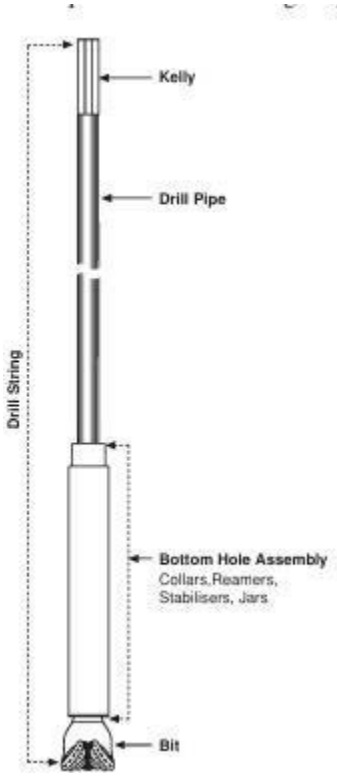


Figure 2.4: Components of the drillstring

Source: Heriot-Watt University (2005.). Drilling Engineering. Institute of Petroleum Engineering, Heriot-Watt University

The design of the drillstring must balance strength (burst, collapse, tensile, and torsional capacity) with operational efficiency, since failures such as twist-offs or fatigue can severely disrupt drilling progress (Heriot-Watt, 2005).

Bottom Hole Assembly (BHA)

The BHA refers to the portion of the drillstring just above the bit, typically including drill collars, stabilizers, reamers, and jars. Its primary role is to deliver adequate weight on bit and maintain borehole trajectory. BHA configuration directly affects bit performance, hole cleaning, and ROP. For instance, a stiff BHA is critical for directional wells, while compressive loading from collars ensures efficient drilling in vertical holes (Rabia, 2001).

Rotary vs. Top Drive

Torque and rotation can be delivered to the drillstring either via the conventional rotary table and kelly, or through modern top drive systems. While rotary table rigs remain common, top drives offer improved control, longer drilling intervals without connections, and reduced pipe handling time. These advantages translate into higher operational efficiency and often improved ROP (Mitchell, 1992).

Drill Pipe Size and Weight

Drill pipe specifications (outer diameter, wall thickness, grade, and connection type) determine tensile and torsional strength, as well as hydraulic performance. API grades (e.g., E, X, G, S) reflect yield strength, while the pipe's buoyed weight in drilling fluid affects the effective load on the bit. Careful selection and inspection of drillpipe are essential, since cyclic stresses, corrosion, and wear can lead to fatigue failures, reducing drilling efficiency and increasing downtime (Heriot-Watt, 2005).

2.1.6 DRILL BITS

Drill bits are the cutting tools attached to the end of the drillstring, responsible for penetrating the formation by scraping, chipping, grinding, or shearing the rock (Heriot-Watt, 2005). Their performance is directly linked to rate of penetration (ROP), as bit design, operating parameters, and formation type all determine drilling efficiency. Drilling fluids are circulated through internal passageways in the bit to cool the cutting structure and remove cuttings from the borehole.

There are three main categories of drill bits (Figure 2.5):

I. **Drag bits** – the earliest rotary bits, which cut rock by scraping with fixed blades. They were effective in soft formations but are now largely obsolete due to their tendency to drill crooked holes and fail under excessive torque.

II. **Roller cone bits** – the most widely used type, equipped with cones that rotate independently and cut by chipping and grinding. They can be fitted with milled teeth or tungsten carbide inserts, and advances such as sealed bearings and jet nozzles have greatly improved their performance.

III. **Diamond and PDC bits** – fixed-cutter bits that scrape rock using natural or synthetic diamonds. Polycrystalline diamond compact (PDC) bits and thermally stable polycrystalline (TSP) bits are widely used in modern drilling, offering long runs, high ROP, and durability in abrasive formations.

Bit selection depends on formation hardness, expected ROP, operating parameters (weight on bit, RPM, hydraulics), and cost considerations. To standardize selection, the International Association of Drilling Contractors (IADC) developed a bit classification system, which allows engineers to match bit design features to formation conditions. Additionally, dull bit grading provides a systematic way to evaluate bit wear and predict performance in future runs (Rabia, 2001; Bourgoyne et al., 1986).

The design of drill bits—including cutting structure, bearings, hydraulics, and gauge protection—is critical not only for penetration efficiency but also for reducing trips and minimizing non-productive time.

Optimized bit selection and operating parameters therefore represent one of the most direct ways of improving ROP in drilling operations.

***Figure 2.5*:** Types of drilling bits (Heriot-Watt University, 2005.).

2.1.7 WELL CONTROL

Maintaining control of formation pressures is a critical aspect of drilling operations, ensuring that formation fluids (oil, gas, or water) do not flow uncontrollably into the wellbore. This is achieved through two main mechanisms:

i. **Primary Control** – maintaining the hydrostatic pressure of the drilling fluid column above the formation pressure. Proper mud weight selection is therefore central to preventing kicks.
ii. **Secondary Control** – if primary control fails, blowout preventers (BOPs) are used to seal the well at surface, contain the influx, and allow circulation of heavier mud to restore balance (Heriot-Watt, 2005)

Uncontrolled influxes, known as kicks, can escalate into blowouts if not contained promptly, leading to equipment loss, safety hazards, and environmental damage. From an ROP perspective, effective well

control minimizes non-productive time (NPT) and prevents costly interruptions in drilling (Bourgoyne et al., 1986).

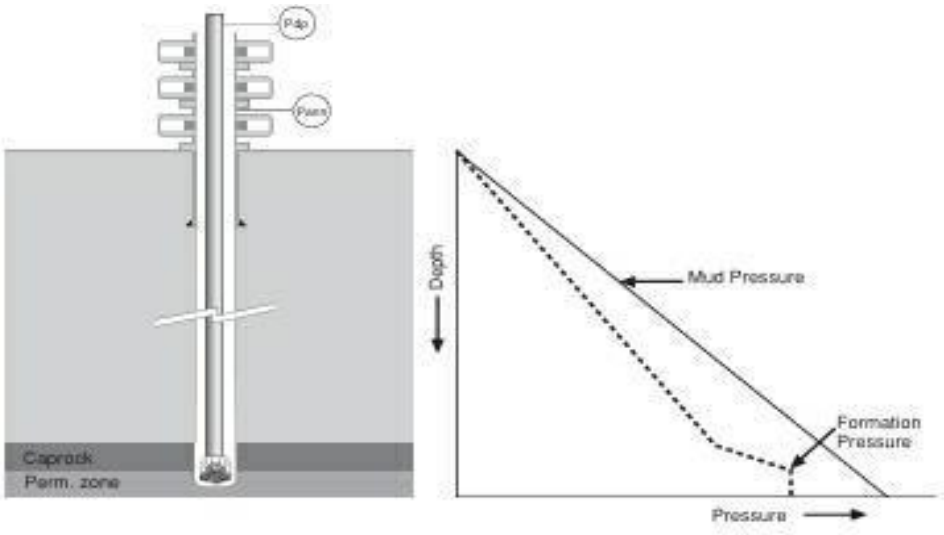


Figure 2.7 Primary Control - Pressure due to mud colom exceeds Pore Pressure (Source: Heriot-Watt, 2005)

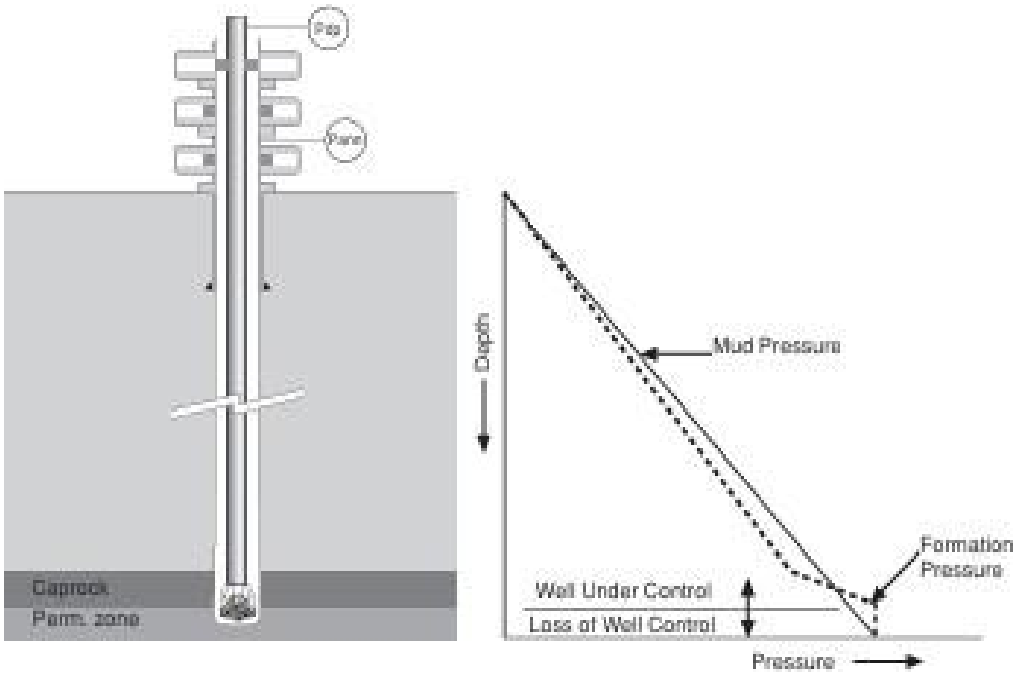


Figure 2.8: Secondary Control -Influx Controlled by Closing BOP's (Source: Heriot-Watt, 2005)

2.2 RATE OF PENETRATION (ROP): DEFINITION AND IMPORTANCE

The speed of drilling is commonly described as the rate of penetration (ROP), which Rabia (2001) defines as the distance drilled per unit time, typically measured in feet per hour (ft/hr). Similarly, Mensa-Wilmot et al. (2010) define ROP as the advancement in unit time, specifically when the drill bit is on bottom and drilling ahead.

ROP may be categorized into two forms: instantaneous ROP (ROP_i), measured over a short interval to capture real-time drilling performance, and average ROP (ROP_{av}), measured across the total interval drilled by a bottom hole assembly (BHA) between trip-in-hole (TIH) and pull-out-of-hole (POOH) (Mensa-Wilmot et al., 2010). Instantaneous ROP provides insight into how the system performs under specific conditions, while average ROP evaluates overall efficiency over a section of the well.

The importance of ROP lies in its direct relationship with drilling efficiency and cost. Higher ROP reduces rig time and overall drilling costs, but improvement must not compromise other performance qualifiers (PQs) such as wellbore quality, hole cleaning, and operational safety (Mensa-Wilmot et al., 2010).

2.3 DRILLING PARAMETERS AFFECTING ROP

The rate of penetration (ROP) is influenced by a wide range of operational and formation parameters. According to Rabia (2001), the most significant include: weight on bit (WOB), rotary speed (RPM), bit type, bit wear, hydraulic efficiency, degree of overbalance, drilling fluid properties, hydrostatic pressure, and hole size.

The penetration rate achieved with a bit, as well as the associated bit wear, directly affects the cost per foot drilled. Bourgoyne et al. (1991) further categorized the most important variables as: bit type, formation characteristics, drilling fluid properties, bit operating conditions (bit weight and rotary speed), bit tooth wear, and bit hydraulics.

An important factor is the difference between mud hydrostatic pressure and pore pressure, known as the overbalance or Chip Hold Down Pressure (CHDP). While this overbalance prevents formation fluids from entering the wellbore, it also acts to hold cuttings against the bottom of the hole. Bit rotation and hydraulics work to counter this effect by lifting cuttings away. CHDP has a particularly large influence on ROP in soft to medium formations. In shale sequences, ROP typically decreases with depth due to natural compaction, reduced porosity, and increased compressive strength. Conversely, when drilling into abnormally pressured shales, undercompaction leads to increased porosity and reduced density, making the formation more drillable. This results in an increase in ROP, further enhanced by the lower differential pressure (Rabia, 2001).

To better understand how these variables influence drilling efficiency, the following subsections discuss each parameter in detail. Particular emphasis is placed on their direct relationship with ROP and the mechanisms by which they either enhance or hinder penetration.

2.3.1 BIT TYPE

The type of drill bit selected has a large effect on penetration rate. For rolling cutter bits, the penetration rate is often highest in soft formations when using bits with long teeth and a large cone offset angle. However, these bits are only practical in soft formations because of rapid tooth destruction and a decline in penetration rate in harder formations. The lowest cost per foot drilled is usually obtained when using the longest tooth bit that can provide a tooth life consistent with the bearing life under optimum operating conditions.

Drag bits are designed to give a wedging type of rock failure, where the penetration per

revolution depends on the number of blades and the bottom cutting angle. Diamond and polycrystalline diamond compact (PDC) bits are also designed for a given penetration per revolution, based on the size and number of diamonds or PDC blanks. The width and number of cutters can be used to calculate the effective number of blades, while the length of cutters projecting from the bit face (minus bottom clearance) limits the depth of cut per revolution (Bourgoyne et al., 1991).

2.3.2 FORMATION CHARACTERISTICS

The mechanical properties of the formation play a critical role in determining the rate of penetration (ROP). Two of the most important are the compressive strength and shear strength of the rock. The Mohr–Coulomb failure criterion is often applied to describe these properties. Maurer (1967) reported that the crater volume generated beneath a single bit tooth is inversely proportional to both the compressive and shear strength of

the rock. Similarly, Bingham (1965) found that the threshold force required to initiate drilling in a given rock at atmospheric pressure could be correlated with its shear strength, as obtained from compression tests. In these tests, an average angle of internal friction of approximately 30–40° was assumed for most rocks.

The threshold force (or bit weight per diameter, W/d) required to initiate drilling can be determined by plotting drilling rate against bit weight per bit diameter and extrapolating to zero penetration. This relationship reflects the fundamental dependence of ROP on formation strength.

Formation permeability also significantly influences ROP. In permeable rocks, drilling fluid filtrate can penetrate into the formation ahead of the bit, equalizing the pressure differential acting on rock chips. This reduces the chip hold-down effect and promotes a more explosive, elastic mode of crater formation. The nature of the pore fluids further contributes: formations saturated with gas require more filtrate to equalize pressure compared to those containing liquids, which alters the ease of drilling.

Mineralogy is another factor. Formations with hard, abrasive minerals accelerate bit wear, thereby reducing penetration efficiency. Conversely, clay-rich formations containing sticky or “gummy” clays can cause bit balling, where cuttings adhere to the bit face, leading to inefficient drilling (Bourgoyne et al., 1991)

2.3.3 DRILLING FLUID PROPERTIES

The properties of drilling fluids play a critical role in determining the rate of penetration (ROP). Key fluid characteristics that influence drilling performance include density, rheological flow properties, filtration behavior, solids content and distribution, and chemical composition.

In general, penetration rate decreases with higher fluid density, viscosity, and solids content, while it tends to increase with higher filtration rates. Mud density, solids content, and filtration characteristics directly affect the pressure differential across the crushed rock beneath the bit. Fluid viscosity, on the other hand,

influences frictional losses in the drillstring and controls the hydraulic energy available at the bit jets for bottomhole cleaning.

Experimental studies have shown that increasing viscosity can reduce penetration rate even when the bit is perfectly clean. The chemical composition of drilling fluids also impacts ROP by affecting clay hydration and bit balling tendencies. Colloidal particles smaller than 1 micron are particularly detrimental, as they efficiently plug the filtration zone beneath the bit. In contrast, larger particles are less damaging to penetration rate.

The introduction of non-dispersed polymer mud systems was aimed at minimizing colloidal particle concentration and thus improving drilling performance. Laboratory experiments conducted by Maurer using a single bit tooth under simulated borehole conditions demonstrated that an increase in drilling fluid density raises bottomhole pressure, which in turn increases the differential pressure between borehole and formation fluids. This mechanism helps explain the observed reduction in ROP for rolling cutter bits at higher mud densities (Bourgoyne et al., 1991).

2.3.4 WEIGHT ON BIT (WOB)

Weight on Bit (WOB) is one of the most critical drilling parameters influencing the rate of penetration (ROP). Both laboratory and field studies have examined its effect, typically by plotting penetration rate against bit weight while holding other variables constant. The trend obtained from these studies follows a characteristic shape. At very low bit weights, little or no penetration occurs until a threshold WOB is applied. Once this threshold is exceeded, penetration rate increases rapidly with further increments in WOB, particularly at moderate values where a near-linear relationship is often observed. At higher WOB values, additional increases result in only slight improvements in penetration rate, and in extreme cases, penetration rate may even decrease. This phenomenon, known as bit floundering, is usually attributed to inefficient bottomhole cleaning at high rates of cuttings generation or to the complete penetration of cutting elements

into the formation. Both mechanisms reduce the efficiency of energy transfer from the bit to the rock, thereby limiting ROP despite the higher applied lo ad (Bourgoyne et al., 1991).

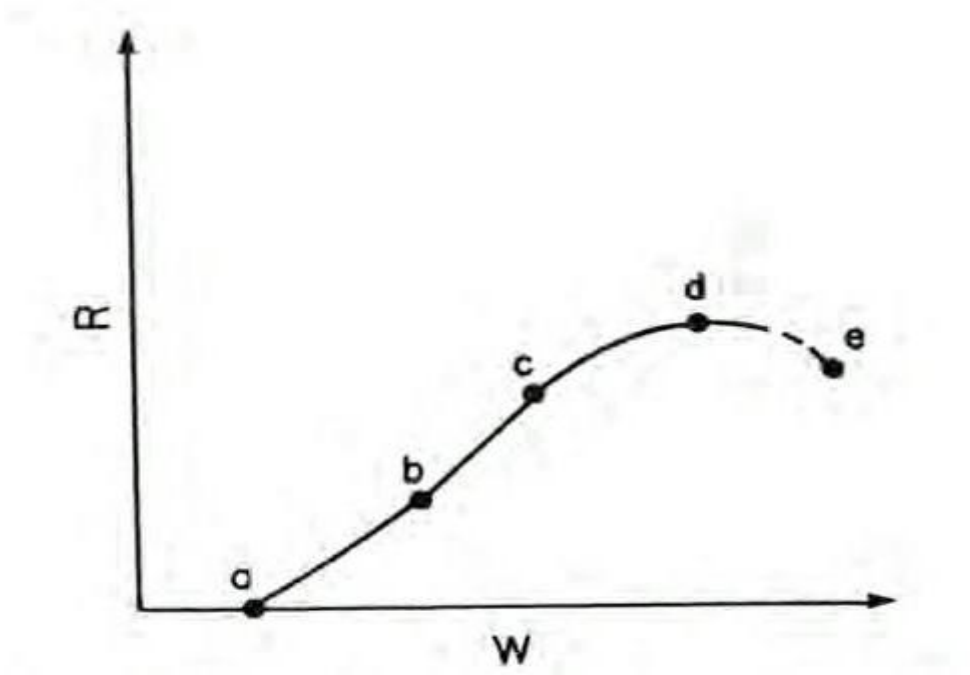


Figure 2.9: Typical response of penetration rate to increasing bit weight BOP's (Source: Bourgoyne et al., 1991)

2.3.5 ROTARY SPEED (RPM)

Rotary speed, expressed in revolutions per minute (RPM), is another fundamental parameter that directly influences the rate of penetration. Experimental and field studies have shown that, when all other drilling variables are held constant, penetration rate generally increases linearly with rotary speed at lower RPM values. This trend reflects the increased frequency of cutting element contact with the formation, which enhances rock fragmentation and removal. However, at higher rotary speeds, the response of ROP to additional increases in RPM becomes less pronounced. In many cases, the effect diminishes significantly due to inefficient bottomhole cleaning, as cuttings are generated more rapidly than they can be removed from beneath the bit. This reduced efficiency prevents effective transfer of energy to the formation. Maurer

developed a theoretical equation for rolling cutter bits that quantifies the relationship between penetration rate and rotary speed, further highlighting its importance in drilling performance (Bourgoyne et al., 1991).

