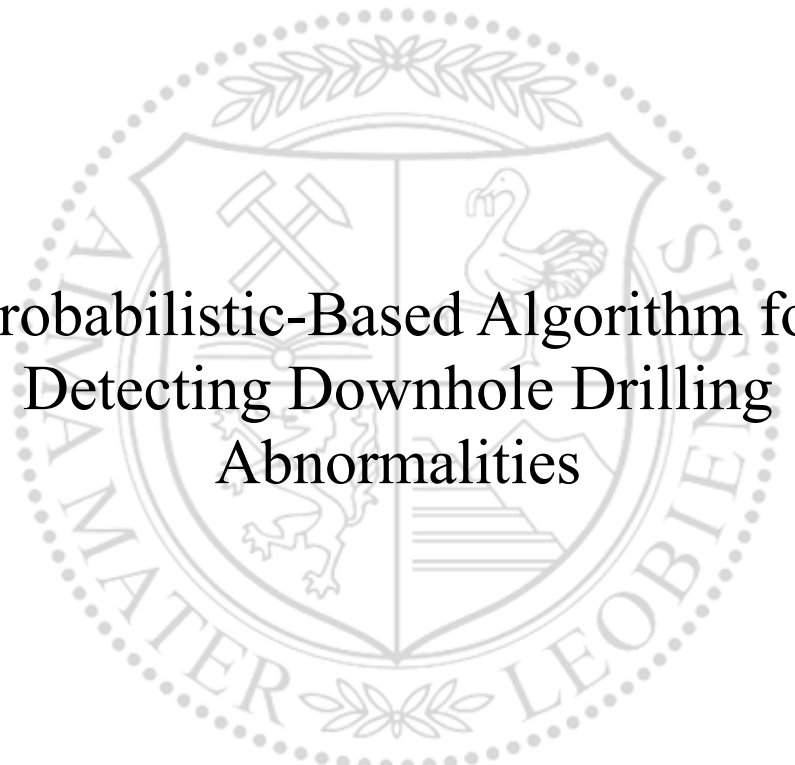Chair of Drilling and Completion Engineering

# Master's Thesis

# Probabilistic-Based Algorithm for Detecting Downhole Drilling Abnormalities

Philipp Grasser, BSc

January 2024

**MONTANUNIVERSITÄT LEOBEN**
www.unileoben.ac.at

## EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich diese Arbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt, den Einsatz von generativen Methoden und Modellen der künstlichen Intelligenz vollständig und wahrheitsgetreu ausgewiesen habe, und mich auch sonst keiner unerlaubten Hilfsmittel bedient habe.

Ich erkläre, dass ich den Satzungsteil „Gute wissenschaftliche Praxis" der Montanuniversität Leoben gelesen, verstanden und befolgt habe.

Weiters erkläre ich, dass die elektronische und gedruckte Version der eingereichten wissenschaftlichen Abschlussarbeit formal und inhaltlich identisch sind.

Datum  30.01.2024

Unterschrift Verfasser/in
Philipp Grasser

Philipp Grasser
Master Thesis 2024
Petroleum Engineering / Geoenergy Engineering

# Probabilistic-Based Algorithm for Detecting Downhole Drilling Abnormalities

Supervisor: Gerhard Thonhauser
Co-supervisor: Sahar Keshavarz, Asad Elmgerbi

Chair of Drilling and Completion Engineering

*Dedicated to my parents and grandparents.*

# Acknowledgments

I would like to express my deepest gratitude to my thesis supervisor M.Sc. Sahar Keshavarz for her continuous guidance and the extraordinary support throughout this project. I am deeply indebted to M.Sc. Sahar Keshavarz for her tireless efforts providing constructive feedback and numerous discussions shaping the outcome of the thesis.

Furthermore, I would like to express my sincere appreciation to Dipl.-Ing. Dr.mont Asad Elmgerbi for his invaluable guidance and support throughout the process of writing this thesis and my academic journey. His dedication to education, his experience and his insightful contributions have been a true inspiration to me throughout my studies, significantly influencing my intellectual growth.

Eventually, I want to thank my friends which I made during my studies in Leoben. I am very grateful that you were part of my academic journey and I appreciate the memories we have created together.

Furthermore, I want to thank my beloved girlfriend Nadine for her continuous support throughout the last years of my academic journey. She has been a key pillar, in encouraging me to reach my set targets, broadening my horizon in multiple ways.

Finally, and most importantly, I want to thank my dear family – my parents, sister, and grandparents – for their enduring encouragement, support, and love. Their financial but moreover mental support throughout my studies laid the necessary foundation for my academic success. Thank you for never stop believing in me.

# Abstract

The oil and gas industry, like many others, is facing challenges brought about by the energy transition, demanding the optimization of operations within defined boundaries. In terms of geo-energy exploration related activities, this is associated with a cost-effective and safe drilling operation. In this regard, it is crucial to minimize the occurrence of undesired downhole problems, which may delay the drilling process, potentially causing non-productive time. One of the essential keys to achieving that is the early detection of anomalous downhole behaviour by continuously monitoring the surface-measured drilling parameters.

The hydraulic system, along with other key surface parameters, plays a crucial role in successful drilling operations. It not only facilitates the circulation of drilling fluids, hole cleaning, and bit colling, but also provides valuable insights of the current downhole condition. Accurate modelling and monitoring of the surface-measured standpipe pressure can serve as a reliable indicator of potential anomalous downhole behaviour. However, the conventional physics-based approach for modelling standpipe pressure faces limitation in accurately representing the dynamic and complex nature of the downhole condition.

Regarding this issue, the ultimate goal of this thesis is to develop a data-driven model based on a machine learning concept to predict standpipe pressure with only three controllable surface parameters as input for the model and still provide robust estimates of the target variable. The models are trained with trouble-free drilling data, which should allow the model to represent the normal trend and thus provide means for analysis and trend identification by comparison of the actual value with the modelled values.

In conclusion, the applied methodology and algorithm can provide acceptable estimates of the target variable utilizing minimum required datapoints stemming from the same well. However, an optimization of the applied approach can possibly lead to improved results. The provided confidence interval provides a range of possible values for the target variable, thus can be useful for analysis and real-time monitoring. However, the predicted confidence interval cannot be directly interpreted as a safe operation window.

# Zusammenfassung

Die Öl- und Gasindustrie steht, wie viele andere auch, vor den Herausforderungen der Energiewende und erfordert eine Optimierung der Abläufe innerhalb definierter Grenzen. Im Hinblick auf geoenergiebezogene Explorationsaktivitäten ist dies mit einem kostengünstigen und sicheren Bohrvorgang verbunden. In diesem Zusammenhang ist es von entscheidender Bedeutung, das Auftreten unerwünschter Bohrlochprobleme zu minimieren, die zu Verzögerungen oder unproduktiver Zeit führen. Einer der wesentlichen Schlüssel zum Erreichen dieses Ziels ist die frühzeitige Erkennung von anomalem Verhalten im Bohrloch durch eine kontinuierliche Überwachung der an der Oberfläche gemessenen Bohrparameter.

Das Hydrauliksystem spielt zusammen mit anderen wichtigen Oberflächenparametern eine entscheidende Rolle für erfolgreiche Bohrarbeiten. Es ermöglicht nicht nur die Zirkulation von Bohrflüssigkeiten, die Bohrlochreinigung und das Kühlen des Bohrkopfes, sondern liefert auch wertvolle Einblicke über den aktuellen Bohrlochzustand. Eine genaue Modellierung und Überwachung des an der Oberfläche gemessenen Standrohrdrucks kann als zuverlässiger Indikator für mögliches anomales Verhalten im Bohrloch dienen. Der herkömmliche physikbasierte Ansatz zur Modellierung des Standrohrdrucks stößt jedoch bei der genauen Darstellung der dynamischen und komplexen Natur der Bohrlochbedingungen an Grenzen.

In Bezug auf dieses Problem besteht das Ziel dieser Arbeit darin, ein datengesteuertes Modell zu entwickeln, das auf einem maschinellen Lernkonzept basiert, um den Standrohrdruck mit nur drei kontrollierbaren Oberflächenparametern als Eingabe für das Modell vorherzusagen und dennoch robuste Schätzungen der Zielvariablen bereitzustellen. Die Modelle werden mit störungsfreien Bohrdaten trainiert, die es dem Modell ermöglichen sollen, den normalen Trend darzustellen und somit Mittel zur Analyse und Trenderkennung durch Vergleich des tatsächlichen Werts mit den modellierten Werten bereitzustellen.

Zusammenfassend lässt sich sagen, dass die angewandte Methode und der Algorithmus akzeptable Schätzungen der Zielvariablen unter Verwendung der minimal erforderlichen Datenpunkte aus demselben Bohrloch liefern können. Jedoch kann eine Optimierung des angewandten Ansatzes möglicherweise zu besseren Ergebnissen führen. Das bereitgestellte Konfidenzintervall bietet einen Bereich möglicher Werte für die Zielvariable und kann daher für die Analyse und Echtzeitüberwachung nützlich sein. Das vorhergesagte Konfidenzintervall kann jedoch nicht direkt als sicheres Betriebsfenster interpretiert werden.