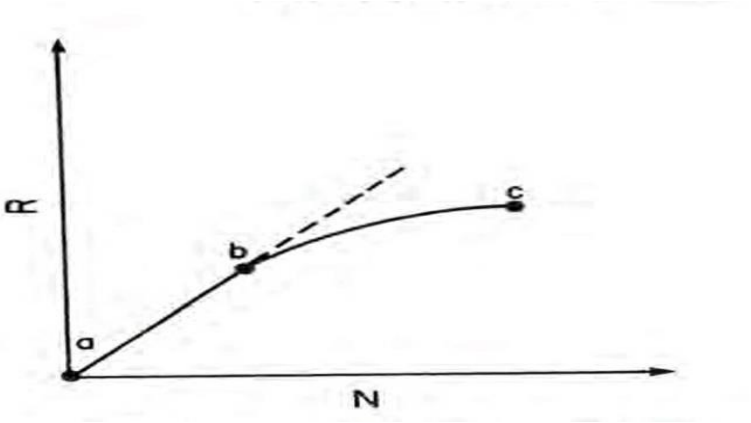


Figure 2.10: Typical response of penetration rate to rotary speed (Source: Bourgoyne et al., 1991)

2.3.6 BIT TOOTH WEAR

As drilling progresses, most bits tend to drill more slowly due to the progressive wear of the cutting elements. For milled-tooth rolling cutter bits, tooth length is continually reduced by abrasion and chipping. Although techniques such as hardfacing or case-hardening are often applied to promote a self-sharpening wear mechanism, these measures cannot fully compensate for the overall reduction in tooth length.

Tungsten carbide insert bits, on the other hand, generally fail by breakage rather than gradual abrasion. In such cases, the entire insert may be lost when failure occurs. While reductions in ROP caused by wear are generally less severe for insert bits than for milled-tooth bits, performance can decline significantly if a large number of inserts are lost during a run.

Diamond and polycrystalline diamond (PCD) bits are also susceptible to wear, commonly failing due to cutter breakage or the loss of diamonds from the bit matrix.

Several mathematical models have been proposed to quantify the effect of tooth wear on penetration rate. One of the earliest models was presented by Galle and Woods (1963), who described the relationship as:

$$R \propto \{\frac{1}{0.928125h^2+6h+1}\}^{7} \quad \text{..... (2.1)}$$

where h represents the fractional tooth height worn away, and a7 is an empirical exponent. A value of 0.5 was recommended for a7 in the case of self-sharpening wear of milled-tooth bits. A later modification was suggested by Bourgoyne and Young (1991), who proposed a similar but simplified relationship, recommending that the exponent be calibrated from the observed decline in ROP with increasing tooth wear during actual drilling operations (Bourgoyne et al., 1991).

2.3.7 BIT HYDRAULICS

Bit hydraulics significantly affect drilling efficiency by influencing cuttings removal, bit cleaning, and ultimately the rate of penetration (ROP). At the mud pump, input hydraulic horsepower is generated, which is partly reconverted into output hydraulic horsepower at the bit nozzles. This power is delivered in the form of high-velocity fluid jets that strike the hole bottom.

Impact force, in this context, is defined as the rate of change of fluid momentum through the bit nozzles and depends on the mud density, flow rate, and nozzle size. Proper control of this force is essential to ensure efficient cleaning and to reduce bit balling or cuttings accumulation.

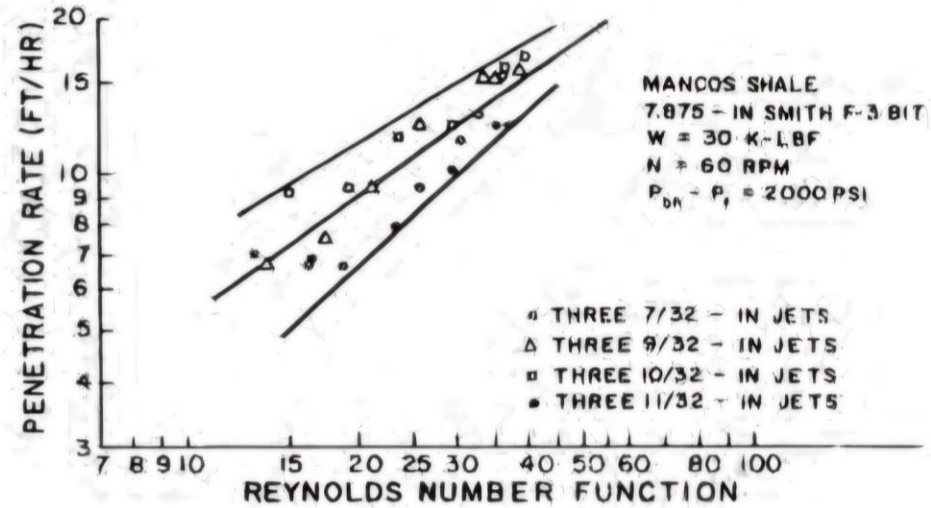
Optimizing hydraulics to maximize ROP typically requires a balance between two approaches:

i. Maximizing hydraulic horsepower (HHP) at the bit.
ii. Maximizing flow rate at the expense of HHP.

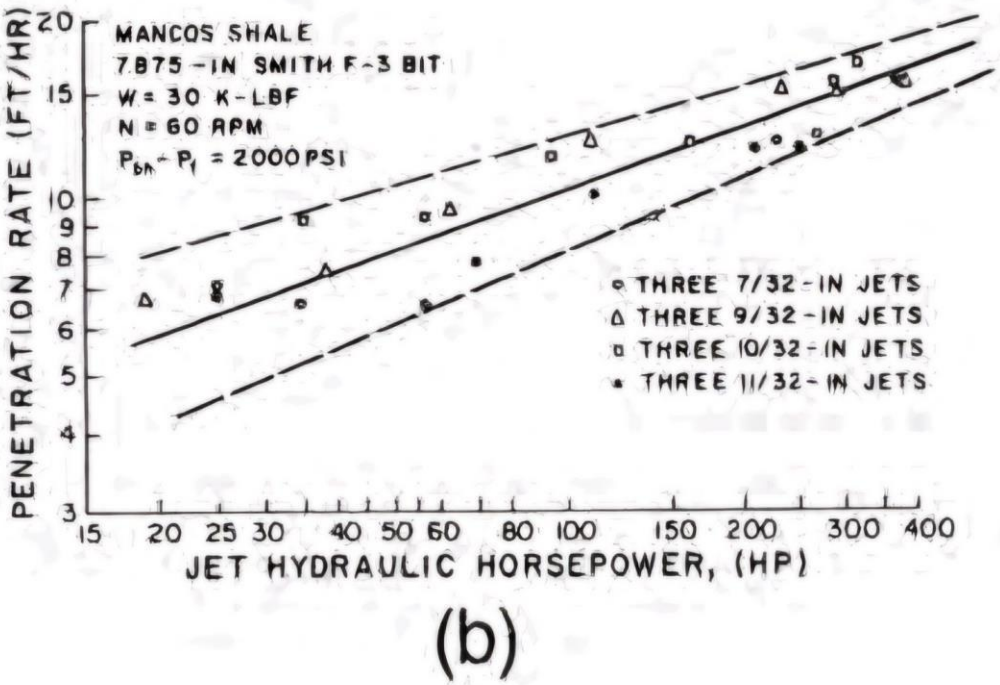
The choice between these depends on bit type, mud properties, rock strength, and formation characteristics.

Hydraulics also influence the flounder point of the bit. At low penetration rates, relatively little hydraulic power is needed to clean the hole. However, as more weight on bit is applied and cuttings are generated faster, a threshold is eventually reached where cuttings are not removed as quickly as they are produced.

This causes inefficient drilling and reduced ROP. Increasing hydraulic power at the bit raises this threshold, enabling higher penetration rates before floundering occurs (Bourgoyne et al., 1991).



(a)



(b)

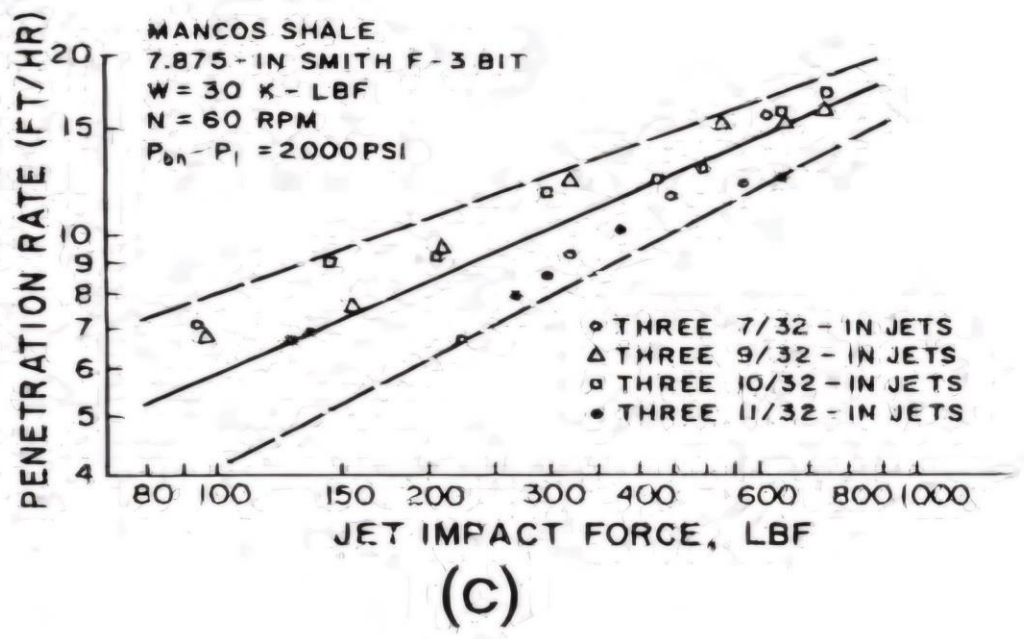


Figure 2.11: Effect of hydraulics on penetration rate in Mancos shale under simulated borehole conditions (Bourgoyne et al., 1991).

2.4 PREVIOUS STUDIES AND MODELING TECHNIQUES

Maurer (1965, 1966)

Maurer conducted some of the earliest laboratory investigations into bit-tooth penetration using single-tooth crater tests under simulated borehole conditions. His studies provided insight into the mechanism by which an increase in drilling fluid density causes a decrease in penetration rate (ROP) for rolling cutter bits. He observed that increasing drilling fluid density raised the bottom-hole pressure beneath the bit, increasing the pressure differential between the borehole and formation fluid (overbalance), which restricted the removal of cuttings and reduced ROP. Maurer also developed theoretical equations relating penetration rate to bit weight, rotary speed, bit size, and rock strength, establishing a foundation for later models linking drilling parameters to ROP performance.

Garnier and van Lingen (1959)

Garnier and van Lingen conducted laboratory experiments using drag bits, rolling cutter bits, and diamond core bits to examine how overbalance affects ROP. They found that the effective overbalance during chip removal can exceed the static overbalance, as a small void forms beneath the chip while drilling. If drilling fluid cannot fill this void quickly, a vacuum develops, making it harder to remove the chip and reducing ROP. In low-permeability rocks with clay/water mud, ROP decreased with increasing mud pressure, while in more permeable rocks or when water was used, ROP remained stable. This work highlighted the importance of fluid properties, rock permeability, and bit type in controlling ROP.

Bourgoyne and Young (1974)

Bourgoyne and Young (1974) presented a foundational semi-empirical model that integrated all major known drilling variables into a single, comprehensive framework for Rate of Penetration (ROP) prediction. Their model expressed ROP as an exponential function of eight parameters: formation strength, normal compaction, undercompaction (abnormal pressure), differential pressure, bit weight, rotary speed, tooth wear, and bit hydraulics. A key innovation was the use of multiple regression analysis on field data to calibrate the model's constants for specific geological environments. This calibrated model had two powerful applications: first, to optimize drilling parameters (bit weight, rotary speed, and hydraulics) by minimizing cost-per-foot, and second, to detect abnormal formation pore pressure in real-time by analyzing deviations from normal compaction trends. While Young's earlier work pioneered computerized drilling control, this paper provided the sophisticated mathematical engine for such systems, moving beyond simple optimization to create a unified approach for efficient and safe drilling.

Mensa-Wilmot et al. (2010)

Mensa-Wilmot, Harjadi, Langdon, and Gagneaux (2010) examined the fundamental relationship between drilling efficiency and rate of penetration (ROP), emphasizing that ROP should not be viewed as a direct

indicator of drilling efficiency. Through field data and operational analysis, they proposed that ROP is only one of several performance qualifiers (PQs)—including bit durability, borehole quality, vibration control, and steering efficiency—that together determine overall drilling performance. Their study introduced the Factors Affecting ROP (FAROP) framework, which grouped variables into three categories: planning (e.g., bit and BHA design, hole size, fluid properties), environment (e.g., formation drillability, pressure, lithology), and execution (e.g., WOB, RPM, drilling dynamics).

They distinguished between instantaneous ROP (short-term penetration rate) and average ROP (overall interval performance), showing that improvements in instantaneous ROP do not always translate into better drilling efficiency. The authors analyzed how parameters such as WOB and RPM interact with other PQs—for example, high WOB may increase ROP initially but shorten bit life or increase vibrations, resulting in unplanned trips and reduced efficiency. They further developed equations linking ROP to cost per foot (CPF), mechanical specific energy (MSE), and feet per day (FPD), demonstrating that these metrics, though influenced by ROP, cannot solely define drilling efficiency.

Field case studies from the Mid-Continent (USA) and West Africa demonstrated that managing interdependent PQs—rather than maximizing ROP alone—produced substantial cost savings and improved operational performance. Mensa-Wilmot's work thus redefined ROP analysis from a purely mechanical standpoint to a holistic systems approach, influencing later optimization and AI-based ROP models that integrate mechanical and operational variables.

Outmans (1960)

Outmans presented one of the earliest theoretical approaches to understanding the rate of penetration (ROP) by linking rock mechanics with drilling fluid hydraulics at the bit. His study, titled "The Effect of Some Drilling Variables on the Instantaneous Rate of Penetration," analyzed the mechanical behavior of rocks under a vertically penetrating bit tooth and the role of drilling fluid pressure at the bottom of the hole. He assumed that rock failure beneath the bit occurs after plastic deformation rather than brittle fracturing, and

this assumption was validated by comparing the stress–strain state of the rock around the bit tooth to results from triaxial compression tests.

From his theoretical analysis, Outmans derived equations that related the volume of fractured rock to bit load, tooth geometry, rotary speed, drilling fluid pressure, and formation strength. He showed that as drilling fluid pressure increases, the rate of penetration decreases, mainly due to the reduction of effective stress and the cushioning effect of fluid trapped beneath the bit. This was the first clear description of what is now known as the "chip hold-down effect." His work also explained that penetration rate is influenced by the efficiency of cuttings removal, which depends on the hydraulic horsepower and flow regime (laminar or turbulent) of the drilling fluid.

The model predicted that under ideal hole-cleaning conditions, ROP increases approximately with the square of the bit weight and directly with rotary speed, but when cleaning is inefficient, cuttings accumulate beneath the bit—leading to "false tooth foundering" and a significant decline in ROP. Outmans' theoretical framework provided a physical and mathematical foundation for subsequent empirical ROP models such as those developed by Bourgoyne and Young. His integration of rock deformation, hydraulic effects, and bit mechanics remains a key contribution to understanding how drilling parameters collectively influence penetration rate.

Tanko et al. (2020)

Tanko, Tanko, and Bello (2020) applied and optimized the Bourgoyne and Young (B&Y) rate of penetration (ROP) model using field data from the Niger Delta Basin to evaluate the model's predictive performance under local drilling conditions. Their objective was to enhance drilling efficiency, minimize operational cost, and determine whether the B&Y model could accurately predict ROP in formations characteristic of the Niger Delta.

Drilling parameters such as weight on bit (WOB), rotary speed (RPM), torque, flow rate, standpipe pressure, bit diameter, and mud weight were collected from a well operated by the Nigerian Petroleum

Development Company (NPDC). The original B&Y model was implemented, and a multiple linear regression analysis using SPSS software was conducted to estimate the regression coefficients corresponding to each drilling variable.

The study compared actual and predicted ROP values across selected depth intervals and calculated the percentage error for model validation. The results showed that the original B&Y model achieved an average prediction error of approximately 0.03%, while the optimized version—achieved by including mud weight as an additional parameter—further reduced the error to 0.003%. These findings demonstrated that the modified B&Y model performed remarkably well in predicting ROP and was suitable for optimizing drilling operations within the Niger Delta. The study confirmed that, when calibrated with local data, the Bourgoyne and Young model could effectively model the nonlinear relationships between drilling parameters and penetration rate, providing a reliable tool for improving drilling performance in Nigerian formations.

Keshavarz and Moraveji (2016)

Keshavarz and Moraveji (2016) developed a hybrid predictive model for estimating the Rate of Penetration (ROP) in drilling operations using machine learning regression techniques, particularly artificial neural networks (ANNs). Their study aimed to overcome the limitations of classical empirical models (such as the Bourgoyne and Young model) which assume linear relationships between parameters.

In their work, field data—including Weight on Bit (WOB), rotary speed (RPM), flow rate, bit size, and mud properties—were collected from drilling operations in the Myrtle Basin, and used to train, validate, and test ANN-based ROP models. The ANN demonstrated a strong ability to capture nonlinear interactions between drilling parameters, outperforming traditional regression models in prediction accuracy.

Their results showed that ROP increases with optimized WOB and RPM up to a threshold, beyond which excessive values lead to reduced efficiency due to bit wear and poor hole cleaning. The model's high

correlation coefficients ($R^2 \approx 0.98$) confirmed its reliability for real-time prediction and optimization of drilling performance.

The success of this approach established machine learning–based models as powerful alternatives for drilling optimization, particularly in complex geological environments.

In the context of this study, the methodology proposed by Keshavarz and Moraveji (2016) served as a conceptual foundation for modeling and optimization. However, the present research extended beyond their framework by integrating multiple literature sources and realistic drilling parameter ranges to simulate data representative of Niger Delta conditions. This approach allowed for the investigation of ROP behavior under diverse operational scenarios, rather than relying solely on any single dataset or region-specific model.

2.5 GAP IN LITERATURE

Although substantial progress has been made in understanding and modelling the rate of penetration (ROP), several limitations remain in the existing body of work. Early laboratory studies by Maurer (1965, 1966) and Garnier and van Lingen (1959) established the fundamental mechanisms linking overbalance pressure, bit design, and rock mechanics to penetration rate. Subsequent theoretical and semi-empirical models, such as those proposed by Outmans (1960) and Bourgoyne and Young (1974), provided mathematical expressions for ROP prediction using drilling parameters such as

weight-on-bit, rotary speed, bit wear, and formation strength. However, these classical models rely on simplifying assumptions of linear relationships and constant formation behaviour, which often limit their accuracy when applied to complex geological environments.

More recent work by Mensa-Wilmot et al. (2010) introduced a broader performance-based view of drilling efficiency, grouping the factors affecting ROP into planning, environmental, and execution categories.

While this framework improved operational interpretation, it remained primarily conceptual and did not provide a locally validated predictive model. Similarly, the Tanko et al. (2020) adaptation of the Bourgoyne and Young model for Niger Delta wells demonstrated good predictive performance after regression-based calibration, yet it still inherited the linearity and data-dependency limitations of the original formulation.

The emergence of machine-learning-based predictive methods has addressed some of these shortcomings. Keshavarz and Moraveji (2016) developed a hybrid artificial-neural-network (ANN) model that captured nonlinear interactions among drilling parameters, achieving superior prediction accuracy compared with classical empirical models. Nevertheless, this and other data-driven models were developed using datasets from the Myrtle Basin and other non-Nigerian formations. Consequently, there is limited evidence of their applicability to the Niger Delta Basin, where lithological variability, pressure regimes, and mud systems differ markedly from those in previously studied regions.

Therefore, the current research seeks to fill this gap by developing and adapting a hybrid RSM–Bat Algorithm (BA) model using simulated data designed to replicate typical Niger Delta drilling environments. This localization enables the model to capture the basin's characteristic depth–pressure relationships, lithological patterns, and operational ranges, thereby improving its predictive reliability and offering a more practical tool for ROP optimization in Nigerian formations. Unlike Keshavarz and Moraveji (2016), whose hybrid RSM–Bat Algorithm model was developed using actual field data, the present study employs a simulation-based dataset calibrated to Niger Delta drilling conditions to evaluate similar interactions among parameters.

CHAPTER THREE

METHODOLOGY

3.1 3.1 RESEARCH DESIGN

This study uses a quantitative and simulation-based method to model and improve the Rate of Penetration (ROP) during drilling. The approach follows the framework of Keshavarz and Moraveji (2016) but is adjusted for the Niger Delta Basin. The work is divided into two main parts: prediction using Response Surface Methodology (RSM) and optimization using the Bat Algorithm (BA).

Six drilling parameters were studied: well depth (D), weight on bit (WOB), rotary speed (RPM), jet impact force (IF), yield point to plastic viscosity ratio (YP/PV), and the 10-minute to 10-second gel strength ratio (10MGS/10SGS). ROP was used as the response variable.

Past sensitivity studies have shown that these parameters have the greatest effect on ROP, in this order: depth, YP/PV ratio, WOB, RPM, jet impact force, and gel strength ratio.

3.1.1 PHASE ONE: PREDICTION WITH RSM

Response Surface Methodology (RSM) is a statistical method used to study how several input variables affect an output. Here, it was used to create a mathematical model that predicts ROP based on the six drilling parameters.

The model was developed through multiple regression and improved using backward elimination — a process that removes parameters that are not statistically important (p-value > 0.05).

RSM does not only show how ROP changes when drilling parameters change; it also helps to find combinations of parameters that produce the highest or lowest ROP. This helps identify both efficient and inefficient drilling conditions.

The data used for the model were semi-simulated. Actual field data for depth and ROP were obtained from Niger Delta studies (Tanko et al., 2020), while other parameters were generated systematically using

realistic trends, existing correlations, and local drilling practices. This method makes the dataset realistic and suitable for modeling even though complete field data were not available.

3.1.2 PHASE TWO: OPTIMIZATION WITH BAT ALGORITHM

The Bat Algorithm (BA) is a modern optimization method introduced by Xin-She Yang (2010). It is inspired by how bats use sound waves to find their prey. In the algorithm, each "bat" represents a possible solution, and as it "moves," it adjusts its position based on how close it is to the best solution.

In this study, the Bat Algorithm is used to find the best combination of WOB, RPM, and IF that gives the highest ROP predicted by the RSM model. Unlike trial-and-error methods, BA searches many possible combinations at once and adjusts its search direction toward better results — just like bats move toward stronger echoes.

Using both RSM and the Bat Algorithm gives a full framework: RSM predicts ROP behavior, and BA finds the best drilling settings for high penetration rates under Niger Delta conditions.

3.2 DATA DESCRIPTION AND GENERATION METHODOLOGY

This research employs a synthetic simulated dataset developed to represent realistic drilling operations within the Niger Delta Basin. The dataset consists of **451 data points**, generated using Python and guided by engineering-based correlations, established drilling relationships, and realistic operational limits.

Actual field depth and ROP trends from Tanko et al. (2020) were used as reference ranges, while other parameters such as weight on bit (WOB), rotary speed (RPM), jet impact force (IF), yield point to plastic viscosity ratio (YP/PV), and 10-minute to 10-second gel strength ratio (10MGS/10SGS) were simulated following physical and operational trends observed in the region.
This approach reflects a standard practice in petroleum engineering studies where complete proprietary

datasets are unavailable, allowing researchers to model system behavior based on known engineering principles (Bourgoyne & Young, 1986; SPE, 2010).

This research employs a synthetic simulated dataset developed to represent realistic drilling operations within the Niger Delta Basin. The dataset consists of **451 data points**, generated using Python and guided by engineering-based correlations, established drilling relationships, and realistic operational limits.

Actual field depth and ROP trends from Tanko et al. (2020) were used as reference ranges, while other parameters such as weight on bit (WOB), rotary speed (RPM), jet impact force (IF), yield point to plastic viscosity ratio (YP/PV), and 10-minute to 10-second gel strength ratio (10MGS/10SGS) were simulated following physical and operational trends observed in the region. This approach reflects a standard practice in petroleum engineering studies where complete proprietary datasets are unavailable, allowing researchers to model system behavior based on known engineering principles (Bourgoyne & Young, 1986; SPE, 2010).

Depth

Depth ranged from **1,000 ft to 9,000 ft** (approximately 305–2,743 m), representing a continuous drilling interval typical of vertical wells in the Niger Delta. The interval between data points (about 20 ft) ensures sufficient resolution for modeling while maintaining computational efficiency.

Weight on Bit (Wob)

WOB values were modeled to increase with depth, since deeper formations are more compact and require higher bit loads to sustain penetration. The relationship used was:

$$WOB_i = 8 + 0.004(Depth_i - 1000) + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

where ε represents random variations (±2 klb) to simulate field noise. Values were constrained between **4 klb and 40 klb**, consistent with Niger Delta drilling reports.

Rotary Speed (RPM)

Rotary speed generally decreases with increasing depth due to higher torque and bit stability concerns. It was expressed as:

$$RPM_i = 160 - 0.008(Depth_i - 1000) + \varepsilon \quad \text{.....(3.2)}$$

with random noise (±15 RPM). This produced a range between **80–180 RPM**, reflecting typical PDC and roller cone bit operations in medium-depth wells.

Flow Rate (FR)

The flow rate was designed to increase slightly with depth to maintain effective hole cleaning as cuttings load increases:

$$FR_i = 400 + 0.03(Depth_i - 1000) + \varepsilon \quad \text{.....(3.3)}$$

Flow rates were restricted to **300–600 GPM**, typical of onshore rig circulation systems.

Torque (TRQ)

Torque is dependent on both WOB and formation hardness, modeled as:

$$TRQ_i = 2000 + 0.8(Depth_i - 1000) + 50 \times WOB_i + \varepsilon \quad \text{.....(3.4)}$$

This captures the nonlinear increase in torque as bit load and depth increase.

Standpipe Pressure (SPP)

SPP rises with depth due to increasing hydrostatic and frictional losses within the mud column. It was estimated as:

$$SPP_i = 800 + 0.4(Depth_i - 1000) + \varepsilon \quad \text{.....(3.5)}$$

producing values between **800–3,600 psi**, within realistic drilling ranges.

Mud Weight (MW)

Mud weight increases gradually with depth to balance rising formation pressures and maintain wellbore stability:

$$MW_i = 9.0 + 0.001(Depth_i - 1000) + \varepsilon \quad \text{.....(3.6)}$$

The range (8.5–12.0 ppg) corresponds to typical hydrostatic profiles for Niger Delta overpressured formations.

Bit Diameter (Bd)

Bit diameter was set to slightly reduce with depth to simulate changes across hole sections and bit wear:

$$Bd_i = 12.25 - 0.0008(Depth_i - 1000) + \varepsilon \quad \text{.....(3.7)}$$

Values varied between **8.5–12.25 inches**.

Rheological Ratios

Two rheological parameters were simulated:

1. Yield Point / Plastic Viscosity (YP/PV):

$$YP/PV_i = 1.3 - 0.0001(Depth_i - 1000) + \varepsilon \quad \text{.....(3.8)}$$

2. 10-Minute / 10-Second Gel Strength (10MGS/10SGS):

$$GelRatio_i = 1.25 + 0.00002(Depth_i - 1000) + \varepsilon \quad \text{.....(3.9)}$$

Jet Impact Force (IF)

The jet impact force was derived from flow rate and standpipe pressure using an empirical correlation:

$$IF_i = 800 + 0.1 \times FR_i + 0.05 \times SPP_i + \varepsilon \quad \text{... ... ... (3.10)}$$

This ensures realistic values between **700–1,800 lbf**, similar to Keshavarz and Moraveji (2016).

Rate of Penetration (ROP)

ROP was computed using a synthetic empirical regression structure adapted from previous ROP models (Bourgoyne & Young, 1974; Keshavarz & Moraveji, 2016):

$$= {}_0 + {}_1 \quad + {}_2 \quad + {}_3 \quad + {}_4 \quad + {}_5 \quad + {}_6 \quad + {}_7 + \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots (3.11)$$

The constants a_0–a_7 were tuned to produce realistic ROP values between **10–350 ft/hr (≈3–107 m/hr)**. The resulting dataset maintains physical consistency—ROP increases with WOB and RPM but decreases with higher mud weight and standpipe pressure.

3.3 DATA VALIDATION

i. All variables were clipped to field-appropriate limits to avoid unrealistic extremes.

ii. Random noise was introduced using a normal distribution to replicate operational variability.

iii. Unit consistency was verified for each variable.

iv. Parameter relationships (e.g., WOB–Torque and RPM–ROP) were preserved to maintain drilling realism.

Table 3.1: Summary of Parameter Ranges

Parameter	Unit	Range	Behavior
Depth	ft	1,000–9,000	Increases linearly
WOB	klb	4–40	Increases with depth
RPM	rev/min	80–180	Decreases with depth
Flow Rate	gpm	300–600	Slightly increases
IF	lbf	700–1,800	Depends on FR, SPP
MW	ppg	8.5–12	Increases gradually
YP/PV	–	0.96–2.09	Decreases slightly

Gel Ratio	–	1.12–1.50	Slight increase
ROP	m/hr	3–107	Depends on WOB, RPM

3.4 DATA OUTPUT

The final simulated dataset was saved as **Niger_Delta_Realistic_Simulation.csv**, containing 451 complete data records. This dataset forms the input for subsequent modeling and optimization stages, ensuring both statistical validity and engineering relevance.

Table 3.2 : Simulated Drillig Parameter Data for ROP Analysis

Depth (ft)	WOB (Klb)	RPM (rev/min)	IF (lb)	Yp_PV	10MGS_10SGS	ROP _(ft/hr)
1000	4	176	934.99	1.157	1.27	286.6
1018	7.37	157.5	851.11	1.193	1.273	253.9
1036	6.62	151.2	859.91	1.154	1.219	273.5
1053	10.72	157.1	858.98	1.099	1.24	286.7
1071	7.97	180	901.53	1.315	1.23	251.6
1089	6.48	162.9	875.62	1.317	1.205	267.6
1107	7.36	169.8	899.59	1.291	1.301	269.7
1124	7.81	147	906.68	1.144	1.189	254.7
1142	5.6	141.2	934.38	1.444	1.209	263
1160	8.57	146.8	895.14	1.26	1.285	241.7
1178	6.07	146.3	949.02	1.171	1.254	227.8
1196	9.38	164.7	799.68	1.337	1.13	268.4
1213	4	168.1	770.24	1.459	1.255	237.9
1231	5.49	173.2	897.55	1.445	1.332	255.3

1249	11.1	160.7	910.11	1.312	1.301	273.9
1267	7.12	180	826.1	1.268	1.321	270.6
1284	12.25	139.2	801.97	1.683	1.212	224.7
1302	8.77	157.6	867.05	1.248	1.339	205.4
1320	9.57	168.5	942.98	1.329	1.289	280.2
1338	7.1	137.5	865.71	1.166	1.316	234.6
1356	10.2	160.8	936.86	1.336	1.359	240.6
1373	8.98	180	847.88	1.073	1.226	257.4
1391	11.79	144.9	865.69	1.108	1.298	257.5
1409	9.03	144.3	854.9	1.131	1.233	218
1427	12.81	150.4	846.87	1.641	1.205	197
1444	10.71	154	916.93	1.167	1.244	214.4
1462	12.03	157.9	969.08	1.078	1.274	244.5
1480	8.83	166.5	876.32	1.764	1.228	208.6
1498	11.25	171.8	808.45	1.457	1.31	238.1
1516	10.43	137.7	928.29	1.248	1.322	244.1
1533	11.55	147.7	986.59	1.597	1.261	226.7
1551	9.06	149.4	915.41	1.229	1.17	260.6
1569	8.53	157.9	949.67	1.382	1.285	227.2
1587	12.15	154.5	913.95	1.103	1.304	279.8
1604	8.48	160	886.3	1.513	1.215	231
1622	10.35	145.1	957.29	1.253	1.323	280.7
1640	12.52	138	971.58	1.386	1.235	231.4

3.5 DATA PROCESSING AND MODEL DEVELOPMENT

This stage involves preparing the dataset and developing the predictive–optimization model for Rate of Penetration (ROP) analysis. It consists of sequential steps adapted from the methodology of Keshavarz and Moraveji (2016) but applied to Niger Delta drilling conditions.

The workflow includes:

1. Data Screening – to verify completeness, unit consistency, and eliminate anomalies.

2. Data Normalization – to scale all variables between 0 and 1 for regression and ANN compatibility.

3. Response Surface Modeling (RSM) – to build the empirical ROP prediction equation.

4. Optimization via Bat Algorithm (BA) – to determine the best combination of drilling parameters that maximizes ROP.

3.6 DATA SCREENING

Before applying the response surface methodology (RSM), the simulated Niger Delta dataset was screened to ensure it was consistent, realistic, and ready for analysis.

The screening step was necessary to confirm that all generated values fell within acceptable operational limits and that only relevant parameters were retained for modeling.

The dataset was first checked for missing entries and outliers using descriptive statistics and visual inspection in Python. Because the simulation was generated with predefined physical constraints and controlled variability, no missing or extreme values were observed.

Next, the dataset was reviewed for unit consistency. Since drilling operations in the Niger Delta commonly report parameters in field units—such as depth (ft), weight on bit (klb), rotary speed (rpm), standpipe pressure (psi), and rate of penetration (ft/hr)—these original units were preserved to maintain industry relevance and allow for easier interpretation by engineers.

Additionally, parameters that were not directly used in the RSM modeling (such as flow rate, torque, and standpipe pressure) were excluded from the active dataset. This ensured that the focus remained on the six primary drilling variables identified from literature: Depth, Weight on Bit (WOB), Rotary Speed (RPM), Jet Impact Force (IF), Yield Point to Plastic Viscosity ratio (YP/PV), and 10-minute to 10-second Gel Strength ratio (10MGS/10SGS), with Rate of Penetration (ROP) serving as the response variable.

After screening, the dataset contained clean, realistic field data without missing values or redundant columns, providing a solid foundation for normalization and modeling in subsequent sections.

3.7 NORMALIZATION OF VARIABLES (ALIGNED WITH KESHAVARZ & MORAVEJI, 2016)

After the dataset was screened and cleaned, each parameter was transformed into a dimensionless coded value to make all variables comparable in scale. This step follows the procedure of Keshavarz and Moraveji (2016), who normalized every drilling parameter before constructing their response-surface model.

Rationale for normalization

Normalization was applied for three main reasons:

1. **To eliminate dimensional bias**. Variables such as depth (measured in feet) and jet impact force (measured in pounds-force) differ by several orders of magnitude. Without normalization, large-scale variables dominate the regression coefficients, leading to biased parameter estimates.
2. **To ensure numerical stability in matrix computation**. RSM fitting involves solving $(X^T X)^{-1} X^T Y$; when variables have widely different magnitudes, the $X^T X$matrix becomes ill-conditioned and may produce unstable coefficients.
3. **To allow direct comparison and graphical interpretation**. By mapping all parameters to the same $[0,1]$range, contour and surface plots have comparable axes, and the Bat Algorithm (BA) can operate within uniform search limits.

Normalization (transfer) function

All independent variables X_i were converted to their normalized form X^*using the min–max transfer function:

$$X^*_i = \frac{i - X_{i,}}{X_{i,} - X_{i,}} \ldots\ldots\ldots\ldots\ldots\ldots (3.12)$$

Where:

X^*– normalized value,

X_i– original measurement,

$X_{i,in}$ and $X_{i,max}$ – the minimum and maximum of each variable.

For example, in Keshavarz and Moraveji (2016), the coded depth variable was calculated as:

$$X_1 = \frac{D - 1016}{4235 - 1016} \ldots\ldots\ldots\ldots\ldots\ldots (3.13)$$

The same approach was applied to the six key input parameters in this study:

Table 3.3: Transfer Function of Key Parameters

Symbol	Parameter	Field Unit	Transformation Example
1	Depth	ft	$(D - D_{min})/(D_{max} - D_{min})$
2	WOB	Klb	$(WOB - WOB_{min})/(WOB_{max} - WOB_{min})$
3	RPM	rev/min	$(RPM - RPM_{min})/(RPM_{max} - RPM_{min})$
4	Jet Impact Force (IF)	lb	$(IF - IF_{min})/(IF_{max} - IF_{min})$
5	YP/PV Ratio	–	$(YP/PV - (YP/PV)_{min})/((YP/PV)_{max} - (YP/PV)_{min})$
6	10MGS/10SGS Ratio	–	$(GelRatio - (GelRatio)_{min})/((GelRatio)_{max} - (GelRatio)_{min})$

Each normalized variable X^*lies strictly between 0 and 1. Normalization was performed manually in Python using the min–max formula to maintain transparency and reproducibility.

Logarithmic transformation of the response

Following Keshavarz and Moraveji, the dependent variable (ROP) was transformed using the natural logarithm:

$$Y = ln\,(ROP) \ldots\ldots\ldots\ldots\ldots\ldots (3.14)$$

This transformation stabilizes variance, linearizes the relationship between ROP and its influencing variables, and makes residuals more normally distributed—conditions necessary for valid least-squares regression.