# Table of Contents

# Chapter 1

# Introduction

## 1.1  Background and Motivations

Energy extraction from geological prospects has become more challenging in terms of the remoteness of the sites and complex well designs, leading to higher overall costs, making it necessary for the drilling operation to be cost-effective. This challenge requires an optimized well construction and drilling process, which can be reached by optimizing the drilling efficiency to reduce non-productive time (NPT), and, subsequently, the overall costs of the operation. The implementation of smart solutions supporting the drilling process has gained a lot of attention in recent years and has become a measure to address this challenge (Zhong et al., 2022).

A modern rig is usually equipped with various sensors that provide real-time measurements of surface drilling parameters. Among these sensors are hook-load, standpipe pressure, flowrate, rotational speed, and torque, which are the key sensors for observing downhole drilling conditions and detecting undesired drilling events in a conventional well construction process. Monitoring these key drilling parameters is crucial for efficient and safe drilling operations. A deviation from standard behaviour could indicate an undesired downhole event, which may cause an interruption in the operation. Therefore, detecting abnormalities by comparing a modelled value with the real-time sensor measurement should provide early trend recognition. Based on that information, countermeasures can be taken to mitigate the possible undesired consequences.

The standpipe pressure reading can be used as an indicator for several common downhole problems. There are conventional models for predicting standpipe pressure, which are physics-based. These models cannot catch the dynamics of the drilling operation to a full extent and,

hence, are less suitable as a real-time solution in terms of efficiency and accuracy (Erge & van Oort, 2020). Therefore, data-driven models should fill the gap to catch the dynamic behaviour, discover hidden knowledge from the data, recognize patterns and trends, and make more accurate predictions of the main drilling parameters (Karthik et al., 2018).

From this perspective, this thesis focuses on developing a data-driven model to predict the standpipe pressure trend employing a supervised machine learning approach using Gaussian process regression.

## 1.2   Scope and Objectives

This thesis aims to develop a purely data-driven model using Gaussian process regression to predict and model the hydraulic trend during drilling by utilizing three controllable surface drilling parameters from the same well as input. Moreover, the thesis will investigate if a machine learning model based on Gaussian process regression can be applied to make accurate and robust predictions of the target variable of interest, aiming to provide not only a predicted mean but also a confidence interval for the predictions. The model is based on three controllable drilling surface parameters, reducing model complexity, and required input parameters.

To achieve the objectives, it was essential to perform a number of steps before the actual predictive modelling task. The research started with a comprehensive literature review of several necessary topics, such as a review of the of parameters of interest with a focus on standpipe pressure, downhole problem causes and consequences, and the concepts of machine learning to get a fundamental understanding of the overarching subject. Then, the focus shifted towards data analysis and data preparation. Therefore, following a machine learning workflow was necessary to correctly extract, analyse, and prepare the desired data to build the predictive model. The model should predict the expected value of the hydraulic trend during the actual drilling operation and be representative of the current trouble-free downhole condition to later allow for comparison between the actual and the modelled value. Therefore, extracting the information accordingly and classifying the datapoints based on different routine rig operations for analysis and modelling was crucial.

## 1.3   Novelty of Work

The thesis introduces a novel approach to predict standpipe pressure in drilling operations using Gaussian process regression. Unlike other machine learning methods that rely on historical data, the proposed approach utilizes real-time sensor data stemming from the same wellbore with a minimum number of required datapoints to provide accurate and up-to-date predictions of the hydraulic trend, hence enabling means to make decision making based on comparing the

actual sensor measurement with the predicted values of the model. This approach should not only eliminate the need for historical data but also significantly simplify the implementation process of the proposed method. Following this approach, the model should be able to make predictions of the hydraulic trend and be representative for the current downhole condition and at the same time capture unique characteristics of the current drilling process, including operational parameters, BHA characteristics, formation dip, trajectory, etc.

## 1.4 Thesis Structure

The thesis contains a literature review providing insights to the concepts of machine learning, downhole drilling problems and a comprehensive review of the modelled parameter standpipe pressure. It then carefully outlines the methodology employed for data preparation and analysis within a machine learning workflow. Finally, the outcomes of this methodology are presented and validated showcasing the application of this approach for three distinct cases, demonstrating the practical applicability.

# Chapter 2

# Literature Review

## 2.1 Chapter Overview

The following chapter introduces to three core topics that are crucial for the thesis. Starting with a brief introduction to machine learning and its systematic approach covering important terminologies and common practices used in a machine learning workflow. This is followed by an overview of common drilling problems, their causes, detection, and mitigation measures. Finally, a closer look is taken on the parameter of interest standpipe pressure and how the surface-measured standpipe pressure can be utilized to detect downhole abnormalities through monitoring and modelling.

## 2.2 Introduction to Machine Learning

Machine Learning (ML) is a subset of the extensive field of artificial intelligence, which describes the capacity of a system to learn from training data to automate the process of analytical model building without being explicitly programmed nor requiring predetermined equations as a model (MathWorks). Machine learning aims to learn information from the available data, where different algorithms and computational methods serve as tools to solve problems such as regression, classification, or clustering. Many different ML algorithms are available to solve data problems, but there is no universal solution (Mahesh, 2018). In order to enable adequate modelling, it is crucial to investigate which algorithm and model fits the underlying data and problem best. Therefore, there are different types of machine learning algorithms, and these can be classified in supervised, unsupervised, and reinforcement learning.

## 2.2.1 Supervised Learning

"Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs" (Mahesh, 2018). Supervised learning algorithms are used to solve either regression or classification tasks, depending on the problem, with the use of labelled datasets. The task in supervised learning is to fit a model that can relate the input to the output and enable to make predictions for new observations (Gareth et al., 2021). Therefore, the data needs to be partitioned into a training dataset, and a test dataset (Mahesh, 2018). For supervised learning the training dataset itself consists of inputs, which are also known as independent variables, features, predictors, or just variables and are usually denoted with the symbol $X$ (Hastie et al., 2008). In case of numerous features, the set of features can be extended to be of any size $p$ for which $X_1, X_2, …, X_P$ represents the number of features measured at $n$ observations. The desired output is also called target, response, dependent variable, or regressor across the literature and, in most cases, denoted as $Y$ (Gareth et al., 2021). The labelled input-output pairs from the training set are used to calibrate the open model parameters, which are referred as model training. When the training phase of the model is completed, it can generate predictions. Therefore, to make predictions, an unseen dataset, also called a test set, is required. New or unseen datapoints of input features $X$ are given to the trained model to make predictions of new target variable $Y$ (Janiesch et al., 2021). Depending on the problem and desired output, supervised machine learning algorithms are used for classification or regression. One of the main differences between classification and regression problems lies in their respective outputs. The task of a classification problem is normally to predict a discrete label or qualitative output, whereas the task of a regression problem lies in predicting a quantitative output (Hastie et al., 2008). There are numerous algorithms available which can be applied to solve this kind of problems. However, in the context of regression problems, some commonly used supervised machine learning algorithms comprise of linear regression, ensemble methods, decision trees, neural networks, support vector machines, and Gaussian Process Regression (GPR). Amongst these, the thesis emphasizes exploring the utilization of Gaussian process regression to solve the desired research task.

### 2.2.1.1 Gaussian Process Regression

Gaussian process regression models have a good ability to deal with uncertainty in a probabilistic framework, working well on small amount of data, are highly flexible, simple to implement and provide confidence intervals of the predictions (Leco & Kadirkamanathan, 2021; Quinonero-Candela & Rasmussen, 2005). Since there are several publications that focus on explaining the nature of GPR´s and Gaussian Processes (GPs), this part will now turn its

focus solely to the key concept (Bousquet et al., 2004; Rasmussen & Williams, 2006; Wang, 2022).

Gaussian process regression is a form of supervised machine learning in a probabilistic framework. Gaussian process models are non-parametric and assume that the function of interest is one realisation of a Gaussian process (Bachoc et al., 2020). Gaussian process models are fully specified by a mean function and covariance function, also called kernel function (Rasmussen & Williams, 2006). The choice of the kernel function determines the assumption about the underlying function´s behaviour. Covariance functions establish a framework for assessing the similarity between two function values (Duvenaud, 2014). Some commonly used kernel functions comprise of exponential kernel, squared-exponential kernel (also known as Radial Basis Function) and the Matérn kernel. The MATLAB environment provides a broad selection of different kernel functions, making it a well-suited tool for the research task. Also, in terms of hyperparameter optimization, MATLAB provides some helpful functionalities. There is the possibility of simultaneous hyperparameter optimization and training.