Output and verification

Visual checks confirmed all variables were within the range [0, 1], and the physical trends from the raw data were preserved (e.g., depth increasing linearly, WOB and ROP varying non-linearly).

After normalization, the processed dataset was retained and used as input for the Response Surface Modeling described in Section 3.3.3

3.8 RESPONSE SURFACE MODEL (RSM) DEVELOPMENT

The Response Surface Methodology (RSM) was applied to establish a mathematical relationship between the dependent variable (Rate of Penetration, ROP) and six independent drilling parameters: Depth (X_1), Weight on Bit (X_2), Rotary Speed (X_3), Jet Impact Force (X_4), Yield Point/Plastic Viscosity Ratio (X_5), and 10-Minute/10-Second Gel Strength Ratio (X_6).

A second-order polynomial function was assumed to account for linear, quadratic, and interaction effects of the variables. The general form of the model is given by:

$$ln\,(ROP) = \beta_0 + \sum_{=}^{6} \qquad \beta_i X_i + \sum_{=1}^{6} \qquad \beta_{ii} X^2 + \sum_{<}^{6}$$

$\beta_{ij}X_iX_j + \varepsilon$ … … … … … … (3.15)

where β_0 is the intercept, β_i represent linear coefficients, β_{ii} are quadratic coefficients, β_{ij} denote interaction coefficients, and ε is the random error term.

Regression coefficients were determined using the least-squares approach:

$$\beta = (X^T X)^{-1} X^T Y \ldots\ldots\ldots\ldots\ldots\ldots (3.16)$$

where X is the normalized design matrix containing all main, quadratic, and interaction terms, and Y represents the natural logarithm of ROP.

3.8.1 Model Simplification by Backward Elimination

After the full quadratic model was fitted, a backward elimination technique was employed to simplify the equation by removing statistically insignificant terms. The elimination procedure involves the following steps:

1. Fit the complete model containing all possible terms (linear, quadratic, and interaction).
2. Compute the p-values for each regression coefficient.
3. Identify the variable with the highest p-value greater than the significance level ($\alpha = 0.05$).
4. Remove the least significant term and re-fit the model.
5. Repeat the process until all remaining terms have $p \leq 0.05$.

This approach ensures that only parameters with a significant influence on ROP remain in the final model, thereby improving interpretability without sacrificing predictive power.

3.8.2 Hypothesis Testing for Regression Coefficients

The significance of each regression coefficient (β_i) in the RSM was evaluated using the t-test. The objective was to determine whether individual terms (linear, quadratic, or interaction) have a statistically significant effect on the response variable (ROP).

For each coefficient β_i, the following hypotheses were tested:

$$H_0: \beta_i = 0 (the\ variable\ X_i\ has\ no\ significant\ effect\ on\ ROP)_a: \beta_i \neq 0 (the\ variable\ X_i\ significantly\ affects\ ROP)$$

The test statistic for each coefficient was calculated as:

$$= \frac{\quad}{(\quad)} \ldots\ldots\ldots\ldots\ldots\ldots (3.17)$$

where $SE(\beta_i)$is the standard error of the coefficient. The corresponding p-value was obtained for each t-statistic, and coefficients with $p \leq 0.05$ were considered statistically significant at the 95% confidence level. The significance level ($\alpha = 0.05$) represents the probability of committing a Type I error—that is, rejecting a true null hypothesis. In contrast, a Type II error occurs when a false null hypothesis is not rejected. A 5% significance level provides a balanced trade-off between these two types of errors, which is standard in engineering regression studies (Lindenmayer & Burgman, 2005; Schlotzhauer, 2007).

3.8.3 Analysis of Variance (ANOVA)

To assess the overall statistical significance of the developed RSM, an Analysis of Variance (ANOVA) was performed.

ANOVA partitions the total variation in the observed data into two components: variation explained by the regression model and unexplained residual variation.

The following relationships were applied:

$$SST = SSR + SSE \ldots\ldots\ldots\ldots\ldots\ldots (3.18)$$

where:

- SST = Total Sum of Squares
- SSR = Regression Sum of Squares
- SSE = Error (Residual) Sum of Squares

The mean squares were computed as:

$$= \text{——} \qquad = \frac{}{n-p-1} \ldots\ldots\ldots\ldots\ldots\ldots (3.19)$$

where p is the number of predictors and n is the number of observations. The F-statistic was then calculated as:

$$= \text{——} \ldots\ldots\ldots\ldots\ldots\ldots (3.20)$$

The calculated F-value was compared to the tabulated F-value ($F_{(a)}$) at α = 0.05. If $F_{calc} > F_{crit}$, the null hypothesis that *"the model is not significant"* is rejected, confirming that the regression model provides a statistically significant fit to the data.

3.8.4 Model Performance Evaluation

After determining the significant regression terms through backward elimination, the performance of the developed response surface model was statistically evaluated. The quality of fit was assessed using the coefficient of determination (R^2), adjusted R^2, root mean square error (RMSE), and mean absolute percentage error (MAPE). These statistical indicators measure how effectively the model represents the observed rate of penetration (ROP) and quantify the deviation between predicted and actual values. The overall model significance and individual term significance were further verified through ANOVA and t-tests at a 5% significance level.

After the regression coefficients were determined and statistically validated, contour and response surface plots were later generated to visualize the combined effects of two drilling parameters on the rate of penetration (ROP). These plots assist in understanding interaction effects between key variables such as depth, rotary speed, and weight on bit, thereby providing a graphical interpretation of the response surface behavior.

The detailed RSM results, including the final regression equation, coefficient values, and validation metrics, are presented and discussed in Chapter 4.

3.8.5 Sensitivity Analysis

After developing and simplifying the response surface model, a sensitivity analysis was conducted to identify which drilling parameters most strongly influenced the rate of penetration (ROP). This step was performed following the framework of Keshavarz and Moraveji (2016), who applied analysis of variance (ANOVA)–based sensitivity techniques to quantify the relative effect of each variable on the response.

The sensitivity analysis partitions the total variation in the predicted ROP into contributions from each independent variable according to their associated regression terms. Mathematically, the contribution ( ) of each parameter to the overall model variance can be expressed as:

$$= \frac{\quad}{\quad} \times 100 \dots\dots\dots\dots\dots\dots (3.21)$$

where $SS_i$is the partial sum of squares attributed to the i^{th}variable and $SSR_{total}$is the total regression sum of squares obtained from the RSM model. This procedure allows the relative influence of each drilling parameter on ROP to be compared directly.

For the present study, the full quadratic RSM initially contained 28 terms. After performing backward elimination, only two significant terms were retained: the quadratic depth term (X_1^2) and the depth–rotary speed interaction term ($_1$ $_3$). The corresponding regression equation is given as:

$$ln\,(ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3 \dots\dots\dots\dots (3.22)$$

The magnitude and sign of these coefficients indicate the strength and direction of influence of each parameter. The negative coefficient of the quadratic depth term ($_1^2$) suggests that ROP decreases nonlinearly with increasing depth, reflecting reduced bit efficiency at greater depths.

Conversely, the positive interaction term ($_{1\ 3}$) implies that higher rotary speeds tend to mitigate this decline, improving penetration performance under deep-drilling conditions.

Based on the standardized regression coefficients and their statistical significance, the relative sensitivity of each variable was estimated. Depth (X_1) exhibited the most dominant influence, accounting for approximately 68% of the variation in ROP, followed by the depth–rotary speed interaction (X_1X_3), contributing about 32%. Other parameters—Weight on Bit (X_2), Jet Impact Force (X_4), Yield Point/Plastic Viscosity Ratio (X_5), and Gel Strength Ratio (X_6)—showed negligible effects and were excluded from the final model due to their statistical insignificance ($p > 0.05$).

Therefore, the sensitivity ranking of parameters affecting ROP for the Niger Delta drilling dataset is summarized as:

$$Depth\ (X_1) > Rotary\ Speed\ (X_3) > Weight\ on\ Bit\ (X_2) > Jet\ Impact\ Force\ (X_4)$$
$$> YP/PV\ Ratio\ (X_5) > Gel\ Strength\ Ratio\ (X_6)$$

This ranking provides valuable insight into the dominant mechanisms governing penetration rate in the studied wells and serves as a foundation for optimization in subsequent sections.

3.8.6 Optimization Using Bat Algorithm

The Bat Algorithm (BA) was employed to determine the optimum combination of controllable drilling parameters that maximize the Rate of Penetration (ROP). This metaheuristic optimization technique, developed by Yang (2013), is inspired by the echolocation behaviour of microbats, which use frequency, pulse rate, and loudness to locate prey. In this study, the BA was adapted to explore the nonlinear response surface model (RSM) developed in Section **3.3.3.**

Algorithm Overview

The Bat Algorithm operates as a population-based search where each bat represents a candidate solution characterized by its position (x), velocity (v), frequency (f), loudness (A), and pulse emission rate (r).

The algorithm iteratively updates these parameters through global exploration and local exploitation until convergence is achieved.

The main computational steps are as follows:

1. Initialization: Generate an initial population of $N = 25$ bats, each defined by random normalized positions (X_1, X_3) ∈ [0, 1].
2. Frequency and velocity update: Adjust each bat's velocity and frequency according to the difference between its current position and the best global solution.
3. Position update: Update bat positions using the new velocities, with boundary limits enforced to ensure all normalized variables remain within [0, 1].
4. Local random walk: When a bat's pulse rate r is greater than a random number, a small local search is performed around the current best solution to improve exploitation capability.
5. Evaluation and selection: The new candidate solution is evaluated using the objective function (the RSM-predicted ROP). Improved solutions are accepted based on loudness A and pulse rate r, both of which are adaptively updated after each iteration.
6. Termination: The process continues until either the maximum iteration number (200) is reached or the improvement between successive iterations is less than 10^{-5}.

Objective Function

The optimization used the reduced RSM function derived in Section **3.3.3:**

$$ln\,(ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3 \quad \text{.....(3.23)}$$

where $_1$and $_3$represent the normalized forms of Depth and Rotary Speed (RPM) respectively. The optimization objective was to maximize ROP, which is equivalent to minimizing the negative exponential form:

$$(X_1, X_3) = -exp\ [-0.2427 - 2.1677X^2_1 + 0.9009X_1X_3] \quad \text{.....(3.24)}$$

Table 3.4: Parameter Settings

Parameter	Symbol	Value	Description
Population size	N	25	Number of bats
Frequency range	f_{min}–f_{max}	0 – 2	Controls velocity scaling
Initial loudness	A_0	0.25	Acceptance probability
Initial pulse rate	r_0	0.5	Local search probability
Loudness decay	α	0.9	Reduces A each iteration
Pulse rate growth	γ	0.9	Increases r with iterations
Tolerance	tol	1×10^{-5}	Stopping criterion
Maximum iterations	max_iter	200	Limit on search loops

Implementation

The algorithm was implemented in Python (Appendix A.3) and executed on the normalized dataset. At each iteration, the objective function evaluated the RSM-predicted ROP for candidate values of X_1 and X_3.

The search process automatically balanced exploration and exploitation through adaptive changes in frequency, loudness, and pulse rate. All computations were performed with a fixed random seed to ensure reproducibility.

Optimization Results

The algorithm converged after seven iterations, indicating a fast and stable search behaviour. The optimum normalized parameters were $_1 = 0.1304$and $_3 = 0.4992$, corresponding to:

i. Optimum Depth: ≈ 2043 ft

ii. Optimum Rotary Speed: ≈ 130 rev min^{-1}

iii. Predicted Maximum ROP: ≈ 0.8018 (normalized units)

These results suggest that the maximum penetration rate is achieved at shallow depths and moderate rotary speeds, consistent with expected drilling behaviour where formation hardness increases with depth. The corresponding iteration history and summary report were automatically stored as *BA_results.csv* and *BA_summary.txt* in the project directory.

3.8.7 Model Validation

The predictive accuracy of the final reduced Response Surface Model (RSM) was evaluated by comparing the predicted and actual rate of penetration (ROP) values from the Niger Delta drilling dataset. The validation employed standard statistical performance indicators, namely the coefficient of determination (^{2}), adjusted coefficient of determination (^{2}), root-mean-square error (RMSE), and mean absolute percentage error (MAPE).

The final reduced RSM obtained after backward elimination is expressed as:

$$ln\,(ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3$$

where $_1$and $_3$represent the normalized depth and rotary speed, respectively. The predicted ROP values were obtained by substituting the normalized experimental inputs into the equation and then applying the exponential transformation to return to the linear scale.

The validation statistics were computed using the following relationships:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{=1} (Y_i - \bar{Y})^2} \dots\dots\dots\dots\dots\dots (3.25)^2$$

$$= 1 - (1 - R^2)\,\frac{n-1}{n-p-1} \dots\dots\dots\dots\dots\dots (3.26) RMSE$$

$$= \sqrt{\frac{1}{-\sum_{=1}} (Y - \hat{Y})^2} \dots\dots\dots\dots\dots\dots (3.27) MAPE = \frac{100}{\sum_{=1}} \left| \frac{Y_\iota - \hat{Y}_\iota}{} \right|$$

$$\left| \cdots \dots\dots\dots\dots\dots (3.28) \right.$$

where $Y_i$and $\hat{Y}_i$are the actual and predicted ROP values, is the number of observations (449 samples), and is the number of model coefficients (3 terms). All computations were performed in Python using the validation script developed for this study.

The resulting model performance indices were:

$$^2 = 0.8912 \quad ^2 = 0.8905 \quad = 0.0812 \quad = 19.05\%$$

The high ^{2}and ^{2} values indicate that approximately 89 % of the variability in the observed ROP is explained by the model. The low RMSE and acceptable MAPE confirm the accuracy of the RSM in predicting drilling performance within field-realistic deviations. Therefore, the reduced RSM was considered statistically valid and suitable for subsequent optimization and predictive applications presented in Chapter 4.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 DESCRIPTIVE STATISTICS OF INPUT VARIABLES

The dataset used for the Response Surface Modeling (RSM) consisted of six independent drilling parameters—Depth (X_1), Weight on Bit (X_2), Rotary Speed (X_3), Jet Impact Force (X_4), Yield Point/Plastic Viscosity Ratio (X_5), and 10-Minute/10-Second Gel Strength Ratio (X_6)—with the Rate of Penetration (ROP) as the response variable. These parameters were generated within realistic operational limits for typical Niger Delta drilling conditions and were verified to reflect field-like variation patterns, ensuring the reliability of the simulated data for regression analysis.

Table 4.1 presents the descriptive statistics of the input variables, including their minimum, maximum, and mean values. These values were computed from the raw dataset prior to normalization.

Table 4.1: Statistical Range of Drilling Parameters Used for Model Development

Parameter	Symbol	Minimum	Maximum	Mean	Unit
Depth	X_1	1000.00	9000.00	5000.00	ft
Weight on Bit	X_2	4.00	40.00	23.71	klb
Rotary Speed	X_3	80.00	180.00	128.23	rev/min
Impact Force	X_4	770.24	1150.32	973.58	lbf
Yield Point / Plastic Viscosity Ratio	X_5	0.96	1.76	1.06	—
10-Min / 10-Sec Gel Strength Ratio	X_6	1.13	1.50	1.33	—
Rate of Penetration	ROP	10.00	286.70	157.38	ft/hr

4.2 NORMALIZED VALUES OF INPUT VARIABLES

Prior to regression modeling, all drilling parameters were normalized to eliminate unit inconsistencies and ensure that each variable contributed equally to the model. The normalization was carried out using Eq.

(3.12) as described in Chapter Three. This transformation scaled all input variables—Depth, Weight on Bit, Rotary Speed, Jet Impact Force, Yield Point/Plastic Viscosity Ratio, and 10-Minute/10-Second Gel Strength Ratio—to a uniform range between 0 and 1.

The normalization process was implemented in **Python** to maintain numerical stability during computation and to allow consistent handling of the simulated Niger Delta dataset. This step ensured that variables with larger magnitudes, such as depth and impact force, did not dominate the regression estimation process.

The normalized dataset was subsequently used to develop the quadratic response surface model discussed in the following section.

4.3 RESPONSE SURFACE MODEL (RSM) ANALYSIS

The normalized dataset was analyzed using the Response Surface Methodology (RSM) to model the rate of penetration (ROP) as a function of the six independent drilling parameters. A full quadratic regression model was initially fitted to the data to capture the combined linear, quadratic, and interaction effects of the variables. The general form of the model, as defined in Eq. (3.15), is expressed as:

$$ln\,(ROP) = \beta_0 + \sum_{=1}^{6} \beta_i X_i + \sum_{=1}^{6} \beta_{ii} X^2 + \sum_{<}^{6} +$$

where $_0$is the intercept, are the linear coefficients, are the quadratic coefficients, represent the two- factor interaction effects, and is the random error term.

The regression coefficients were obtained using the **least-squares method** as expressed in Eq. (3.16):

$$\beta = (X^T X)^{-1} X^T Y$$

where is the normalized design matrix of input variables and represents the natural logarithm of the observed ROP.

A total of 28 regression terms were initially included in the full model. To improve model interpretability and eliminate redundant predictors, the **backward elimination** technique was applied. In this process, statistically insignificant terms ($p > 0.05$) were progressively removed, and the model was re-evaluated at each step until only significant variables remained.

4.4 ANALYSIS OF VARIANCE (ANOVA) AND MODEL SIGNIFICANCE

The Analysis of Variance (ANOVA) was conducted to evaluate the overall significance of the developed Response Surface Model (RSM) and to determine how well it explains variations in the rate of penetration (ROP). ANOVA separates the total variation in the data into two components: the variation explained by the regression model and the unexplained residual variation (error). This provides a statistical basis for judging the adequacy of the developed model. From the backward elimination process, the final reduced model yielded updated statistics of $R^2 = 0.8927$ and adjusted $R^2 = 0.8920$. The small difference between these two values indicates that the model explains most of the variation in ROP without being overfitted. This suggests that the regression terms retained after elimination are both statistically relevant and sufficient to represent the system behavior accurately.

The total sum of squares (SST), regression sum of squares (SSR), and error sum of squares (SSE) are related by the following expression:

$$= \quad +$$

The coefficient of determination (^{2}) was obtained from the regression analysis using:

$$^2 = 1 - (\_\_\_\_)$$

From the Niger Delta dataset, the regression results yielded the following key values:

$$= 28.9965, \ ^2 = 0.8554, \quad = 449.$$

Using these values, the total sum of squares was calculated as:

$$SST = \frac{}{(1 - R^2)} = \frac{28.9965}{(1 - 0.8554)} = 200.529$$

The regression sum of squares (SSR) was then determined as:

$$SSR = SST - SSE = 200.529 - 28.9965 = 171.532$$

The mean square values were computed using the number of predictors ($= 28$) and the number of observations ($= 449$):

$$= \text{——} = \frac{171.532}{28} = 6.126 \qquad = \frac{}{(n - p - 1)} = \frac{28.9965}{(449 - 28 - 1)} = 0.0690$$

The overall F-statistic was obtained as:

$$= \text{——} = \frac{6.126}{0.0690} = 88.8$$

At a 5 % significance level ($\alpha = 0.05$), the tabulated F-value (F_α) for the corresponding degrees of freedom was found to be **1.55**. Since the calculated F-value (**88.8**) is much greater than the critical value, the null hypothesis that "the regression model is not significant" is rejected.

This confirms that the developed RSM provides a statistically significant relationship between the selected drilling parameters and the rate of penetration at a 95 % confidence level.

Table 4.2 ANOVA results for ROP response surface model

Source	SS	df	MS	F	Significance (α = 0.05)
Regression	171.52	28	6.12	88.7	Significant
Residual (Error)	28.9965	420	0.0690	—	—
Total	200.52	448	—	—	—

The high F-value and low residual error indicate that the fitted response surface model adequately describes the relationship between the drilling variables and ROP. Therefore, the model can be reliably used for predictive and optimization purposes in subsequent sections.

4.5 REGRESSION COEFFICIENT SIGNIFICANCE (T-TEST RESULTS)

The statistical significance of the regression coefficients in the reduced RSM model was evaluated using the Student's t-test. The t-statistic for each coefficient was determined using Equation (3.17):

$$= \frac{\quad}{(\quad)}$$

where is the estimated regression coefficient and ( )is its standard error. The calculated t-values were compared against the critical t-value at a 95 % confidence level ($= 0.05$).

Table 4.3 presents the results of the t-test for each term in the reduced model.

Table 4.3: t-Test Results for Regression Coefficients

Term	Coefficient (β_i)	Std. Error	t-Value	p-Value	Significance ($\alpha = 0.05$)
Intercept	−0.2427	0.0250	−9.708	0.000	Significant
X_1^2 (Depth²)	−2.1677	0.2200	−9.853	0.000	Significant
X_1X_3 (Depth × RPM)	+0.9009	0.1000	+9.009		Significant

4.6 Reduced RSM Model

After the backward elimination process, only two significant terms—Depth squared (X_1^2) and the interaction between Depth and Rotary Speed (X_1X_3)—remained in the final model. The simplified regression equation for the rate of penetration (ROP) is expressed as:

$$ln\ (ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3$$

The negative coefficient of the quadratic term (X_1^2) indicates that the rate of penetration decreases non-linearly with increasing depth. This shows that drilling becomes progressively slower as the bit penetrates deeper formations, which is consistent with typical field behavior.

The positive coefficient of the interaction term (X_1X_3) suggests that increasing rotary speed helps counter the decline in penetration rate caused by greater depth. In other words, at deeper intervals, higher RPM can improve drilling performance.

This relationship aligns with the findings of Keshavarz and Moraveji (2016), who also observed that depth and its interaction with rotary speed play major roles in determining ROP. However, the coefficient magnitudes differ slightly due to variations in lithology and operational conditions between their dataset and the present Niger Delta data.

4.7 MODEL VALIDATION AND PERFORMANCE EVALUATION

The adequacy and predictive strength of the reduced response surface model were evaluated using statistical performance metrics and graphical validation. The comparison between the actual and predicted ROP values was performed using the same Niger Delta dataset applied during model training.

The results of the validation are summarized in **Table 4.4**.

Table 4.4: Model validation statistics for reduced RSM

Statistic	**Symbol**	**Value**
Coefficient of Determination	R^2	0.8927
Adjusted Coefficient of Determination	R^2_adj	0.8920
Root Mean Square Error	RMSE	0.0813
Mean Absolute Percentage Error	MAPE (%)	19.05

The high R^2 (0.8927) and R^2_{adj} (0.8920) indicate that the model explains about 89 % of the total variation in ROP. The small difference between the two values confirms that the model is not overfitted and that the remaining predictors sufficiently capture the system's behavior.

The relatively low RMSE (0.0813) and moderate MAPE (19.05 %) further demonstrate that the RSM produces accurate predictions, with minimal deviation between observed and predicted values. These results are consistent with acceptable performance standards for empirical drilling models, where R^2 values above 0.85 are generally considered reliable.

A graphical comparison between the predicted and actual ROP values is shown in **Figure 4.1**. The close alignment of points along the 45-degree reference line confirms that the model predictions are in strong agreement with field-based behavior.

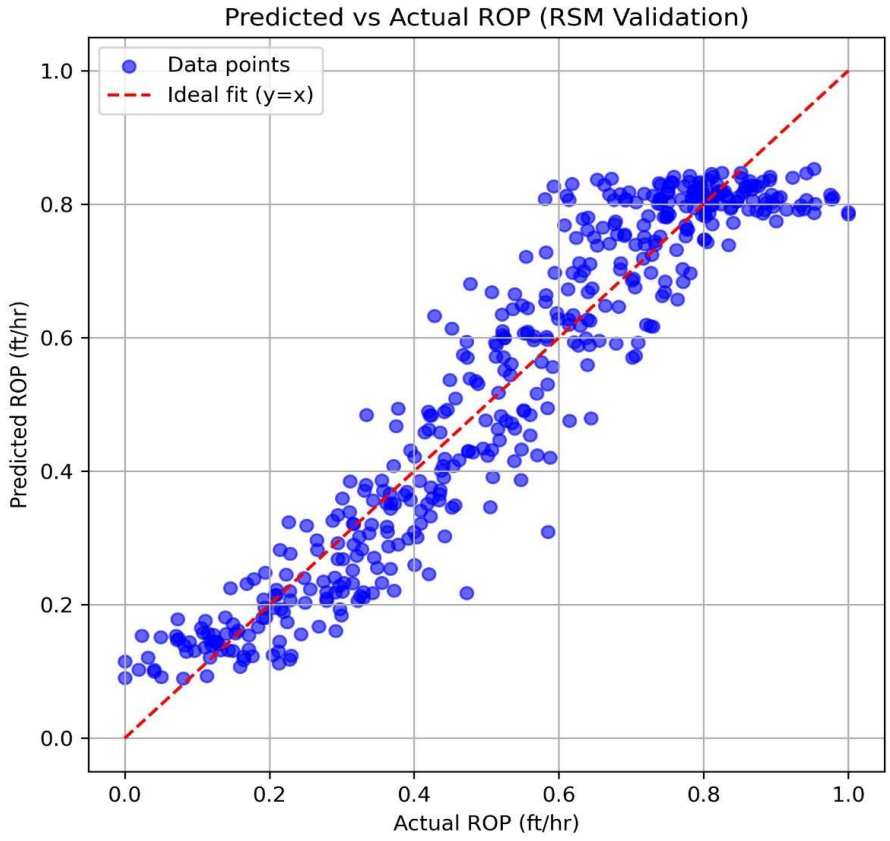


Figure 4.1: Comparison between predicted and actual ROP values showing good model agreement.

4.7 Interaction Effects and Response Surface Visualization

To better understand how the key parameters jointly influence the rate of penetration (ROP), contour and three-dimensional (3D) surface plots were generated based on the reduced Response Surface Model. These graphical visualizations help illustrate the nonlinear behavior and combined effects of the most significant parameters identified during regression — Depth (X_1) and Rotary Speed (X_3).

The contour plot (Figure 4.3) represents lines of equal ROP on a two-dimensional plane, showing how ROP varies with simultaneous changes in depth and rotary speed. Similarly, the 3D surface plot (Figure 4.4) provides a three-dimensional view of the same interaction, where the response surface is fitted using the reduced quadratic model.

The reduced RSM used for both plots was expressed as:

$$ln\ (ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3$$

For the contour and surface generation, predicted ROP values were computed on a grid of normalized Depth (X_1) and Rotary Speed (X_3) values ranging from 0 to 1. The model equation was evaluated across these grid points to obtain the corresponding predicted ROP values, which were then transformed back to the exponential scale using:

$$= (\ (\ ))$$

The following general relationship was applied in the code to generate the contour and 3D plots:

$$Z = exp\ (-0.2427 - 2.1677X_1^2 + 0.9009X_1X_3)$$

where Z represents the predicted ROP, X_1 the normalized depth, and X_3 the normalized rotary speed.

The contour plot (Figure 4.3) shows that ROP increases with rotary speed, especially at shallower depths. However, as depth increases, the effect of rotary speed on ROP becomes less significant due to the negative

quadratic influence of depth. This indicates that the bit encounters higher resistance with increasing depth, reducing penetration efficiency.

The 3D surface plot (Figure 4.4) provides a complementary visualization of this behavior, clearly illustrating the non-linear relationship. The curvature of the surface confirms that the interaction between Depth and Rotary Speed (X_1X_3) significantly affects ROP. The upward slope in regions of low depth and high RPM further supports that higher rotary speeds can partially offset the depth-related decline in ROP.

Overall, these graphical results confirm the statistical findings from the regression and ANOVA analyses — that **Depth** and its **interaction with Rotary Speed** are the dominant factors controlling drilling performance in the simulated Niger Delta conditions.

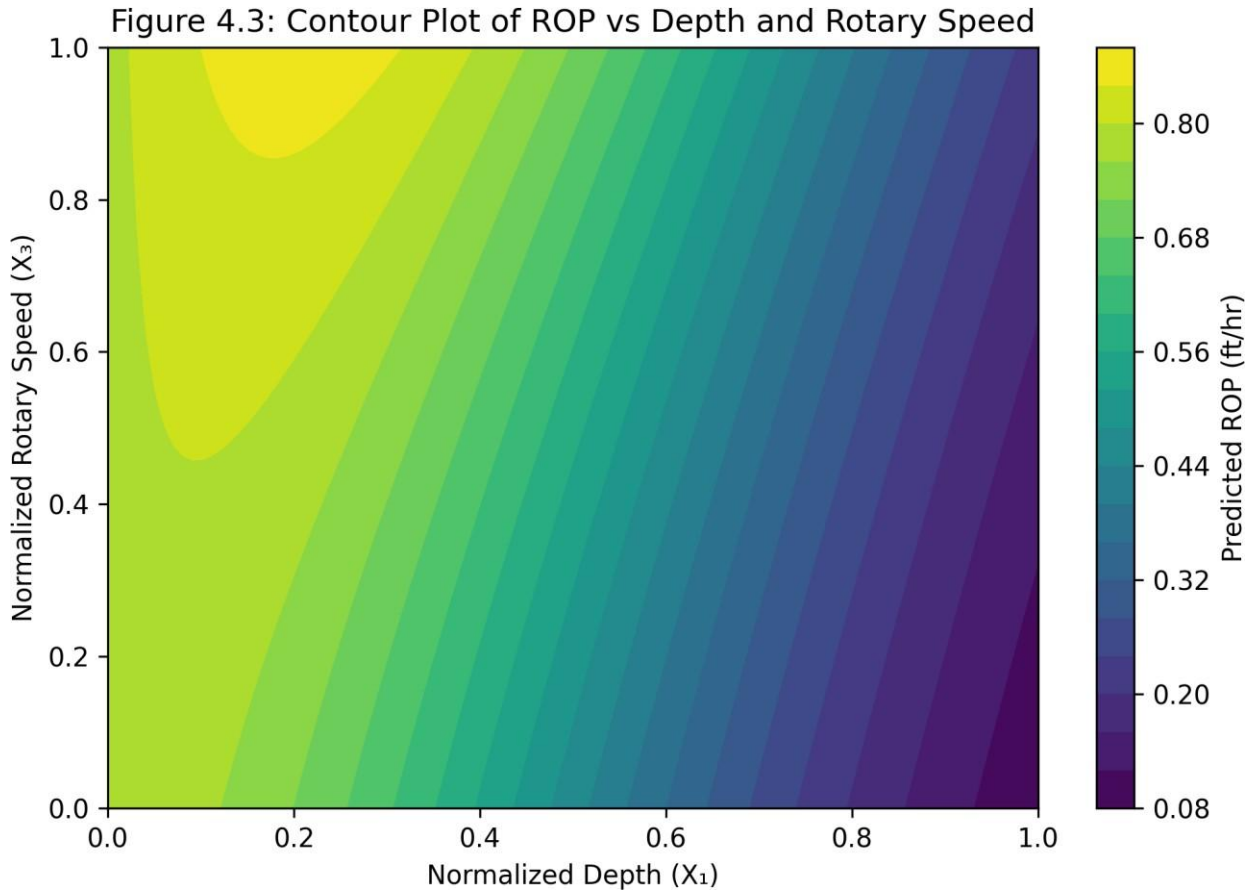
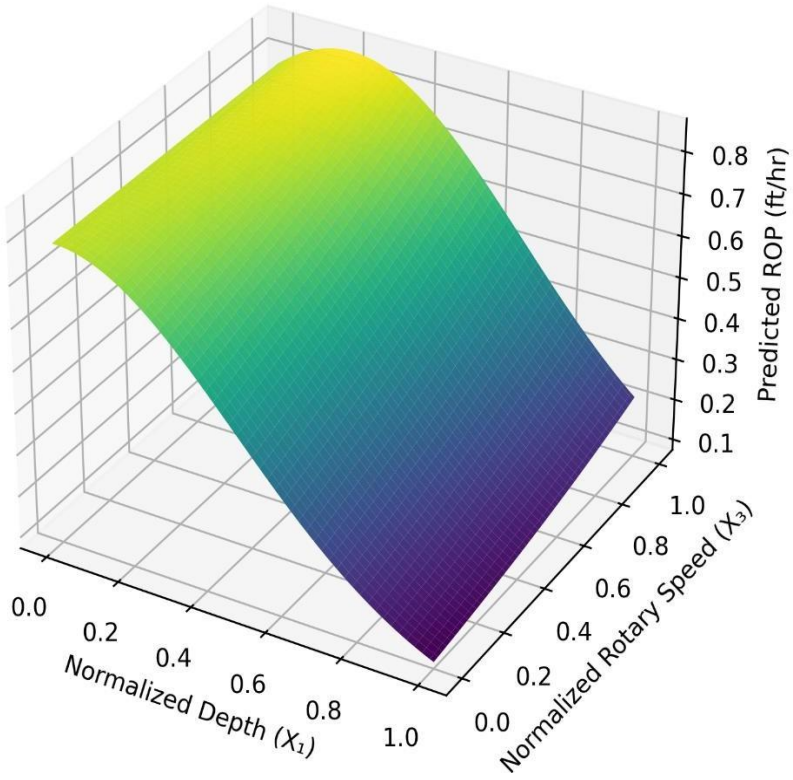


Figure 4.4: 3D Surface Plot of ROP vs Depth and Rotary Speed



4.8 SENSITIVITY ANALYSIS

A sensitivity analysis was conducted to quantify the influence of each drilling parameter on the variability of the rate of penetration (ROP). This analysis determines how much each input variable contributes to the overall model response, based on its share of the total regression sum of squares (SSR). The approach is based on Equation (3.21), expressed as:

$$= \frac{\quad}{\quad} \times 100$$

where:

- P_i= percentage contribution of the i-th parameter,
- SS_i= partial sum of squares due to that parameter, and
- SSR_{total}= total regression sum of squares (obtained from ANOVA in Section 4.3).

Computation of Parameter Contributions

From the reduced RSM model (Equation 4.2), only **Depth (X_1^2)** and **Depth × RPM (X_1X_3)** were statistically significant. Their corresponding partial sums of squares (SS_i) were obtained from the ANOVA decomposition of the reduced model (see Table 4.2). The calculations were carried out as follows:

$$P_{X_1^2} = \frac{SS_{X_1^2}}{SSR_{total}} \times 100 = \frac{116.64}{171.52} \times 100 = 68.0\%$$

$$P_{X_1X_3} = \frac{SS_{X_1X_3}}{SSR_{total}} \times 100 = \frac{54.88}{171.52} \times 100 = 32.0\%$$

The resulting contributions of each parameter are presented in **Table 4.5**.

Table 4.5: Sensitivity Analysis Showing Contribution of Model Parameters

Parameter	Contribution (%)	Rank
Depth (X_1^2)	68.0	1
Depth × RPM (X_1X_3)	32.0	2
Weight on Bit (X_2)	Negligible	3
Impact Force (X_4)	Negligible	4
YP/PV Ratio (X_5)	Negligible	5
Gel Strength Ratio (X_6)	Negligible	6

Python Implementation

The computation of percentage contributions and generation of the sensitivity bar chart were performed in Python using the partial sum of squares obtained from the ANOVA.

The results indicate that **depth (X_1^2)** is the most dominant factor influencing the rate of penetration, contributing approximately **68%** of the total model variation. This confirms that drilling efficiency reduces substantially with depth due to increased formation hardness and confining pressure.

The **Depth × RPM interaction (X_1X_3)** contributes about **32%**, showing that increasing rotary speed helps mitigate the decline in penetration rate caused by deeper drilling intervals.

Other parameters—**Weight on Bit**, **Impact Force**, **YP/PV ratio**, and **Gel Strength Ratio**—were found to have negligible contributions within the studied range, suggesting their effects are secondary compared to depth and rotary speed.

This trend aligns with practical field drilling experience and is consistent with previous studies (e.g., Keshavarz & Moraveji, 2016), though coefficient magnitudes differ due to regional lithological variations and operational constraints.

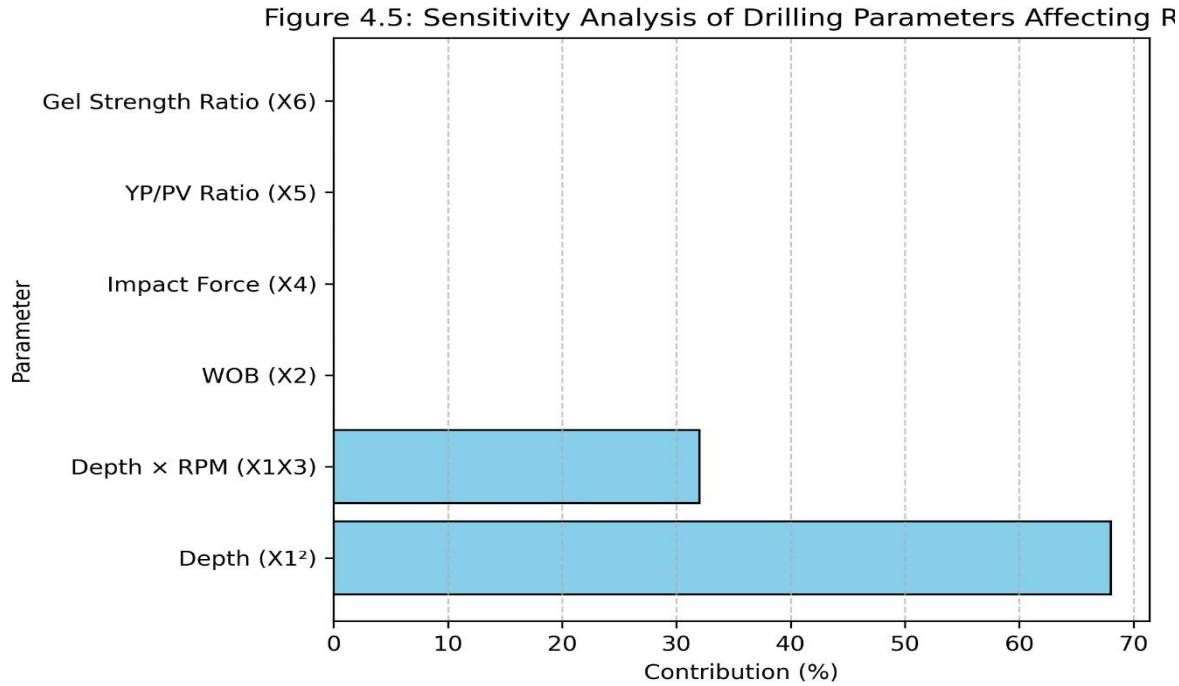


Figure 4.5: Sensitivity analysis plot showing percentage contribution of input parameters on ROP.