## 2.2.2  Unsupervised Learning

Unsupervised learning differentiates from supervised learning in a sense that there is no "teacher", and the input data has no labelled responses. The input or training data consists of a setting of features from $X_1$, $X_2$, …, $X_P$ with $n$ observations, but there is no interest in making predictions because there is no associated target variable $Y$ to predict (Gareth et al., 2021). Instead, unsupervised learning techniques aim to find patterns, intrinsic structures or groupings of data that share common properties. This function is commonly known as clustering. Besides that, unsupervised learning techniques are also applied to reduce the dimensionality of the data, for example using Principal Component Analysis (PCA) (Gareth et al., 2021). Besides PCA, there are numerous other unsupervised learning algorithms such as K-Means, Neural Networks or Gaussian Mixture models, which can be used to solve the unsupervised learning task (MathWorks). Pre-processing, data visualization, dimensionality reduction and clustering are main topics in unsupervised learning (Gareth et al., 2021).

## 2.2.3  Reinforcement Learning

Different to supervised and unsupervised learning is the technique of reinforcement learning, as it does not rely on predefined input-output pairs to train the model, but instead the system learns by interacting with the environment and making decisions based on the outcomes (Sutton & Barto, 1998). The system has a specified goal and a list of constraints and is trained based on the principle of trial and error to maximize reward. This task is done by rewarding desired actions and penalizing undesired behaviours and letting the model itself decide how to

maximize the reward (Janiesch et al., 2021). Table 1, taken and adapted from Kumar (2020), provides an overview of the three different types of machine learning, outlining their distinct features and differences.

*Table 1 – Supervised, unsupervised and reinforcement learning algorithms compared. Table adapted from (Kumar, 2020).*

| Feature | Supervised Learning | Unsupervised Learning | Reinforcement Learning |
|---|---|---|---|
| Definition | Learning by using labelled data | Learning by using unlabelled data | Learning with trial and error |
| Labelled data | yes | no | no |
| Supervision | Model is trained with labelled data, including input data and a corresponding target value | Model is trained with unlabelled data, no target values included | Model is trained through interaction with environment, where desirable actions are rewarded |
| Problem type | Regression and classification | Clustering, pattern recognition, dimensionality reduction | Exploration or exploitation |
| Algorithms | Linear Regression, Logistic Regression, Support Vector Machines, GPR, Decision Trees, Neural Networks, etc. | K-Means, Gaussian Mixture, Neural Networks, Hidden Markov Model, Hierarchical, etc. | Q-Learning, SARSA, etc. |

## 2.2.4 Machine Learning Workflow

To develop or deploy a data-driven model based on a machine learning concept, typically involves a structured process and requires several steps as depicted in Figure 1. The first step is data acquisition and definition of the problem. The following step is data preparation, one of the most important steps. Because the performance of ML models relies highly on the quality of the underlying data. Therefore, during this step, data analysis is performed by several means, such as visual data exploration and statistical methods, to get an overview of the data type, shape, and distribution. To end up with a clean dataset missing values have to be identified as well as undesired outliers have to be removed. Depending on the problem and the nature of the data, it can be beneficial to reshape the data and perform scaling in the form of standardization or normalization of identified features. For example, scaling is necessary if identified features heavily differ in range or several orders of magnitude (Brownlee, 2019). Proper pre-processing of the input data can also lead to enhanced performance in terms of computational speed (Bishop, 2006). Model selection comprise of selecting a particular algorithm for the model in accordance to solve the predefined problem. To train a machine learning model a segment of

the total dataset namely the training dataset is used to fit the model´s parameters. The remaining part of the data, the so-called test set is then used to test the model and assess the performance of it on unseen data (Gareth et al., 2021).
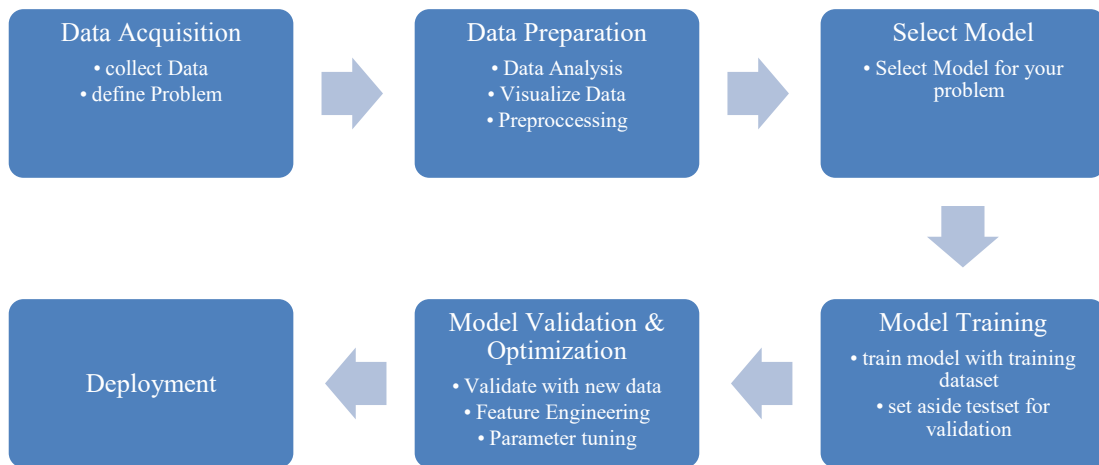


*Figure 1 – Common workflow of development of a machine learning model.*

During training, the model should find a fit between the input features ($X$) and the corresponding target ($Y$). The held-out data points are used to test the model later on and evaluate the fit. The test set has the same general form as the training dataset, containing the same number of features ($X_{Test}$) and the corresponding target variable ($Y_{Actual}$). The purpose of model testing is that only the features of the test dataset ($X_{Test}$) are given to the model to make predictions ($Y_{Pred}$). Then, comparing the actual value of the target variable ($Y_{Actual}$) with the predicted value from the model ($Y_{Pred}$) enables evaluation of the model performance. However, when evaluating model performance, a simple train-test split, as mentioned above, may not be suitable in all cases. Another approach is the so-called procedure of cross-validation (CV), or k-fold cross-validation. For k-fold cross-validation, a test set is held out to evaluate unseen data. The available training dataset is split in "k" equally sized folds or subsets of data samples. The model is trained and evaluated "k" times. Throughout this iterative process, k-1 folds are utilized for training and optimization of the model, and the remaining fold is reserved as a test dataset. This process is repeated "k" times, ensuring that each unique combination of train-test folds is used for training and evaluating the model once. (Lyashenko & Jha, 2022). The final performance is often summarized by calculating the average of the performance indicators obtained across "k" iterations (scikit-learn, 2023). Figure 2 shows the approach of a 5-fold CV, where in the first iteration, fold #1 is used for testing, and folds #2-5 are used for model training. Therefore, this thesis applies k-fold cross-validation to assess and represent model performance results.
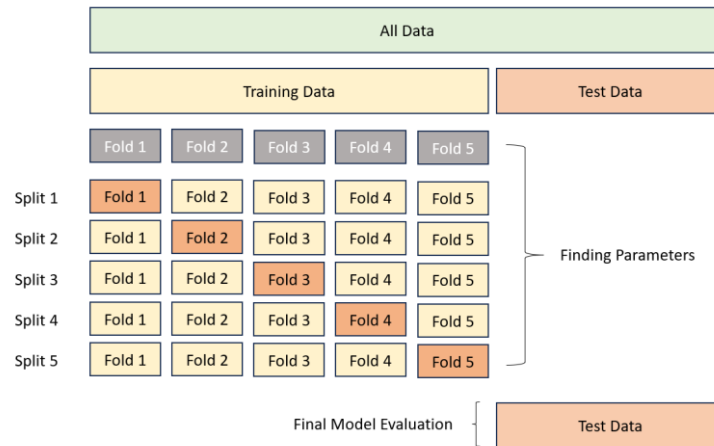
*Figure 2 – Visual representation of 5-fold cross-validation technique. Adapted from (scikit-learn, 2023).*

## 2.3 Downhole Drilling Problems

Various problems can occur during well construction related to the drilled formation, the wellbore, drilling equipment, technical operational parameters, or difficult downhole situations. These problems can result in an unwanted and unplanned interruption of the drilling process, causing downtime and non-productive time (NPT). According to Forshaw et al., 30% of total upstream oil and gas costs are related to NPT, whereas half of these account for downhole drilling problems (Forshaw et al., 2020). Thus, knowing the symptoms to identify downhole problems as early as possible or before they happen can have a crucial impact on well construction costs. Therefore, the following paragraphs provide an overview of the most common downhole-related drilling problems, their causes, and their operational consequences. At this point, it needs to be stated that the following list of problems does not cover all problematic events which may happen during a drilling operation.