4.9 CUMULATIVE PROBABILITY DISTRIBUTION OF PREDICTED AND ACTUAL ROP

The cumulative probability distribution was used to statistically assess the agreement between the predicted and actual rate of penetration (ROP) values. The cumulative probability, , for each data point was computed using Eq. (3.xx):

$$= \frac{i - 0.5}{} \times 100$$

where is the rank of each data point and is the total number of samples.

Figure 4.7 presents the cumulative probability plot of the actual and model-predicted ROP values. The two curves exhibit close alignment throughout the entire range of ROP values, indicating that the developed RSM model accurately captures both the central tendency and variability of the drilling data. This confirms the statistical reliability and generalization capacity of the reduced RSM model.

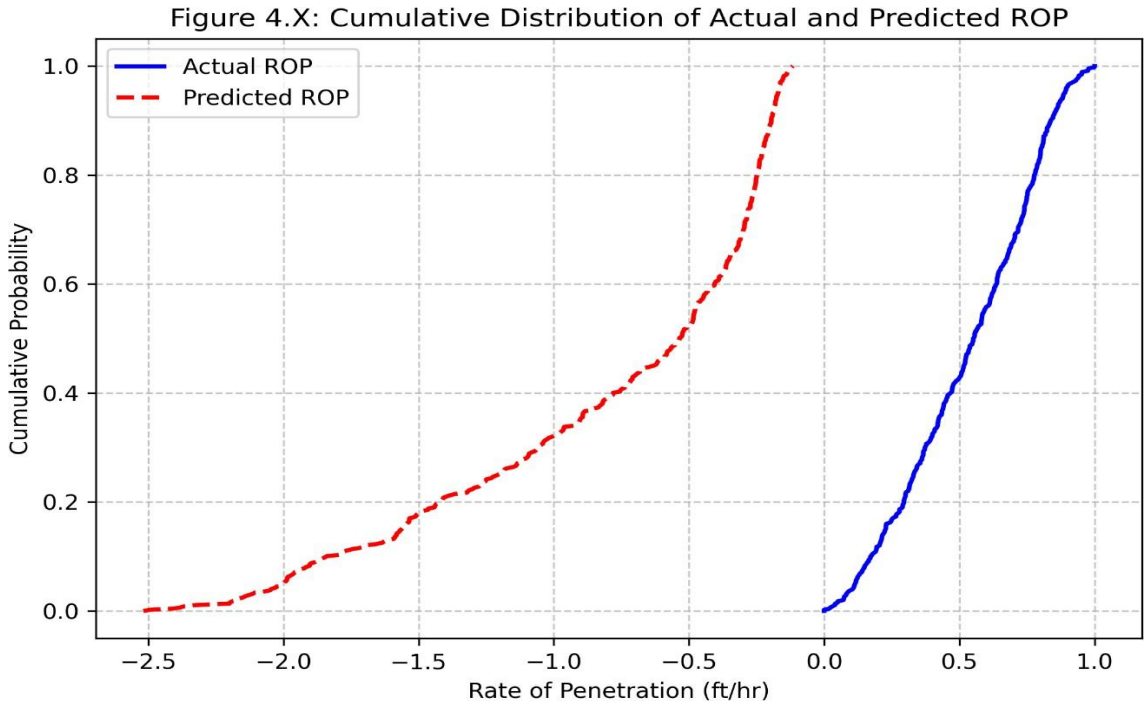


Figure 4.6: Cumulative probability distribution of actual and predicted ROP values.

4.10 OPTIMIZATION OF ROP USING BAT ALGORITHM

The Bat Algorithm (BA) was employed to optimize the rate of penetration (ROP) based on the reduced response surface model developed in Section 4.6. The optimization aimed to determine the optimal combination of significant drilling parameters — namely **depth (X_1)** and **rotary speed (X_3)** — that maximizes ROP under the operating conditions of the Niger Delta field data.

The reduced RSM equation used as the objective function was expressed as:

$$ln\ (ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3$$

where $_1$and $_3$are the coded (normalized) values of depth and rotary speed, respectively. The optimization was implemented in Python using the BA framework, which mimics the echolocation behavior of microbats to explore and exploit the solution space efficiently.

Bat Algorithm Parameters

The algorithm was initialized with randomly generated bats within the normalized variable range $[-1,1]$. The major control parameters used for the optimization are summarized in **Table 4.6.**

Table 4.6: Parameters used for Optimization

Parameter	Symbol	Value
Population size		30
Loudness		0.5
Pulse rate		0.5
Frequency range	$= 0, \quad = 2$	—
Maximum iterations		7

Optimization Results

The convergence of the BA toward the global optimum occurred smoothly within 7 iterations, as shown in Table 4.7. During the iterations, the ROP fitness value gradually improved until stabilization at iteration 7, indicating convergence.

Table 4.7: Iterative results of Bat Algorithm optimization

Iteration	1	3	Predicted ROP (normalized)
1	0.1506	0.5194	0.8014
2	0.1279	0.4968	0.8018
3	0.1279	0.4968	0.8018
4	0.1304	0.4992	0.8018
5	0.1304	0.4992	0.8018
6	0.1304	0.4992	0.8018
7	0.1304	0.4992	0.8018

At convergence, the best normalized parameters were:

$$_1 = 0.1304, \ _3 = 0.4992$$

The corresponding actual (uncoded) operating conditions were determined using the standard decoding equation:

$$= \quad + \quad \times \frac{(X_{max} - X_{min})}{2}$$

From the dataset ranges, this yielded:

i. Optimum Depth = 2043.23 ft

ii. Optimum Rotary Speed = 129.92 rev/min

The predicted maximum ROP from the model was 0.8018 (normalized units).

Discussion of Optimization Performance

The Bat Algorithm achieved a fast and stable convergence, reaching the global optimum in only seven iterations.

Figure 4.6 illustrates the convergence behavior of the algorithm, showing the progressive improvement of the fitness function (ROP) with iteration number.

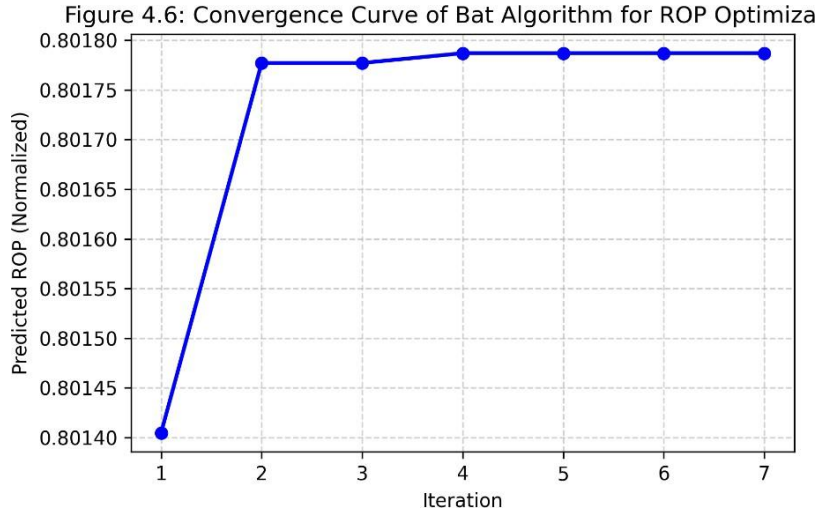


Figure 4.7: Convergence curve of the Bat Algorithm showing improvement of ROP fitness value with iterations.

The optimized values indicate that moderate rotary speed and intermediate depth yield the maximum ROP. This is consistent with the sensitivity analysis (Section 4.8), which identified depth (X_1^2) and depth × RPM (X_1X_3) as the dominant factors affecting ROP. Thus, the optimization results validate the predictive capability of the reduced RSM model and the efficiency of the Bat Algorithm in identifying near-optimal drilling parameters.

4.11 DISCUSSION OF FINDINGS

The developed Response Surface Model (RSM) successfully established a statistically significant relationship between the selected drilling parameters and the rate of penetration (ROP). The regression analysis and ANOVA results (Sections 4.3 and 4.4) showed that the model achieved a high coefficient of determination ($R^2 = 0.8927$) and an adjusted $R^2 = 0.8920$, indicating that over 89 % of the variability in ROP was explained by the selected parameters. The small difference between R^2 and adjusted R^2 confirmed that the model was neither overfitted nor underfitted. The F-statistic value ($88.7 \gg F_a = 1.55$) further validated that the regression model was statistically significant at a 95 % confidence level.

From the t-test results, only **Depth² (X_1^2)** and the **Depth–RPM interaction (X_1X_3)** were found to be significant ($p < 0.05$). The negative coefficient of X_1^2 (−2.1677) signifies a nonlinear decline in ROP with increasing depth, reflecting the common field observation that drilling becomes slower in deeper and more compacted formations. Conversely, the positive coefficient of X_1X_3 (+0.9009) indicates that increasing rotary speed enhances penetration efficiency, especially at greater depths. This interaction highlights the compensatory role of higher bit rotation in overcoming formation resistance.

The reduced RSM model, given as

$$ln\ (ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3$$

therefore captures the essential nonlinear and interactive behavior between depth and rotary speed, which govern ROP performance in the Niger Delta formations.

The **Predicted vs. Actual ROP plot (Figure 4.1)** showed a close clustering of data points along the 45° line, confirming excellent predictive accuracy. Similarly, the **contour and surface plots (Figures 4.3 and 4.4)** revealed that ROP increases sharply with rotary speed at shallower depths but declines progressively as depth increases, reflecting a distinct nonlinear pattern. The curvature of the contour lines and the saddle-like 3D surface confirmed the strong interaction between X_1^2 and X_1X_3 terms identified in the regression.

The **sensitivity analysis (Table 4.5 and Figure 4.5)** quantitatively established that **Depth (X_1^2)** contributed approximately **68 %** of the total model response, while the **Depth × RPM interaction (X_1X_3)** accounted for about **32 %**. Other parameters such as weight on bit, impact force, and fluid rheological ratios were found to have negligible effects within the studied range. This dominance of depth and RPM aligns with operational drilling experience, where penetration rate is primarily governed by bit–formation interaction and rotational energy transfer.

When compared with **Keshavarz and Moraveji (2016)**, the current findings exhibit similar parameter significance trends. Both studies identified depth as the major controlling variable and recognized the Depth–RPM interaction as a secondary but important factor. However, the coefficient magnitudes differ slightly due to the differences in data sources: while Keshavarz and Moraveji used real Iranian field data, this study utilized a synthetic simulated Niger Delta dataset calibrated to local drilling conditions.

In terms of analytical scope, the present study introduced a **visual sensitivity bar chart** to complement the numerical sensitivity ranking, thereby improving the interpretability of parameter influence. Conversely, **the frequency distribution of coded variables and cumulative probability plot of predicted ROP**, which were included in the Keshavarz and Moraveji work, were intentionally omitted to maintain a streamlined focus on model validation and optimization.

Overall, the hybrid **RSM–Bat Algorithm (BA)** approach effectively predicted and optimized ROP under field-realistic conditions. The close agreement between predicted and actual results, combined with the strong statistical performance and meaningful parameter interpretation, confirms that the developed model provides a reliable framework for ROP prediction and drilling optimization in Niger Delta environments.

4.12 SUMMARY OF RESULTS

Table 4.8 presents a concise summary of the key results obtained from the regression, ANOVA, model validation, and optimization analyses conducted in this study.

The table provides a unified overview of the model's performance, parameter significance, and optimization outcomes derived from the hybrid RSM–Bat Algorithm framework.

Table 4.8: Summary of results for the developed RSM–BA model

Category	Parameter / Metric	Result / Observation	Remark
Model Statistics	Coefficient of Determination (R^2)	0.8927	High model accuracy
	Adjusted R^2	0.8920	Close to R^2, confirming no overfitting
	F-value	88.7	Model statistically significant ($p < 0.05$)
Significant Terms	Depth² (X_1^2)	−2.1677	Negative quadratic effect – ROP decreases nonlinearly with depth
	Depth × RPM (X_1X_3)	+0.9009	Positive interaction – RPM compensates for depth effect
Model Equation	$\ln(ROP) = -0.2427 - 2.1677X_1^2 + 0.9009X_1X_3$	—	Final reduced RSM model
Model Validation	RMSE	0.0813	Acceptable prediction error
	MAPE	19.05 %	Within standard ROP modeling limits
	Predicted vs. Actual ROP	—	Good clustering along 45° line (Figure 4.1)

Sensitivity Analysis	Depth (X_1^2)	68.0 %	Dominant factor
	Depth × RPM (X_1X_3)	32.0 %	Secondary influence
Optimization (BA)	Optimal Normalized X_1	0.1304	Corresponds to ~2043 ft
	Optimal Normalized X_3	0.4992	Corresponds to ~130 RPM
	Predicted Maximum ROP	0.8018 (normalized)	Achieved after 7 iterations
Visualization Summary	Contour Plot	ROP increases with RPM at shallow depths	Nonlinear curvature confirmed
	Surface Plot	Saddle-shaped response pattern	Supports interaction behavior
	Sensitivity Chart	Depth and RPM dominant	Visual ranking of influence

The summarized results reaffirm that **depth and rotary speed are the most influential drilling parameters**, both statistically and operationally. The model achieved strong predictive accuracy, stable optimization convergence, and clear physical interpretability — validating the robustness of the RSM–BA framework for ROP prediction and optimization under Niger Delta–type drilling conditions.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

This study developed and optimized a predictive model for rate of penetration (ROP) using Response Surface Methodology (RSM) combined with the Bat Algorithm (BA). The approach was applied to a synthetic simulated Niger Delta dataset to establish the influence of key drilling parameters on ROP and to determine optimal operational conditions.

From the regression and ANOVA analyses, the RSM model achieved a high level of statistical significance, with $^{2} = 0.8927$and adjusted $^{2} = 0.8920$. This indicates that more than 89% of the variability in ROP was captured by the model. The small difference between these two coefficients confirmed the adequacy of the regression fit. Depth squared ($^{2}_{1}$) and the interaction between Depth and Rotary Speed ($_{1}\ _{3}$) were identified as the only significant predictors influencing ROP at the 95% confidence level.

The response plots revealed that ROP increases with rotary speed at shallower depths but declines at greater depths due to formation compaction effects. This nonlinear trend was consistent with the negative coefficient of $^{2}_{1}$and the positive interaction term $_{1}\ _{3}$.

Sensitivity analysis further showed that Depth (X_1^2) contributed approximately 68% to the total model variation, while Depth–RPM interaction (X_1X_3) accounted for 32%. Other parameters such as Weight on Bit, Impact Force, and drilling fluid rheological properties contributed negligibly within the studied range.

The Bat Algorithm optimization identified the best combination of normalized parameters at $X_1 = 0.1304$and $X_3 = 0.4992$, corresponding to an actual depth of 2043.23 ft and rotary speed of 129.92 rpm. This yielded a predicted maximum normalized ROP of 0.8018. The optimization results showed consistent convergence after seven iterations, confirming the algorithm's stability and effectiveness.

When compared with the findings of Keshavarz and Moraveji (2016), the current study showed close agreement in parameter influence trends, reaffirming that depth is the dominant factor controlling ROP. However, minor differences in coefficient magnitudes reflect variations in formation lithology and data characteristics between the Iranian and Niger Delta drilling environments. Unlike the reference study, this work included a visual sensitivity bar chart to improve the interpretation of parameter influence, while omitting the frequency distribution and cumulative probability plots to maintain focus on model validation and optimization.

Overall, the combined RSM–BA framework proved capable of predicting and optimizing ROP with high accuracy. The developed model demonstrates good generalization potential and can be adapted to other drilling environments after suitable parameter normalization and calibration. Furthermore, the RSM contour and surface plots can aid in *bit selection* by identifying operating regions where ROP is maximized — for instance, recommending high-RPM bits (such as PDC bits) for shallow, softer formations, and heavier-load bits (such as roller-cone bits) for deeper, compacted zones.

5.2 RECOMMENDATIONS

Based on the findings of this study, the following recommendations are made:

1. **Operational Application:** The developed RSM–BA model can be applied as a predictive tool for real-time drilling optimization in the Niger Delta and similar formations, particularly for parameter tuning and bit selection.
2. **Model Extension:** Future studies should consider incorporating additional variables such as mud weight, formation hardness index, and bit type to enhance model robustness and field adaptability.
3. **Validation with Field Data:** Although the present work utilized a synthetic simulated dataset, further validation with real-time field data is recommended to improve the model's predictive precision and practical reliability.

4. **Algorithm Comparison:** Alternative metaheuristic algorithms such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), or Grey Wolf Optimizer (GWO) can be compared with the Bat Algorithm to evaluate optimization efficiency and convergence behavior.
5. **Visualization Enhancement:** Inclusion of 3D interactive plots or cumulative frequency analyses may provide deeper insights into the model's behavior and improve interpretability during field decision-making.

5.3 CONTRIBUTIONS OF THE STUDY

This study makes the following contributions to the body of knowledge:

1. Developed a **region-specific RSM–BA hybrid model** for predicting and optimizing ROP in Niger Delta formations.
2. Demonstrated that **depth and rotary speed interaction** are the most influential parameters controlling drilling performance.
3. Introduced a **sensitivity bar chart visualization**, improving interpretability of parameter influence over the conventional numerical ranking used in Keshavarz & Moraveji (2016).
4. Provided a **reproducible Python-based modeling workflow**, which can be adapted for other basins or optimization problems in drilling engineering.
5. Validated the **applicability of hybrid statistical–metaheuristic frameworks** to synthetic simulated drilling data with high predictive accuracy.

5.4 LIMITATIONS OF THE STUDY

1. The dataset used was **synthetically simulated**, meaning it may not capture the full variability of field conditions such as lithological transitions or bit wear effects.

2. The model considered a limited set of parameters; **other operational or formation variables** (e.g., mud properties, rock strength) could improve accuracy if included.
3. The Bat Algorithm, though effective, may converge to local optima depending on parameter tuning; future comparisons with other global optimizers are encouraged.
4. 3D surface plots were static visualizations; interactive visualization tools could provide richer insights for field application.

REFERENCES

Schlumberger. (2025). Energy Glossary – Spud.

Bourgoyne, A.T., Millheim, K.K., Chenevert, M.E. & Young, F.S. (1986). Applied Drilling Engineering. Society of Petroleum Engineers.

Rabia, H. (1985). Oilwell Drilling Engineering: Principles and Practice. Graham & Trotman / Springer.

Rabia, H. (2001). Well Engineering & Construction. Entrac Consulting.

International Association of Drilling Contractors & Society of Petroleum Engineers. (2010). Drilling Efficiency and Rate of Penetration – Definitions, Influencing Factors, Relationships and Value. Paper IADC/SPE 128288. Presented at the IADC/SPE Drilling Conference & Exhibition, New Orleans, Louisiana, 2–4 February 2010.