### 2.3.1 Stuck Pipe

A pipe is considered to be stuck when the upward and downward movement of the drillstring is not possible, nor rotation or circulation, and the retrieval of the string is not feasible without exceeding the maximum allowable tensile strength of the string (Alshaikh et al., 2018). There are various reasons for the string to get stuck, such as differential pressure sticking, unstable formations, accumulation of cuttings due to poor hole cleaning, inadequate drilling fluids, or complex borehole features. Stuck pipe events may result in time-consuming fishing operations or necessary side-tracking, causing significant NPT. The NPT related to stuck pipe events can account for 25% of total NPT during well construction (Muqeem et al., 2012), thus having a

notable effect on the total costs. A stuck pipe can be classified into differential pressure and mechanical sticking, as discussed below.

### 2.3.1.1 Differential Pressure Sticking

Differential pressure sticking can occur across permeable zones if there is a pressure differential between the borehole and the formation. High permeable zones are more prone to cause differential pressure sticking (Mitchell & Miska, 2011). The pressure differential is caused when the mud hydrostatic is greater than the formation pore pressure at the corresponding interval, causing increased forces to act on the drillstring and force the drillstring against the wellbore wall into the mud cake. The filter cake usually has a high friction coefficient, resulting in a higher force required to pull the string across the area where the drillstring is embedded in the mud cake. Figure 3, taken from Alshaikh et al. (2018) shows the top and side view of stuck BHA due to differential sticking. When the drillstring is stuck due to differential pressure, no upward or downward movement or rotation is usually possible, however, circulation is possible (Mitchell & Miska, 2011).
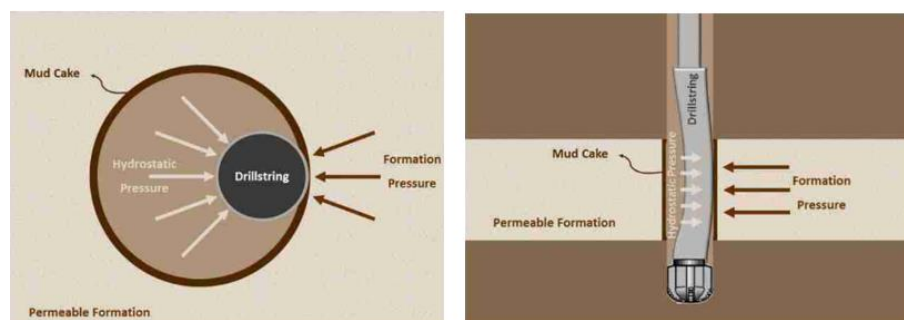


*Figure 3 – Differential pressure sticking of the BHA (Alshaikh et al., 2018).*

### 2.3.1.2 Mechanical Sticking Due to Bridging or Pack-Off

Pack-off or bridging is a type of mechanical sticking that occurs when debris or drilled cuttings accumulate in the wellbore around the drillstring, thus restricting further upward or downward movement of the string and preventing circulation of the drilling fluid. A Pack-off situation occurs when small debris plugs the annular space between drillstring and the wellbore, as shown in Figure 4, while bridging refers to larger pieces of cuttings bridging the annular space (Alshaikh et al., 2018). Inadequate hole-cleaning practices often cause these types of mechanical sticking. One of the primary functions of the drilling fluid is to effectively transport drilled cuttings to the surface, and even when circulation is halted, cuttings suspension must be maintained. Ensuring cuttings suspension in the static fluid column is essential for preventing the accumulation of cuttings that could cause hole fill-up. Early signs of inadequate hole cleaning include increase torque and drag. Additionally, when drilling at high rate of penetration (ROP), drilling fluid rheology may alter due to increased cutting concentration,

leading to increased equivalent circulating density (ECD) and standpipe pressure (SPP) (Brankovic et al., 2021). Many recommended practices for proper hole cleaning exist. These include maintaining proper rheological properties of the drilling fluid, appropriate hydraulic management, eventually cleaning of the wellbore, controlling the ROP, and monitoring surface drilling parameters and their possible deviation (Alshaikh et al., 2018).
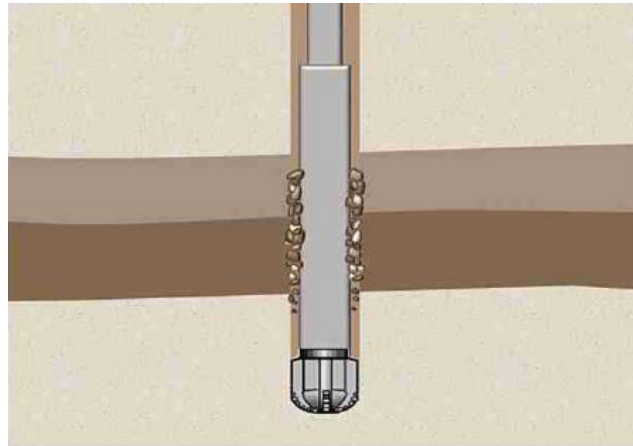


*Figure 4 – Hole pack-off (Alshaikh et al., 2018).*

### 2.3.1.3 Mechanical Sticking Due to Wellbore Geometry and Formation

Variations in the wellbore geometry, shape, and condition, as well as interactions between the drilling fluid and the formation can lead to mechanical sticking of the drill string or BHA components. A perfectly shaped wellbore does not practically exist. Some downhole irregularities, such as doglegs, ledges, or keyseats, may be created during normal drilling operations. A dogleg is caused due to the curvature of the wellbore when the trajectory of the wellbore changes in inclination and/or azimuth. At these doglegs, keyseats may eventually form. A keyseat can form when the drill pipe wears a slot into the borehole wall. This created slot has approximately the same diameter as the drill pipe and a smaller diameter than the borehole itself. When pulling the string upward, the larger diameter BHA can get stuck in such keyseats, requiring increased force (Mitchell & Miska, 2011).

Also, drilling fluid–formation interactions can cause a mechanical sticking problem. Reactive formations such as water-sensitive shales may swell and break when in contact with a water-based drilling fluid. This can result in either restricted diameter of the drilled hole or pack-off if larger parts of the reactive formation break and pack around the drillstring (Mitchell & Miska, 2011).

Regarding the above described reasons which can cause pipe sticking, there is a strong interest in the industry in predicting stuck pipe events through detection of early warning signs and correct interpretation these. Several approaches were taken to tackle this problem utilizing a

machine learning approach. A study by Alshaikh et al. (2019) investigated the use of different machine learning classification algorithms namely Artificial Neural Networks (ANNs), Decision Trees (DTs) and Support Vector Machines (SVMs), to detect early warnings signs of stuck pipe. The results of this study show that stuck pipe events could be accurately identified, and the model can also detect early warning signs in the drilling parameters, indicating impeding downhole problems which eventually may lead to a stuck pipe situation. The work published by Khan et al. (2020) compared the performance of an ANN and a SVM model in stuck pipe prediction. In their study they used data consisting of 268 datasets (123 stuck pipe and 145 non-stuck pipe) and 19 drilling parameters to train the models. The final models of this study show a high accuracy in predicting the two cases, stuck or non-stuck respectively, whereas the ANN model demonstrates better performance. In contrast to the two previous studies that employed supervised machine learning algorithms, the work published by Mopuri et al. (2022) presents a different approach based on an unsupervised learning algorithm. They built Autoencoders on Recurrent Neural Networks (RNNs) to model the normal drilling activity and detecting stuck pipe events as abnormal activity. In this study, field data from 30 wells was compiled and analysed. They demonstrated the applicability of this approach in early sign detection for pipe sticking events. Also, the authors mention that unsupervised learning techniques may be more appropriate for addressing the problem of early stuck sign detection.

## 2.3.2  Lost Circulation

Lost circulation is one of the most common downhole problems encountered during drilling operations, and it is characterized by a continuous loss of the drilling fluid or cement slurry to a formation, which, in the worst case, can result in a total loss of the drilling fluid column, leading not only to economic losses but moreover can result in severe well control situations (Mitchell & Miska, 2011).

Loss of circulation can happen during any routine drilling operation where the pumps are on. Two conditions must be met for lost circulation to occur. First, the formation must have a significant permeability with flow channels, which allow for fluid flow, and secondly, a positive pressure differential between the wellbore and the formation pore pressure is needed (Mitchell & Miska, 2011). High-permeable zones, natural fractured formations, caverns, or drilling-induced fractures can be possible loss zones. Figure 5, reprinted from Alkinani et al. (2020), illustrates possible formations for loss of circulation. Induced fractures can occur due to a high pressure differential between the wellbore and formation, and this can be the case either if the applied ECD is too large or during uncontrolled tripping operation, which both can be controlled by the driller to some extent. Induced fractures commonly appear at the weakest point of the open borehole, at the point where the formation fracture pressure is lowest. This

point is usually found below the last casing shoe (Skalle et al., 2013). However, caverns represent the most severe problem, and such cavernous structures are generally found in leached limestone formations. Most critical because caverns can be so large that a complete loss of circulation can occur (Mitchell & Miska, 2011).

Depending on the volume of fluid lost, lost circulation can be classified into seepage (up to 1 $m^3$/hr mud loss), partial losses (up to 10 $m^3$/hr mud loss), severe losses (more than 10 $m^3$/hr mud loss) and complete loss (Zhong et al., 2022).
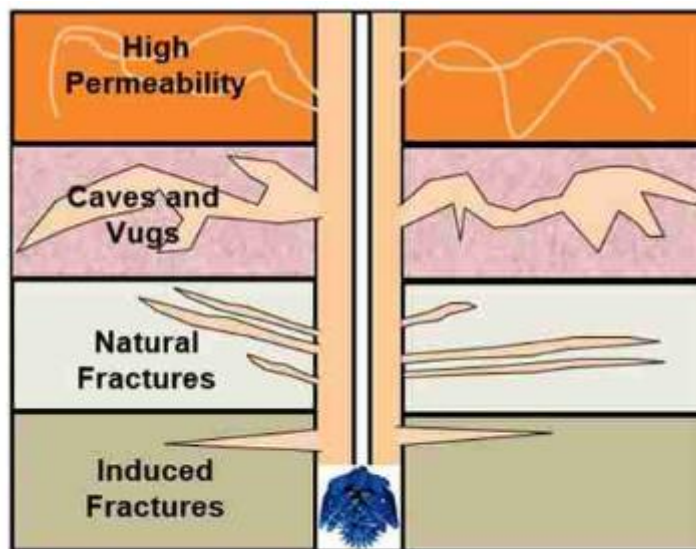


*Figure 5 – Possible loss zones (Alkinani et al., 2020).*

An ongoing loss of circulation situation substantiates on the surface with a decrease of flow out and a decrease in the mud pit level with a constant flowrate. When losses occur, drilling fluid flows into a formation. Depending on the severity of the loss situation, the hydrostatic pressure of the fluid column can decrease to a point where the hydrostatic pressure in the borehole is less than the formation pressure, leading to a well control situation (Islam et al., 2017). In the worst-case scenario, total loss of returns can even lead to a point where circulation is not possible anymore, making well control situations even more problematic.

Handling of lost circulation events happens in two ways. First, the risk of lost circulation is reduced by proper hydraulic pressure management and selecting proper mud systems for the formation to be drilled. However, losses still can occur and therefore the severity, and location of the loss zone need to be evaluated, followed by an appropriate treatment of the situation with lost circulation material (LCM) to seal off the zones where losses occur.

Machine learning approaches are taken to assist in preventing or predicting lost circulation problems. Alkinani et al. (2020) demonstrates the successful application of neural networks to predict mud loss. In this study two networks were created, one for natural fractures and one for

induced fractures respectively. The data used to train the two neural networks was compiled from over 3500 wells worldwide, resulting in high $R^2$ values of 0.956 and 0.925 for the natural fractures model and the induced fractures model. Further investigations were performed on unseen data (wells which were not used for creating the models), comprising of testing the network and their predictive power on 24 new wells. The results of the test show low errors, approximately one barrel per hour in predicting the actual mud loss. Thus, providing a generalized model, applicable worldwide, to estimate mud loss prior drilling.

### 2.3.3 Kick Situations

A well control problem occurs when there is an undesired flow of formation fluids into the wellbore, commonly referred to as a kick (Bourgoyne, Jr, A T et al., 1991). For a kick to happen, three conditions must be present at the same time, and these are as follows: the presence of a mobile fluid in a porous rock, sufficient permeability to allow fluid flow into the wellbore, and the pressure inside the wellbore must be less than the formation pore pressure (Mitchell & Miska, 2011). During an overbalanced drilling operation, the bottomhole pressure (BHP) is maintained to be above the formation pore pressure ($P_{form}$) and below the formation fracture pressure ($P_{frac}$) to prevent the influx of formation fluids and, at the same time, not to fracture the formation.

$$P_{form} < BHP < P_{frac}$$

*(2.1)*

The bottomhole pressure (BHP) itself is given by the Equation (2.2):

$$BHP = BHP_{hydrostatic} + BHP_{circulating} + BHP_{surge/swab}$$

*(2.2)*

Where $BHP_{hydrostatic}$ is the static bottomhole pressure without circulation. So, when no circulation is present, the BHP is equal to the mud hydrostatic. $BHP_{circulating}$ is the annular pressure loss that acts as a back pressure on the formation due to circulation against the direction of the flow. Tripping operation results in the additional pressure term $BHP_{surge/swab}$. Swabbing occurs during an upward movement of the drillstring, causing a reduction in BHP. On the other hand, surging substantiates increased BHP due to the downward movement of the drillstring (Islam et al., 2017).

Given the above relationships, a kick can occur when the formation pressure exceeds the BHP. This can be the case when drilling through formations with abnormal pressure, improper hole fill-up during tripping operation, insufficient mud weight, swab, surge during tripping, gas-cut mud, or lost circulation (Mitchell & Miska, 2011). During normal drilling operation there are

several primary or secondary kick indicators for investigating the situation. Primary indicators provide good knowledge about the situation: an increase in return flow rate with a constant flow-in rate and an increase in mud pit volume. Although secondary kick indicators do not directly indicate that a well control situation is jeopardizing, they have to be monitored cautiously and must not be overseen. Secondary kick indicators are drilling breaks (an increase of penetration rate), increased torque and drag, and change in cutting size. In case of a gas influx into wellbore, standpipe pressure may temporarily increase, followed by a gradual decrease of SPP with constant flowrate. This happens due to the lower density formation fluid entering the wellbore and displacing the heavier fluid inside the wellbore (Sule et al., 2018; Yin et al., 2021). In case of an ongoing kick event, several well control procedures exist to get the well under control again.

Kick detection techniques can be broadly categorized into three main approaches: conventional methods, integrated gas kick detection systems, and machine learning for automated kick detection. Conventional methods are utilizing primary and secondary indicators for kick detection as described above. Among integrated gas kick detection systems, Tang et al. (2019) proposes the utilization of two kick indicators, the drilling parameter group (DGP) and flow parameter group (FPG), respectively. The method can automatically analyse time series data and detect a kick event based on anomalous behaviour of these two indicators. Multiple studies demonstrate the applicability of machine learning approaches for early kick detection on laboratory scales with simulation data. The work from Nhat et al. (2020) demonstrates the application of a data-driven Bayesian Network, or Muojeke et al. (2020) shows the utilization of the simplest artificial neural network which would allow for kick detection. The work published by Yin et al. (2021) showcases the application of ML with actual field data, where through utilization of Long Short-Term Memory (LSTM), a sophisticated data preprocessing, and novel data labelling criterion promising results were achieved. The deployed LSTM model shows an accuracy of 87% in the testing dataset and could detect a gas kick earlier than the "Tank Volume" detection method in several case studies.

## 2.3.4  Drillstring Washout

A drillstring washout occurs when a leak in the drillstring develops and disturbs the pressure integrity of the string (Kuesters et al., 2020). This pipe washout can cause a partition of the drillstring, so-called twist-off, which can lead to a significant increase in NPT due to required fishing operation or even inevitable side-tracking.

During drilling operations, the drillstring and the BHA are subjected to challenging downhole and operational conditions including cyclic stresses, corrosive environments, and wear, which can contribute to mechanical fatigue. Small cracks can develop and result in a leaking drill pipe,

enabling a pathway for the drilling fluid to enter the annulus, as shown in Figure 6, reprinted from (Zhao et al., 2019). When unnoticed, these small cracks may enlarge, eventually leading to a twist-off. An ongoing washout is identified with standard surface drilling parameters, such as flowrate and standpipe pressure and determination of the hydraulic coefficient (Ambrus et al., 2018; Jeong et al., 2020). Depending on the severity of the washout, a significant drop of SPP with constant flowrate can be indicative for an ongoing washout event. However, due to the nature of the problem, a washout can also develop slowly. In this case, SPP may show a gradually decreasing trend over hours, making detection of a washout more complicated and challenging (Kuesters et al., 2020).
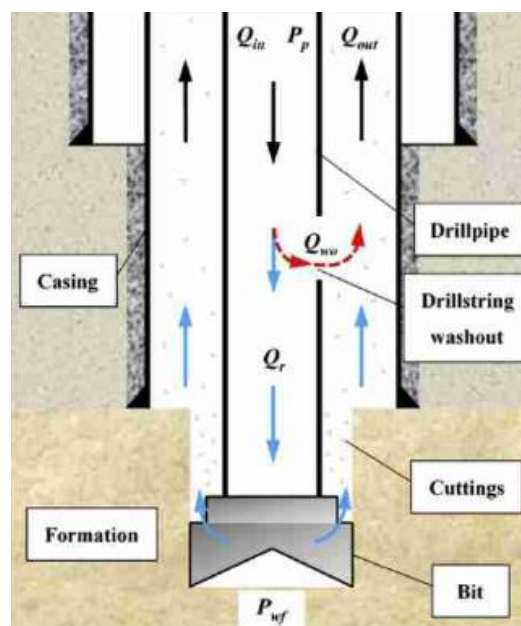