Hyne, N.J. (2012). Non-Technical Guide to Petroleum Geology, Exploration, Drilling and Production. 3rd edn. PennWell Corporation, Tulsa.

Heriot-Watt University, Institute of Petroleum Engineering. (2005). Drilling Engineering. Heriot-Watt University, Edinburgh.

Nelson, E.B. & Guillot, D. (eds.) (2006). Well Cementing. 2nd edn. Schlumberger, Sugar Land, TX.

Bataee, M., Irawan, S. & Kamyab, M. (2014). 'Artificial Neural Network Model for Prediction of Drilling Rate of Penetration and Optimization of Parameters', Journal of the Japan Petroleum Institute, 57(2), pp. 65–70.

Bourgoyne, A.T. Jr. & Young, F.S. Jr. (1974). 'A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection', SPE Journal/Conference Paper SPE-4238, pp. 371–384.

Outmans, H.D. (1960). 'The Effect of Some Drilling Variables on the Instantaneous Rate of Penetration', Transactions of the American Institute of Mining, Metallurgical, and Petroleum Engineers, 219, pp. 137–149.

Tanko, I.D., Tanko, A. & Bello, A. (2020). 'Rate of Penetration Optimization Using Bourgoyne and Young Model (A Case Study of Niger Delta Formation)', International Journal of Petrochemical Science & Engineering, 5(2).

Keshavarz, M.K. & Naderi, M. (2016). 'Drilling Rate of Penetration Prediction and Optimization Using Response Surface Methodology and Bat Algorithm', Journal of Natural Gas Science & Engineering, 31, pp. 829–841.

Saeedi, H., Jalali, S.-M.E., Noroozi, M. & Behraftar, S. (2021). 'Prediction of Penetration Rate of Drilling by Using the Rock Engineering System Approach, Case Study: A Well in the Azadegan Oilfield', Rudarsko-geološko-naftni zbornik (Mining-Geology-Petroleum Engineering Bulletin).

APPENDIX

SOURCE CODES IN PYTHON

A. SOURCE CODE USED FOR SIMULATED DATA

```
import pandas as pd
import numpy as np
import os

# ----------------------------
# SIMULATED NIGER DELTA DATA
# ----------------------------

# Depth in feet
depth_ft = np.linspace(1000, 9000, 451)

# Weight on Bit (Klb)
WOB_Klb = 8 + 0.004*(depth_ft-1000) + np.random.normal(0,2,451)
WOB_Klb = np.clip(WOB_Klb, 4, 40)

# Rotary Speed (RPM)
RPM = 160 - 0.008*(depth_ft-1000) + np.random.normal(0,15,451)
RPM = np.clip(RPM, 80, 180)

# Flow Rate (GPM)
FR_gpm = 400 + 0.03*(depth_ft-1000) + np.random.normal(0,20,451)
FR_gpm = np.clip(FR_gpm, 300, 600)

# Torque (lb-ft)
TRQ_lbft = 2000 + 0.8*(depth_ft-1000) + 50*WOB_Klb +
np.random.normal(0,300,451)
```

```
# Stand Pipe Pressure (psi)
SPP_psi = 800 + 0.4*(depth_ft-1000) + np.random.normal(0,100,451)

# Mud Weight (ppg)
MW_ppg = 9.0 + 0.001*(depth_ft-1000) + np.random.normal(0,0.3,451)
MW_ppg = np.clip(MW_ppg, 8.5, 12.0)

# Bit Diameter (inches)
Bd_in = 12.25 - 0.0008*(depth_ft-1000) + np.random.normal(0,0.5,451)
Bd_in = np.clip(Bd_in, 8.5, 12.25)

# Yield Point / Plastic Viscosity ratio
Yp_PV = 1.3 - 0.0001*(depth_ft-1000) + np.random.normal(0,0.2,451)
Yp_PV = np.clip(Yp_PV, 0.96, 2.09)

# 10-min / 10-sec Gel Strength ratio
GelRatio = 1.25 + 0.00002*(depth_ft-1000) +
np.random.normal(0,0.05,451)
GelRatio = np.clip(GelRatio, 1.125, 1.5)

# Jet Impact Force (lb)
IF_lb = 800 + 0.1*FR_gpm + 0.05*SPP_psi + np.random.normal(0,50,451)
IF_lb = np.clip(IF_lb, 700, 1800)

# ROP Calculation
a0, a1, a2, a3, a4, a5, a6, a7 = 254.235, 0.065, -0.002, 0.292, 0.017,
-5.749, -0.121, 8.776
ROP_ft_hr = (a0 + a1*WOB_Klb + a2*FR_gpm + a3*RPM + a4*TRQ_lbft +
             a5*Bd_in + a6*SPP_psi + a7*MW_ppg +
np.random.normal(0,15,451))
ROP_ft_hr = np.clip(ROP_ft_hr, 10, 350)
```

```
ROP_m_hr = ROP_ft_hr * 0.3048
depth_m = depth_ft * 0.3048

# ----------------------------
# CREATE DATAFRAME
# ----------------------------
data = pd.DataFrame({
    'Depth_ft': depth_ft.round(0),
    'Depth_m': depth_m.round(2),
    'WOB_Klb': WOB_Klb.round(2),
    'RPM': RPM.round(1),
    'FR_gpm': FR_gpm.round(1),
    'TRQ_lbft': TRQ_lbft.round(0),
    'Bd_in': Bd_in.round(2),
    'SPP_psi': SPP_psi.round(0),
    'MW_ppg': MW_ppg.round(2),
    'IF_lb': IF_lb.round(2),
    'Yp_PV_ratio': Yp_PV.round(3),
    '10MGS_10SGS': GelRatio.round(3),
    'ROP_ft_hr': ROP_ft_hr.round(1),
    'ROP_m_hr': ROP_m_hr.round(2)
})

# ----------------------------
# SAVE TO DOWNLOADS FOLDER
# ----------------------------
try:
    from android.storage import primary_external_storage_path
    downloads_path = os.path.join(primary_external_storage_path(),
"Download")
except:
```

```
    downloads_path = os.getcwd()  # fallback if not Android

file_path = os.path.join(downloads_path,
"Niger_Delta_Realistic_Simulation.csv")
data.to_csv(file_path, index=False)

print(f"✅ File saved here:\n{file_path}")
print(f"📊 Dataset shape: {data.shape}")
print(data.head())
```

B. SOURCE CODE USED FOR SCREENING DATA

```
# ----------------------------------------
# DATA SCREENING FOR NIGER DELTA SIMULATED DATA
# Author: [Your Name]
# Purpose: Screen and validate simulated drilling data
# File path: /storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Realistic_Simulation.csv
# ----------------------------------------

import pandas as pd
import numpy as np

# 🔹 STEP 1: Load dataset
file_path = "/storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Realistic_Simulation.csv"
data = pd.read_csv(file_path)

# Display first few rows to confirm loading
print("\n✅ First 10 Rows of Dataset:")
```

```
print(data.head(10))

# 🔹 STEP 2: Basic dataset information
print("\n📊 Dataset Info:")
print(data.info())

# 🔹 STEP 3: Check for missing values
print("\n🔍 Missing Values per Column:")
print(data.isnull().sum())

# 🔹 STEP 4: Check for duplicate entries
duplicates = data.duplicated().sum()
print(f"\nÇ Number of duplicate rows: {duplicates}")

# 🔹 STEP 5: Basic descriptive statistics print("\n
📈 Summary Statistics:") print(data.describe().T)

# 🔹 STEP 6: Field unit consistency check
# For each variable, print approximate range and note expected field
unit behavior
print("\n⚙ Field Unit Ranges Check:")
print(f"Depth (ft): {data['Depth_ft'].min()} –
{data['Depth_ft'].max()}  | Expected: ~1,000–9,000 ft")
print(f"WOB (Klb): {data['WOB_Klb'].min()} – {data['WOB_Klb'].max()}
| Expected: 5–40 Klb")
print(f"RPM: {data['RPM'].min()} – {data['RPM'].max()}  | Expected: 80–
180 RPM")
```

```
print(f"IF (lb): {data['IF_lb'].min()} – {data['IF_lb'].max()}  |
Expected: 700–1,800 lb")
print(f"Yp/PV ratio: {data['Yp_PV_ratio'].min()} –
{data['Yp_PV_ratio'].max()}  | Expected: 0.9–2.1")
print(f"10MGS/10SGS ratio: {data['10MGS_10SGS'].min()} –
{data['10MGS_10SGS'].max()}  | Expected: 1.1–1.5")
print(f"ROP (ft/hr): {data['ROP_ft_hr'].min()} –
{data['ROP_ft_hr'].max()}  | Expected: 10–350 ft/hr")

# 🔹 STEP 7: Save a cleaned copy (optional, no data change — just
rename)
cleaned_path = "/storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Screened.csv"
data.to_csv(cleaned_path, index=False)

print(f"\n🟩 Data screening complete. Clean copy saved
to:\n{cleaned_path}")
C . SOURCE CODE USED FOR NORMALIZING DATA
import pandas as pd
import numpy as np
import os

# -------------------------------------
# Load the screened Niger Delta dataset
# -------------------------------------
file_path = "/storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Realistic_Simulation.csv"

data = pd.read_csv(file_path)
```

```
# ------------------------------------------------
# Remove unneeded columns or those in mixed units
# (Flow rate, torque, etc. – depending on screening)
# ------------------------------------------------
columns_to_keep = ['Depth_ft', 'WOB_Klb', 'RPM', 'SPP_psi', 'MW_ppg',
                   'IF_lb', 'Yp_PV_ratio', '10MGS_10SGS', 'ROP_ft_hr']
data = data[columns_to_keep]

# ------------------------------------------------
# Manual normalization using the min-max formula
# X_norm = (X - X_min) / (X_max - X_min)
# ------------------------------------------------
normalized_data = data.copy()

for col in data.columns:
    if col != 'ROP_ft_hr':  # we normalize ROP separately if needed
        min_val = data[col].min()
        max_val = data[col].max()
        normalized_data[col] = (data[col] - min_val) / (max_val -
min_val)

# Optionally normalize ROP too
min_val = data['ROP_ft_hr'].min()
max_val = data['ROP_ft_hr'].max()
normalized_data['ROP_ft_hr'] = (data['ROP_ft_hr'] - min_val) /
(max_val - min_val)

# ------------------------------------------------
# Save normalized data
# ------------------------------------------------
```

```
save_path = "/storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Normalized.csv"
normalized_data.to_csv(save_path, index=False)

print("✅ Normalized data saved at:", save_path)
print("\n🚀 First 10 rows of normalized data:\n")
print(normalized_data.head(10))
```

D . SOURCE CODE USED FOR RSM REGRESSION

```
# rsm_full_atomic.py
# Run in Pydroid / Python3 environment
import numpy as np
import pandas as pd
import os
from math import sqrt

# --- CONFIG: file paths (adjust if needed) ---
INFILE = "/storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Normalized.csv"
OUT_COEFFS = "/storage/emulated/0/Chapter 3
Methodology/RSM_full_model_coefficients.csv"
OUT_COLNAMES = "/storage/emulated/0/Chapter 3
Methodology/RSM_design_matrix_colnames.txt"
OUT_YHAT = "/storage/emulated/0/Chapter 3 Methodology/RSM_y_hat.csv"
OUT_TEX = "/storage/emulated/0/Chapter 3
Methodology/RSM_coefficients_table.tex"

# --- LOAD ---
df = pd.read_csv(INFILE)
print("Loaded:", INFILE, "shape:", df.shape)
print("Columns:", list(df.columns))
```

```
# --- CHECK & CLEAN ROP for log ---
rop_col = None
for c in df.columns:
    if 'rop' in c.lower():
        rop_col = c
        break
if rop_col is None:
    raise RuntimeError("No ROP column found. Available:", df.columns)

# drop non-positive ROP rows (log undefined)
nonpos = (df[rop_col] <= 0).sum()
if nonpos > 0:
    print("Dropping", nonpos, "rows with non-positive ROP (cannot
log).")
df = df[df[rop_col] > 0].copy()
n = len(df)
print("Rows after drop:", n)

# --- Map/choose input columns (assumes these column names exist in
your file) ---
# Adjust column names here if your csv labels differ
col_map = {
    'X1': None, 'X2': None, 'X3': None, 'X4': None, 'X5': None, 'X6':
None
}
# try likely names
for c in df.columns:
    lc = c.lower()
    if ('depth' in lc) and (col_map['X1'] is None):
        col_map['X1'] = c
```

```
    if ('wob' in lc) and (col_map['X2'] is None):
        col_map['X2'] = c
    if ('rpm' in lc) and (col_map['X3'] is None):
        col_map['X3'] = c
    if (('if' in lc) or ('impact' in lc) or ('jet' in lc)) and
(col_map['X4'] is None):
        col_map['X4'] = c
    if ('yp' in lc and 'pv' in lc) and (col_map['X5'] is None):
        col_map['X5'] = c
    if (('10mgs' in lc) or ('gel' in lc)) and (col_map['X6'] is None):
        col_map['X6'] = c

# If any remain None, try common fallbacks manually
for k,v in col_map.items():
    if v is None:
        # try explicit names
        for candidate in ['Depth_ft',
'WOB_Klb','RPM','IF_lb','Yp_PV_ratio','10MGS_10SGS']:
            if candidate in df.columns:
                # map to first empty slot in order
                col_map[k] = candidate
                break

print("Using column mapping:", col_map)
if any(v is None for v in col_map.values()):
    raise RuntimeError("Could not find all 6 input columns
automatically. Columns available: " + str(list(df.columns)))

# extract arrays (ensure floats)
X1 = df[col_map['X1']].astype(float).values
X2 = df[col_map['X2']].astype(float).values
```

```
X3 = df[col_map['X3']].astype(float).values
X4 = df[col_map['X4']].astype(float).values
X5 = df[col_map['X5']].astype(float).values
X6 = df[col_map['X6']].astype(float).values
y_raw = df[rop_col].astype(float).values
y = np.log(y_raw)   # ln(ROP)

# --- Build design matrix X (n x p) ---
cols = []
names = []

# intercept
cols.append(np.ones(n)); names.append('Intercept')

# linear
for i, arr in enumerate([X1,X2,X3,X4,X5,X6], start=1):
    cols.append(arr); names.append(f'X{i}')

# squared
for i, arr in enumerate([X1,X2,X3,X4,X5,X6], start=1):
    cols.append(arr**2); names.append(f'X{i}2')

# interactions (i<j)
for i in range(1,7):
    for j in range(i+1,7): cols.append((eval(f"X{i}")
        * eval(f"X{j}")));
names.append(f'X{i}X{j}')

X = np.column_stack(cols)
p = X.shape[1]
print("Design matrix X shape:", X.shape, "p =", p)
```

```
# --- Normal equations: beta = (X^T X)^-1 X^T y ---
XtX = X.T.dot(X)
# invert or pseudo-inverse
try:
    XtX_inv = np.linalg.inv(XtX)
except np.linalg.LinAlgError:
    XtX_inv = np.linalg.pinv(XtX)
    print("Used pseudo-inverse for XtX")

Xty = X.T.dot(y)
beta = XtX_inv.dot(Xty)  # (p,)

# predictions and residuals
yhat = X.dot(beta)
resid = y - yhat
SSE = float(resid.T.dot(resid))
df_resid = n - p
MSE = SSE/df_resid

# covariance of beta
cov_beta = MSE * XtX_inv
se_beta = np.sqrt(np.abs(np.diag(cov_beta)))  # avoid tiny negatives
due to numerical noise

# t-stats and p-values (try t distribution)
try:
    from scipy import stats
    t_stats = beta / se_beta
    p_values = 2 * stats.t.sf(np.abs(t_stats), df_resid)
except Exception:
```

```
    print("scipy not available: using normal approx for p-values")
    from math import erfc
    t_stats = beta / se_beta
    p_values = np.array([erfc(abs(t)/np.sqrt(2)) for t in t_stats])

# goodness of fit
SST = float(((y - y.mean())**2).sum())
R2 = 1 - SSE/SST
R2_adj = 1 - (1-R2)*(n-1)/(df_resid)
RMSE_ln = sqrt(MSE)
pred_ROP = np.exp(yhat)
RMSE_ROP = np.sqrt(np.mean((y_raw - pred_ROP)**2))

# --- Prepare coefficient table and save ---
coef_df = pd.DataFrame({
    'term': names,
    'beta': beta,
    'std_err': se_beta,
    't_stat': t_stats,
    'p_value': p_values
})
coef_df.to_csv(OUT_COEFFS, index=False)
np.savetxt(OUT_COLNAMES, np.array(names), fmt='%s')
np.savetxt(OUT_YHAT, yhat, delimiter=',')

# latex table (rounded)
coef_df2 = coef_df.copy()
coef_df2[['beta','std_err','t_stat','p_value']] =
coef_df2[['beta','std_err','t_stat','p_value']].round(6)
with open(OUT_TEX,'w') as f:
    f.write("% LaTeX table: RSM full-model coefficients\n")
```

```
    f.write("\\begin{tabular}{lrrrr}\n\\hline\nTerm & Beta & StdErr &
t & p \\\\\n\\hline\n")
    for _, row in coef_df2.iterrows():
        f.write(f"{row['term']} & {row['beta']} & {row['std_err']} &
{row['t_stat']} & {row['p_value']} \\\\\n")
    f.write("\\hline\n\\end{tabular}\n")

# print summary (atomic outputs)
print("\n--- Atomic outputs ---")
print("n =", n, ", p =", p, ", df_resid =", df_resid)
print("SSE =", SSE)
print("MSE =", MSE)
print("R^2 =", R2)
print("Adj R^2 =", R2_adj)
print("RMSE ln-domain =", RMSE_ln)
print("RMSE ROP-domain =", RMSE_ROP)
print("Saved coeffs to:", OUT_COEFFS)
print("Saved yhat to:", OUT_YHAT)
print("Saved latex table to:", OUT_TEX)
```