*Figure 6 – Schematic circulation path of drilling fluid in a drill string washout situation (Zhao et al., 2019).*

Different machine learning approaches exist in addressing the detection of drillstring washouts. The work published by Ambrus et al. (2018) utilizes machine learning to identify drillstring washouts and mud pump failures with the use of a Bayesian network. The network is able to aggregate real-time drilling sensor data (SPP, pump rate, flow-out, etc.), contextual information and predictions of the hydraulic trend. All these factors together cause a probabilistic belief system that suggests the presence of washout or mud pump failure situation. The model achieved to detect a failure hours before recognized by the rig crew. Kuesters et al. (2020) demonstrates the successful application of a Change Point Detection and Rules-Based Approach in early detection of washouts, and also assessed a Deep Learning Method utilizing a LSTM network for tackling the same problem with promising results. Zhao et al. (2019) proposes a method for early washout detection utilizing improved pattern recognition. The

model consists of three parts: a dynamic hydraulic model and standard washout modes, an adaptive pattern recognition algorithm, and synthetic probability method. The model uses standard rig measurements (SPP, pump rate and pit gain) as input. The model was tested in a real-world case study, where it successfully detected a drillstring washout 16 minutes earlier than the field detection result.

### 2.3.5 Bit Balling

Bit balling occurs when reactive formations adhere to the BHA, causing possible plugging of the bit nozzles accompanied by a significant reduction in drilling performance. Reactive formations, such as low permeability shales or clays, are predominantly causing this undesirable downhole problem. This is due to the nature of the drilled formation. In contact with water, the surface of clay particles can adsorb water, leading to an effect called hydration. This hydration process causes the particles to swell, and consequently followed by a dispersion of clay platelets into the surrounding water. In the presence of a continuous water supply, hydration and dispersion can continue, eventually causing dispersed clay platelets to adhere to the BHA (Kehinde et al., 2023). The effect of bit balling can be seen at the surface with a significant decrease in ROP, although all other operational parameters are kept constant, due to cuttings accumulating around the bit and BHA. Also, drilling torque will be lower and standpipe pressure will increase due to cuttings plugging the bit nozzles or the reduced annular space because of the adhesion of clay to the BHA.

## 2.4 Utilizing Standpipe Pressure as an Indicator for Downhole Issue Detection

The hydraulic circulating system is an essential part of any standard drilling rig, enabling the circulation of drilling fluids from the surface, down to the bit and up again. The main components of the hydraulic circulating system are the mud pumps, mud tanks, surface lines, drillstring and a solids control system. The drilling fluid moves (1) from the mud tanks to the pump, (2) from the pump through the surface lines to the drillstring, (3) through the drillstring down to the bit, (4) through the bit nozzles, (5) up to the surface through the annular space between drill string and open hole / cased hole, (6) to the solids control system and back to the mud tanks (Bourgoyne, Jr, A T et al., 1991) as shown in Figure 7.
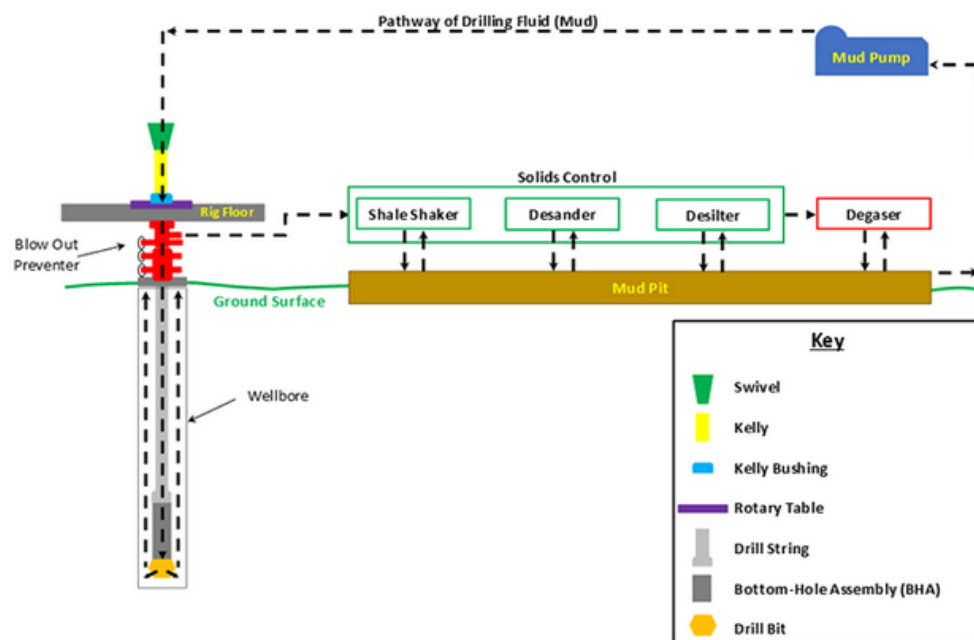


*Figure 7 – Hydraulic circulation system components and pathway of the drilling fluid (King, 2023).*

Some of the main functionalities the hydraulic circulating system and the drilling fluid provide are the removal of cuttings beneath the bit and transport of those to the surface in order to keep the hole clean, lubricate and cool the drill bit, and enable a pressure differential between the formation and the wellbore acting as a well barrier element. Besides that, monitoring and understanding of hydraulic surface parameters is from uppermost importance, because it provides crucial information about the current downhole condition. Amongst these hydraulic parameters, is the surface-measured standpipe pressure (SPP). Normally the standpipe pressure is measured at the high-pressure end of the pump (Erge & van Oort, 2022). When circulating a fluid through the hydraulic circuit, pressure drop takes place at various nodes in the system due to the friction between the fluid and the surface it is in contact with (Chowdhury et al., 2009b).

The pressure needed to circulate the fluid and overcome the friction is supplied by the mud pumps. Therefore, standpipe pressure (SPP) is the cumulative frictional pressure (Erge & van Oort, 2022) drop in the hydraulic circuit and equals the pump pressure (Chowdhury et al., 2009b). This relation is represented in Equation (2.3):

$$P_{pump} = \Delta P_{surface} + \Delta P_{Drill\ Pipe} + \Delta P_{BHA} + \Delta P_{Bit} + \Delta P_{Annulus/OH} + \Delta P_{Annulus/CH}$$

*(2.3)*

$P_{Pump}$ is the pump pressure, and equal to the total frictional pressure loss in the hydraulic system. $\Delta P_{Surface}$ is the pressure drop in the surface lines when fluid moves from the pump to the drill string. $\Delta P_{Drill\ Pipe}$ is the pressure drop in the drill pipe (DP). $\Delta P_{BHA}$ is the pressure drop in the Bottom Hole Assembly (BHA), including drill collars that may be larger than the normal drill pipe. The BHA can also be made up with additional components for directional control, downhole motors, or other special equipment such as Logging-While-Drilling (LWD) tools, for which the pressure drop within these tools must also be accounted for. The pressure differential across the bit is represented by $\Delta P_{Bit}$. This pressure drop is considered to be the only "useful" pressure drop and is carefully optimized in the well planning phase for proper selection of bit nozzles and mud pump dimensions in order to maximize efficiency in terms of ROP and hole cleaning (Chowdhury et al., 2009b). $\Delta P_{Annulus/OH}$ is the pressure drop in the annular space in the open hole section, where the different diameters of the DP and BHA need to be considered. The pressure drop in the annulus between the drill pipe and the casing is represented by $\Delta P_{Annulus/CH}$. Thus, Equation (2.3) can be simplified and rewritten to parasitic pressure losses, which represent the total system pressure losses plus the pressure drop across the bit, resulting in:

$$P_{Pump} = P_{Parasitic} + P_{Bit}$$

*(2.4)*

Now, any change in pump pressure ($P_{Pump}$) may be induced either by pressure losses in the system ($P_{Parasitic}$) or across the bit ($P_{Bit}$) or both – if operational drilling surface parameters such as flowrate and weight on bit are kept constant, indicating a change in the system and possible change in current downhole conditions.

Calculating and modelling these pressure drops across the whole system is not trivial, thus vital for a safe and successful drilling operation.

## 2.4.1  Conventional Standpipe Pressure Modelling

The conventional approach for calculating the frictional pressure loss in the hydraulic circuit is physic-based. These models are derived from initial principles describing fluid flow, such as

conservation of mass, energy, and momentum (Bourgoyne, Jr, A T et al., 1991). Since not all fluids behave the same way, there are different rheological models representing the behaviour of a moving fluid through a circular pipe and the typical rheological models to characterize fluid flow can be classified as Newtonian or non-Newtonian fluids. These rheological models establish and describe the relation between shear stress and shear rate of a moving fluid through a circular pipe (Chowdhury et al., 2009a). The most common rheological models are the Newtonian model, and the non-Newtonian fluid models, comprising of the Bingham Plastic model, the Power Law model, and the Herschel-Bulkley model (also known as the Yield Power Law model). The relationship between shear stress and shear rate for these rheological models is presented in Figure 8.
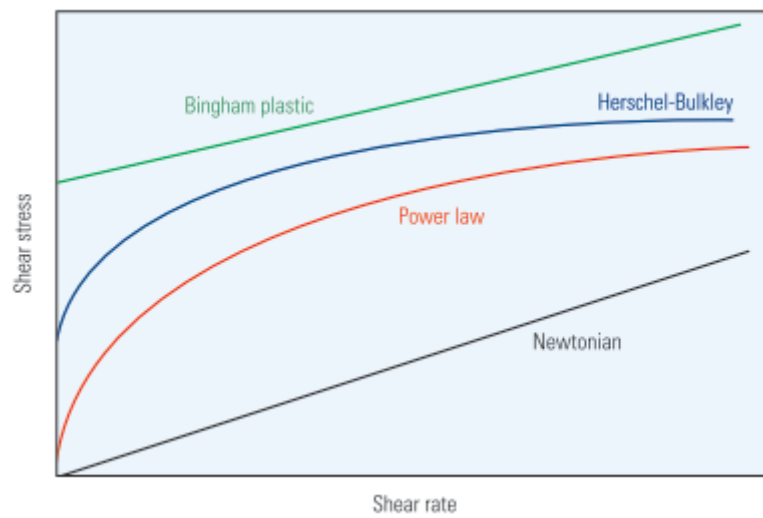


*Figure 8 – Different rheological fluid models (Schlumberger, 2016).*

Newtonian fluids exhibit a linear relationship between shear stress ($\tau$) and shear rate ($\gamma$), with the dynamic viscosity ($\mu$) as constant of proportionality and can be described by Equation (2.5). This linear relationship is valid only for laminar flow (Chowdhury et al., 2009a).

$$\tau = \mu * \gamma$$

*(2.5)*

Bingham plastic model can be described by Equation (2.6) for laminar flow. For a Bingham plastic fluid to flow, the applied shear stress needs to exceed a minimum value which is called the yield point of the fluid ($\tau_\gamma$). When the yield point is exceeded, the change of shear stress ($\tau$) is proportional to the change in shear rate ($\gamma$) and described by the constant of proportionality, the plastic viscosity ($\mu_p$) (Chowdhury et al., 2009a).

$$\tau = \mu_p * \gamma + \tau_y$$

*(2.6)*

The Power law model depicts a non-linear relationship between shear stress and shear rate, as given by Equation (2.7) for laminar flow. In this equation $K$ stands for the consistency index and $m$ denotes the flow behaviour index. Equation (2.7) reduces to the Newtonian fluid model when $m = 1$ (Chowdhury et al., 2009a).

$$\tau = K * \gamma^m$$

*(2.7)*

Herschel-Bulkley (Yield Power Law) fluid model, can be described by Equation (2.8) and valid only for laminar flow. This fluid model also takes the yield point ($\tau_\mu$) of the fluid into account and assumes a non-linear relationship between shear stress and shear rate (Chowdhury et al., 2009a).

$$\tau = \tau_y + K * \gamma^m$$

*(2.8)*

The most basic example of a Newtonian fluid is water. However, well construction fluids need certain rheological properties and functionalities and are usually non-Newtonian. Most drilling fluids are thixotropic non-Newtonian fluids dependent on shear rate, pressure, and temperature, thus preferably modelled with Herschel-Bulkley fluid model (Erge & van Oort, 2020). Thixotropic behaviour can be described as when the flow of a fluid at rest is initiated, the viscosity of the fluid decreases with time, and when the flow is halted, the fluid recovers its viscosity (Erge & van Oort, 2021). This property is vital for the drilling operation because when circulating, the viscosity - which is defined as the fluid's resistance to flow – decreases, and when the circulation is stopped, and the fluid is at rest, viscosity recovers, and gelation of the drilling fluid starts. This development of gel-like structure of the fluid at rest enables the suspension of cuttings, which is very important for keeping the hole clean. However, additional force is required for pump-start, resulting in pressure spikes during this short transitional period from pump start-up to steady state flow. To overcome this, it is recommended to rotate the drillstring first and slowly increase the flowrate (Erge & van Oort, 2021). The rotation of the string induces shear stress, resulting in a decrease of viscosity and, thus, decreased frictional pressure losses.

Once the rheological model is defined, velocity and flow regime must be determined for the drillstring and annulus. The flow regime depends on the geometry, velocity, rheological properties of the fluid (e.g. viscosity), and density, and is calculated based on the Reynolds number for the corresponding rheological model (Bourgoyne, Jr, A T et al., 1991). Equation (2.9) shows the calculation of the mean velocity derived from the conservation of mass. In this equation, $v$ denotes the velocity in m/s, $Q$ represents the flowrate in m³/s, and $A$ is the area in

m$^2$. Equation (2.10) represents the calculation of the Reynolds number for a Newtonian fluid. Where $N_{Re}$ is the dimensionless Reynolds number, $\rho$ is the fluid density in kg/m$^3$, $d$ is the diameter of the flow channel (pipe) in m, and $\mu$ the viscosity in Pa-s. The equation for calculating the Reynolds number is slightly different for other rheological models.

$$v = \frac{Q}{A}$$

*(2.9)*

$$N_{Re} = \frac{\rho * v * d}{\mu}$$

*(2.10)*

Based on the Reynolds number, the prevailing flow regime can be determined, enabling to make a distinction between either laminar or turbulent flow (Mitchell & Miska, 2011). Therefore, depending on the present flow regime, the frictional pressure loss is calculated, either for laminar flow regime or turbulent flow regime. For the Newtonian fluid model, laminar flow occurs when the Reynolds number is less than 2100. However, the transition from laminar to fully developed turbulent flow does not occur abruptly; therefore, for Reynolds numbers ranging from 2000 to around 4000, the flow regime is considered a transitional zone (Bourgoyne, Jr, A T et al., 1991). For Reynolds numbers greater than 4000, the flow regime is turbulent. In non-Newtonian fluids, the critical Reynolds number must be determined to determine whether a turbulent or laminar flow regime is present. This is essential because frictional pressure loss calculations differ not only for the aforementioned flow regimes, but also for the different rheological models discussed earlier. So, when velocity, flow regime, and rheological properties (rheological model) are known, the frictional pressure losses can be calculated for each individual part and geometry within the drillstring and annulus at certain flowrates. More precisely, pressure losses should be calculated for pressure losses inside the drill pipe ($\Delta P_{DP}$), inside the drill collars ($\Delta P_{BHA}$), annular space between drill collars-open hole ($\Delta P_{DC-OH}$), drill pipe open hole ($\Delta P_{DP-OH}$), and drill pipe cased hole ($\Delta P_{DP-CH}$) respectively according to prevailing flow regime.

Pressure loss in surface equipment depends on internal diameter and length of each piece of equipment and is given by Equation (2.11):

$$\Delta P_{Surface} = E * \rho^{0.8} * Q^{1.8} * \mu^{0.2}$$

*(2.11)*

In Equation (2.11) E is a coefficient depending on the length and diameter of the surface equipment, $\rho$ is the density of the fluid in kg/m$^3$, Q is the flowrate in m$^3$/s, and $\mu$ is the viscosity in cP.

Pressure drop at the bit is given by Equation (2.12):

$$\Delta P_{Bit} = \frac{\rho * Q^2}{2 * C^2 * A^2}$$

*(2.12)*

$\Delta P_{Bit}$ is the pressure drop at the bit, $\rho$ is the density of the fluid in kg/m$^3$, Q represents the flowrate in m$^3$/s, C is a dimensionless correction factor or discharge coefficient, and A is the total nozzle area in m$^2$.