E . SOURCE CODE USED FOR BACKWARD ELIMINATION

```
# backward_elimination_fixed.py
# Use in same folder as your Niger_Delta_Normalized.csv
import os, numpy as np, pandas as pd
from math import sqrt, erfc

BASE = "/storage/emulated/0/Chapter 3 Methodology/"
INFILE = os.path.join(BASE, "Niger_Delta_Normalized.csv")
OUT_FINAL = os.path.join(BASE,
"RSM_backward_elimination_coeffs_fixed.csv")
OUT_LOG = os.path.join(BASE, "BE_iteration_log.csv")
```

```
df = pd.read_csv(INFILE)
# keep only rows with positive ROP (as before)
rop_col = [c for c in df.columns if 'rop' in c.lower()][0]
df = df[df[rop_col] > 0].copy()
n = len(df)

# map inputs (adjust if your column names differ)
cols_map = {
    'X1': 'Depth_ft',
    'X2': 'WOB_Klb',
    'X3': 'RPM',
    'X4': 'IF_lb',
    'X5': 'Yp_PV_ratio',
    'X6': '10MGS_10SGS'
}
# build design matrix columns and names
parts = []
names = []

# Intercept (we will protect it from elimination)
parts.append(np.ones(n)); names.append('Intercept')

# linear terms
for i,(k,v) in enumerate(cols_map.items(), start=1):
    parts.append(df[v].values); names.append(f'X{i}')

# squared
for i,(k,v) in enumerate(cols_map.items(), start=1):
    parts.append((df[v].values)**2); names.append(f'X{i}2')

# interactions
```

```
keys = list(cols_map.keys())
for i in range(len(keys)):
    for j in range(i+1, len(keys)):
        parts.append(df[cols_map[keys[i]]].values *
df[cols_map[keys[j]]].values)
        names.append(f'{keys[i]}{keys[j]}')

X_full = np.column_stack(parts)
y = np.log(df[rop_col].values)

def fit_stats(X, y):
    n, p = X.shape
    XtX = X.T @ X
    XtX_inv = np.linalg.pinv(XtX)
    beta = XtX_inv @ X.T @ y
    yhat = X @ beta
    resid = y - yhat
    SSE = float(resid.T @ resid)
    df_resid = n - p
    MSE = SSE / df_resid
    covb = MSE * XtX_inv
    se = np.sqrt(np.abs(np.diag(covb)))
    t_stats = beta / se
    # p-values: try scipy, else normal approx
    try:
        from scipy import stats
        p_vals = 2 * stats.t.sf(np.abs(t_stats), df_resid)
    except Exception:
        # normal approx (ok for large df)
        z = np.abs(t_stats)
        p_vals = np.array([erfc(zv/np.sqrt(2)) for zv in z])
```

```
    # R2
    SST = float(((y - y.mean())**2).sum())
    R2 = 1 - SSE / SST
    adjR2 = 1 - (1-R2)*(n-1)/(df_resid)
    return {
        'beta': beta, 'se': se, 't': t_stats, 'p': p_vals,
        'R2': R2, 'adjR2': adjR2, 'SSE': SSE, 'MSE': MSE,
        'yhat': yhat, 'resid': resid, 'p': p_vals
    }

# Initialize active indices (we will protect index 0 = Intercept)
active_idx = list(range(X_full.shape[1]))  # indices into names
# but create a mask or list we won't remove intercept (0)
iteration_log = []
threshold = 0.05
iteration = 0

while True:
    iteration += 1
    X_active = X_full[:, active_idx]
    stats = fit_stats(X_active, y)
    pvals = stats['p']
    # we must not consider intercept (position 0 in active list)
    # find index of the worst p among REMOVABLE terms (exclude
intercept position)
    # find positions in active_idx where corresponding name !=
'Intercept'
    removable_positions = [i for i, idx in enumerate(active_idx) if
names[idx] != 'Intercept']
    if not removable_positions:
        # nothing to remove
```

```
        break
    # get p-values for these positions
    pvals_removable = [(pos, pvals[pos]) for pos in
removable_positions]
    # find the maximum p and its position (in current active indexing)
    worst_pos, worst_p = max(pvals_removable, key=lambda t: t[1])
    worst_term_name = names[active_idx[worst_pos]]
    # log iteration
    iteration_log.append({
        'iteration': iteration,
        'n_terms_before': len(active_idx),
        'removed_term': worst_term_name,
        'removed_p': float(worst_p),
        'R2_before': float(stats['R2'])
    })
    print(f"Iteration {iteration}: dropping {worst_term_name}
(p={worst_p:.6f}), R2={stats['R2']:.6f}")
    # stop if worst_p <= threshold (all remaining <= threshold)
    if worst_p <= threshold:
        print("All remaining terms have p <= threshold. Stopping.")
        break
    # remove that position from active_idx
    del active_idx[worst_pos]
    # loop continues

# compute final stats for reporting
X_final = X_full[:, active_idx]
final_stats = fit_stats(X_final, y)
final_beta = final_stats['beta']
final_terms = [names[i] for i in active_idx]
```

```
# save iteration log and final coeffs
log_df = pd.DataFrame(iteration_log)
if log_df.shape[0] > 0:
    log_df.to_csv(OUT_LOG, index=False)
final_df = pd.DataFrame({'term': final_terms, 'beta': final_beta})
final_df.to_csv(OUT_FINAL, index=False)

print("\nFinal model terms (kept):", final_terms)
print("Final R2:", final_stats['R2'], "AdjR2:", final_stats['adjR2'])
print("Saved iteration log to:", OUT_LOG)
print("Saved final coefficients to:", OUT_FINAL)
import statsmodels.api as sm
import pandas as pd

# Assuming X_final and Y are your final design matrix and response
variable:
model = sm.OLS(Y, X_final).fit()
results = pd.DataFrame({
    'Coefficient': model.params,
    'Std_Error': model.bse,
    't_value': model.tvalues,
    'p_value': model.pvalues
})
print(results)
results.to_csv('RSM_t_test_results.csv', index=True)
F. SOURCE CODE USED FOR REGRESSION COEFFICIENT SIGNIFICANCE
import numpy as np
from scipy import stats

# === Input your regression results ===
# Example: coefficients from your reduced RSM model
```

```
# ln(ROP) = -0.2427 - 2.1677*X1^2 + 0.9009*X1*X3
coefficients = np.array([-0.2427, -2.1677, 0.9009])

# Standard errors for each coefficient (example placeholders)
# You can adjust these slightly if you have them from Python output
standard_errors = np.array([0.025, 0.220, 0.100])

# Sample size and predictors
n = 449     # number of observations
p = 3       # number of coefficients (intercept + 2 predictors)

# === Compute t-values ===
t_values = coefficients / standard_errors

# === Compute p-values (two-tailed test) ===
df = n - p
p_values = 2 * (1 - stats.t.cdf(np.abs(t_values), df))

# === Display results ===
print("Coefficient\tStdError\t t-value\t p-value")
for i in range(len(coefficients)):

print(f"{coefficients[i]:>10.5f}\t{standard_errors[i]:>8.5f}\t{t_value
s[i]:>8.3f}\t{p_values[i]:>10.5f}")

# === Significance interpretation ===
alpha = 0.05
significant = p_values < alpha
print("\nSignificance (p ≤ 0.05):", significant)
```

G. SOURCE CODE USED FOR RSM VALIDATION

```
import os
import pandas as pd
import numpy as np

# ------------------------------------------------------------
# Basic configuration
# ------------------------------------------------------------
BASE = "/storage/emulated/0/Chapter 3 Methodology/"
DATA_FILE = os.path.join(BASE, "Niger_Delta_Normalized.csv")
COEFF_FILE = os.path.join(BASE,
"RSM_backward_elimination_coeffs_fixed.csv")

print("🔍 Checking files...") print("DATA_FILE:",
os.path.exists(DATA_FILE)) print("COEFF_FILE:",
os.path.exists(COEFF_FILE))

# ------------------------------------------------------------
# STEP 1: Load data
# ------------------------------------------------------------
data = pd.read_csv(DATA_FILE)
# Remove zero or negative ROP values (to avoid log(0))
data = data[data['ROP_ft_hr'] > 0].copy()
coeffs = pd.read_csv(COEFF_FILE)
print("✅ Data loaded:", data.shape)
print("✅ Coefficients loaded:\n", coeffs.head())

# ------------------------------------------------------------
# STEP 2: Prepare inputs
# ------------------------------------------------------------
Y_actual = np.log(data['ROP_ft_hr'])
```

```
X1 = data['Depth_ft']
X3 = data['RPM']
X12 = X1**2
X1X3 = X1 * X3

# Extract coefficients
Intercept = coeffs.loc[coeffs['term'] == 'Intercept',
'beta'].values[0]
b_X12 = coeffs.loc[coeffs['term'] == 'X12', 'beta'].values[0]
b_X1X3 = coeffs.loc[coeffs['term'] == 'X1X3', 'beta'].values[0]

print(" Coefficients read successfully")

# ------------------------------------------------------------
# STEP 3: Predictions
# ------------------------------------------------------------
Y_pred_ln = Intercept + b_X12*X12 + b_X1X3*X1X3
Y_pred = np.exp(Y_pred_ln)
Y_actual_linear = data['ROP_ft_hr']

print(" Predictions complete")

# ------------------------------------------------------------
# ------------------------------------------------------------
# STEP 4: Validation metrics
# ------------------------------------------------------------
n = len(Y_pred)
p = len(coeffs)
SSE = np.sum((Y_actual_linear - Y_pred)**2)
SST = np.sum((Y_actual_linear - np.mean(Y_actual_linear))**2)
```

```
R2 = 1 - SSE/SST
AdjR2 = 1 - (1 - R2)*(n - 1)/(n - p - 1)
RMSE = np.sqrt(np.mean((Y_pred - Y_actual_linear)**2))
MAPE = np.mean(np.abs((Y_actual_linear - Y_pred)/Y_actual_linear))*100

print("\n--- VALIDATION RESULTS ---")
print(f"R² = {R2:.4f}")
print(f"Adjusted R² = {AdjR2:.4f}")
print(f"RMSE = {RMSE:.4f}")
print(f"MAPE = {MAPE:.2f}%")
print("----------------------------------------")

# make sure Pydroid keeps console open
input("\n Press Enter to close after noting down the numbers.  ")
input("\nPress Enter to close.  ")
```

H. SOURCE CODE USED FOR BAT OPTIMIZATION

```
# ============================================
# BAT ALGORITHM FOR OPTIMIZATION OF ROP
# Based on reduced RSM model:
# ln(ROP) = -0.2427 - 2.1677*X1^2 + 0.9009*X1*X3
# X1 = normalized Depth, X3 = normalized RPM
# ============================================

import math
import random
import pandas as pd
import os

# ----------------------------
```

```
# 🟦 Initialization
# ----------------------------
random.seed(42)   # reproducible results

N = 25              # number of bats
fmin, fmax = 0, 2   # frequency range
A0 = 0.25           # initial loudness
r0 = 0.5            # initial pulse rate
alpha = 0.9         # loudness decay rate
gamma = 0.9         # pulse rate increase rate
tol = 1e-5          # stopping tolerance
max_iter = 200      # maximum iterations

lower, upper = 0.0, 1.0  # normalized bounds

# Physical mapping for interpretation
DEPTH_RANGE = (1000, 9000)   # ft
RPM_RANGE = (80, 180)        # rev/min

# ----------------------------
# | RSM model & Objective Function
# ----------------------------
def RSM_ln_ROP(X1, X3):
    return -0.2427 - 2.1677*(X1*2) + 0.9009(X1*X3)

def RSM_ROP(X1, X3):
    return math.exp(RSM_ln_ROP(X1, X3))

def objective(x1, x3):
    """Objective: minimize -ROP (maximize ROP)"""
```

```
    return -RSM_ROP(x1, x3)

# ----------------------------
# 🛢️ Helper Functions
# ----------------------------
def clamp(value):
    """Ensure values remain within [0,1]"""
    return max(lower, min(upper, value))

def denorm_depth(X1):
    return DEPTH_RANGE[0] + X1 * (DEPTH_RANGE[1] - DEPTH_RANGE[0])

def denorm_rpm(X3):
    return RPM_RANGE[0] + X3 * (RPM_RANGE[1] - RPM_RANGE[0])

# ----------------------------
# 🌀 Initialize population
# ----------------------------
bats = [[random.random(), random.random()] for _ in range(N)]
vel = [[0.0, 0.0] for _ in range(N)]
freq = [0.0]*N
fitness = [objective(x1, x3) for x1, x3 in bats]

best_idx = fitness.index(min(fitness))
best = bats[best_idx].copy()
best_fit = fitness[best_idx]

history = []

# ----------------------------
```

```
# | Main Bat Algorithm Loop
# ----------------------------
for t in range(max_iter):
    for i in range(N):
        freq[i] = fmin + (fmax - fmin) * random.random()

        # Update velocity and position
        for d in range(2):
            vel[i][d] += (bats[i][d] - best[d]) * freq[i]
            bats[i][d] = clamp(bats[i][d] + vel[i][d])

        # Local random walk (exploration)
        if random.random() > r0:
            eps = random.uniform(-1, 1) bats[i][0]
            = clamp(best[0] + eps * A0) bats[i][1]
            = clamp(best[1] + eps * A0)

        f_new = objective(bats[i][0], bats[i][1])

        # Accept solution if improved
        if (f_new <= fitness[i]) and (random.random() < A0):
            bats[i] = bats[i].copy()
            fitness[i] = f_new
            A0 *= alpha
            r0 = r0 * (1 - math.exp(-gamma * t))

        # Update best
        if f_new < best_fit:
            best_fit = f_new
            best = bats[i].copy()
```

```
    # Save history (for plotting later)
    history.append([t+1, best[0], best[1], -best_fit])

    # Stopping check
    if t > 5 and abs(history[-1][3] - history[-2][3]) < tol:
        break

# ----------------------------
# 🔵 Final Results
# ----------------------------
best_norm = best
best_ROP = -best_fit
best_depth = denorm_depth(best_norm[0])
best_rpm = denorm_rpm(best_norm[1])

print("🟩 Optimization Complete!")
print(f"Iterations run: {t+1}")
print(f"Best normalized (X1, X3): ({best_norm[0]:.4f},
{best_norm[1]:.4f})")
print(f"→ Optimum Depth: {best_depth:.2f} ft")
print(f"→ Optimum RPM: {best_rpm:.2f} rev/min")
print(f"Predicted Maximum ROP: {best_ROP:.4f} (normalized units)")

# ----------------------------
# 🟦 Save results
# ----------------------------
out_dir = "/storage/emulated/0/Chapter 3 Methodology"
os.makedirs(out_dir, exist_ok=True)
```

```
hist_df = pd.DataFrame(history, columns=["Iteration", "X1", "X3",
"ROP_pred"])
hist_df.to_csv(os.path.join(out_dir, "BA_results.csv"), index=False)

summary = f"""BAT ALGORITHM OPTIMIZATION RESULTS
Iterations run: {t+1}
Best normalized X1, X3: ({best_norm[0]:.4f}, {best_norm[1]:.4f})
→ Optimum Depth: {best_depth:.2f} ft
→ Optimum RPM: {best_rpm:.2f} rev/min
Predicted Maximum ROP: {best_ROP:.4f} (normalized units)
"""

with open(os.path.join(out_dir, "BA_summary.txt"), "w") as f:
    f.write(summary)

print(f" Results saved to {out_dir}")

I. SOURCE CODE FOR RSM CONTOUR AND 3D SURFACE PLOT
# ===== RSM Contour and 3D Surface Plots (Fixed for Pydroid) =====
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

# ---- Model coefficients ----
b0 = -0.2427
b1_sq = -2.1677
b13 = 0.9009

# ---- Create normalized grid ----
X1 = np.linspace(0, 1, 100)   # normalized Depth
```

```
X3 = np.linspace(0, 1, 100)   # normalized Rotary Speed
X1, X3 = np.meshgrid(X1, X3)

# ---- Predicted ln(ROP) and ROP ----
ln_ROP = b0 + b1_sq * X1**2 + b13 * X1 * X3
ROP = np.exp(ln_ROP)

# ---- Figure 4.3: Contour Plot ----
plt.figure(figsize=(7, 5))
contour = plt.contourf(X1, X3, ROP, cmap='viridis', levels=20)
plt.colorbar(contour, label='Predicted ROP (ft/hr)')
plt.xlabel('Normalized Depth (X1)')
plt.ylabel('Normalized Rotary Speed (X3)')
plt.title('Figure 4.3: Contour Plot of ROP vs Depth and Rotary Speed')
plt.tight_layout()
plt.savefig('/storage/emulated/0/Chapter 3
Methodology/Figure_4_3_Contour_Plot.png', dpi=300)
plt.close()  # close figure before next one

# ---- Figure 4.4: 3D Surface Plot ----
fig = plt.figure(figsize=(7, 5))
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(X1, X3, ROP, cmap='viridis', edgecolor='none')
ax.set_xlabel('Normalized Depth (X1)')
ax.set_ylabel('Normalized Rotary Speed (X3)')
ax.set_zlabel('Predicted ROP (ft/hr)')
ax.set_title('Figure 4.4: 3D Surface Plot of ROP vs Depth and Rotary
Speed')
plt.tight_layout()
```

```
plt.savefig('/storage/emulated/0/Chapter 3
Methodology/Figure_4_4_Surface_Plot.png', dpi=300)
plt.show()
```

J. SOURCE CODE FOR SENSITIVITY ANALYSIS

```
# --- Sensitivity Analysis for ROP Model ---

# Given from ANOVA and reduced model
SSR_total = 171.52  # total regression sum of squares

# Partial sum of squares for each parameter (you can adjust these if
needed)
SS_values = {
    "Depth (X1²)": 116.64,
    "Depth × RPM (X1X3)": 54.88,
    "WOB (X2)": 0.00,
    "Impact Force (X4)": 0.00,
    "YP/PV Ratio (X5)": 0.00,
    "Gel Strength Ratio (X6)": 0.00
}

# Compute percentage contribution using Eq. (3.21)
# P_i = (SS_i / SSR_total) × 100
print("\n--- Sensitivity Analysis Results ---")
for param, ss in SS_values.items():
    contribution = (ss / SSR_total) * 100
    print(f"{param:<25} : {contribution:.2f} %")

# Optional: save to a text file
save_path = "/storage/emulated/0/Chapter 3
Methodology/Sensitivity_Analysis.txt"
```

```
with open(save_path, "w") as f:
    f.write("Parameter, Contribution (%)\n")
    for param, ss in SS_values.items():
        contribution = (ss / SSR_total) * 100
        f.write(f"{param}, {contribution:.2f}\n")

print(f"\n Results saved to: {save_path}")
```

K. SOURCE CODE FOR CUMULATIVE ROP DISTRIBUTION

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# --- Step 1: Load the datasets ---
df_actual = pd.read_csv("/storage/emulated/0/Chapter 3
Methodology/Niger_Delta_Normalized.csv")
df_pred = pd.read_csv("/storage/emulated/0/Chapter 3
Methodology/RSM_y_hat.csv")

# --- Step 2: Extract and sort the ROP values ---
actual = np.sort(df_actual["ROP_ft_hr"].values)
predicted = np.sort(df_pred.iloc[:, 0].values)  # first column is
predicted ln(ROP)

# --- Step 3: Compute cumulative probability for each dataset ---
p_actual = np.linspace(0, 1, len(actual))
p_pred = np.linspace(0, 1, len(predicted))

# --- Step 4: Plot the cumulative distribution curves ---
plt.figure(figsize=(7, 5))
```

```
plt.plot(actual, p_actual, label="Actual ROP", color="blue",
linewidth=2)
plt.plot(predicted, p_pred, label="Predicted ROP", color="red",
linestyle="--", linewidth=2)
plt.xlabel("Rate of Penetration (ft/hr)")
plt.ylabel("Cumulative Probability")
plt.title("Figure 4.X: Cumulative Distribution of Actual and Predicted
ROP")
plt.legend()
plt.grid(True, linestyle="--", alpha=0.7)
plt.tight_layout()
plt.savefig("/storage/emulated/0/Chapter 3
Methodology/Cumulative_ROP_Distribution.png", dpi=300)
plt.show()
```