Combining the individual frictional pressure drops across all components in the hydraulic conduit yields to the total frictional pressure drop (Chowdhury et al., 2009b). Figure 9 shows a schematic representation of the individual frictional pressure loss components along the hydraulic conduit as described in Equation (2.3).
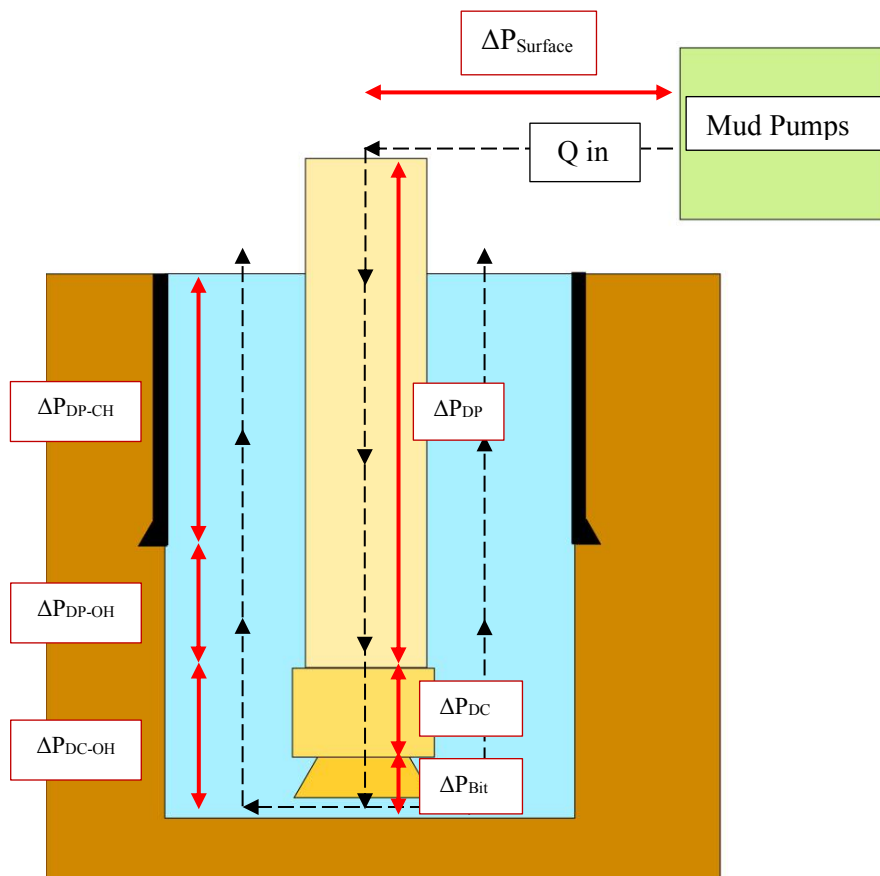


*Figure 9 – Schematic representation of the frictional pressure losses in the hydraulic conduit.*

### 2.4.1.1 Drawbacks of the Conventional Approach in Real Time Application

The conventional approach is vital in the planning phase of the wellbore to identify the operational conditions of the hydraulic parameters enabling the determination of optimized flowrates in terms of hole cleaning, efficient ROP, and safety within the given downhole boundary conditions. However, the conventional approach has several limitations in real-time application. It is very difficult to continuously model SPP with this physics-based approach due to the dynamic nature of the well construction process and uncertainties related to the borehole. Several physics-based hydraulic models allow for real-time simulation of SPP, but they are all based on the same principles. These physics-based models have two major drawbacks (Erge & van Oort, 2020).

The first weakness is related to the general physics-based approach. The physics-based models have difficulties fully representing the physical dynamics in the well construction process, resulting in inaccuracies. This is due to the fact that these models rely on assumptions and simplifications such as a concentric drillstring, a perfectly circular borehole, a non-rotating drillstring, isothermal conditions in the wellbore, and steady-state axial flow (Chowdhury et al., 2009a). While these simplifications are essential to reduce the model complexity, and computational requirements, they also show lower accuracies (Erge & van Oort, 2022). However, in real life, these assumptions are not totally valid (Chowdhury et al., 2009a) and consequently do not represent the actual downhole environment or behaviour of the drillstring and fluid.

Secondly, these models need continuous maintenance and calibration, relying on human interference. This is due to the dynamic nature of the well construction process, where either operational parameters may change or possible change of the equipment in use, which requires manual adaption of the model by an operator, increasing the susceptibility to errors due to human interference (Erge & van Oort, 2022). Also, as stated earlier, accurate modelling of SPP highly relies on the rheological model and rheological properties of the drilling fluid in use as an input. Therefore, the rheological properties must be determined manually on the rig site. However, in reality, measurements of these properties can be infrequent or on a too low frequency, leading to inexact approximations of rheological properties and thus resulting in inaccurate assumptions of the actual frictional pressure drop (Lafond et al., 2021).

The above-mentioned shortcomings of the physics-based approach led to the decision to investigate data-driven approaches to allow for modelling and monitoring of SPP in this complex and dynamic environment. There are several other publications dealing with either purely data-driven or hybrid modelling approaches to predict the hydraulic trend and by utilizing a variety of different algorithms for making predictions (Chowdhury et al., 2009b;

Erge & van Oort, 2020; Varadarajan et al., 2021; Youcefi et al., 2022). However, the approach of this thesis focuses on Gaussian process regression to build the predictive model. This thesis aims to utilize trouble-free drilling data to make predictions of the current SPP trend with the use of controllable surface drilling parameters. This should allow the model to make predictions of the trouble-free case, according to the set operational parameters. And by comparing the modelled value with and actual sensor response, identification of abnormal downhole behaviour should be enabled. The study also aims to determine the minimum number of data points stemming from the same wellbore that are necessary to construct an accurate and efficient model, potentially eliminating the need for historical data. By employing this approach, the model could be applied in real-time, assisting the operator in decision making and enable means to identify potential downhole anomalies by comparing the model´s output to the actual sensor readings.

## 2.4.2  Standpipe Pressure Trend Interpretation

Monitoring and modelling the trend of key surface parameters are essential for detecting deviations from the normal. Unwanted downhole events may cause abnormal behaviours of surface parameters. Identification of these abnormalities happens by modelling the parameter of interest and comparing it with the actual sensor measurement in real time. This provides a methodology for quantifying the deviation of the expected and the actual sensor response, hence enabling early detection of anomalous behaviour. In case of hydraulic parameter monitoring, standpipe pressure is the parameter of interest. Anomalous behaviour of the surface pressure is usually the first indicator for numerous downhole-related problems (Lafond et al., 2021), and therefore, understanding the hydraulic trend is crucial for a safe operation, early event detection, and proper problem diagnosis. As stated earlier, a change in pump pressure with all operational parameters kept within the operational range (weight on bit, ROP, flowrate, RPM) may either be caused by an induced change in the parasitic system pressure losses or can be related to pressure loss at the bit. As a consequence, abnormal surface pressures can be directly related to the most common downhole problems when correctly interpreted. In several cases, standpipe pressure can be used as a main or secondary indicator for several downhole problems (Elmgerbi & Thonhauser, 2022). Standpipe pressure, combined with the readings of other key surface parameters such as surface torque, hook load, and flow-out, provides a systematic approach for downhole problem classification, as shown in Table 2, adapted from Elmgerbi & Thonhauser (2022).

*Table 2 – Downhole problems with main and secondary indicators. Table adapted from Elmgerbi & Thonhauser (2022).*

| Incident | Main Indicator | Secondary Indicator(s) |
|---|---|---|
| Drill String Washout | Standpipe Pressure | Flow-out |
| Hole Cleaning | Standpipe Pressure | Surface Torque, Hook Load |
| Twist Off | Standpipe Pressure | Hook Load |
| Bit Balling | Standpipe Pressure | Rate of Penetration |
| Losses | Flow-out | Standpipe Pressure |
| Kick | Flow-out | Standpipe Pressure, Rate of Penetration |
| Stuck Pipe | Surface Torque | Standpipe Pressure, Hook Load |
| Tight Hole | Surface Torque | Standpipe Pressure, Hook Load |
| Hole Collapse | Surface Torque | Standpipe Pressure, Flow-out |

Depending on the specific downhole problem, SPP trend can either show an increasing or decreasing trend. Within this increasing or decreasing behaviour a further distinction can be made and classified in sudden increase, gradual increase, sudden decrease, and gradual decrease of the standpipe pressure trend.

Assuming operational drilling parameters are kept constant, a sudden increase in SPP can be observed due to plugging of one or more bit nozzles, or due to bit balling (Elmgerbi & Thonhauser, 2022).

Gradually increasing SPP can be caused either by an increased flowrate, a change in drilling fluid rheological properties and/or elongation of the drill string during the drilling process. During drilling, cuttings may accumulate in the annulus due to poor hole cleaning practices, causing the rheological properties of the drilling fluid to change which shows in increased surface pressures. If this gradual increase in SPP is followed by an increase in surface torque, a stuck pipe event is likely to be imminent (Elmgerbi & Thonhauser, 2022). High ROP can also lead to increased frictional pressure losses due to increasing concentrations of cuttings in the drilling fluid. In this regard, the increased volume of generated cuttings during drilling may eventually change the drilling fluid´s rheological properties, causing this effect. Also, with the progress of the drilling operation, the drill string elongates with every stand drilled, which also results in increased frictional pressure losses. Therefore, in normal drilling operations, the flowrate must be periodically reduced as depth or length of the wellbore is increasing (Elmgerbi et al., 2022).

A sudden decrease in SPP can be due to the decrease of bit pressure owing to the loss of a bit nozzle or a drillstring twist-off.

Gradually decreasing SPP indicates an impending kick, loss of drilling fluid, dysfunction of the mud pumps, or an ongoing drillstring washout. An influx of formation fluids causes a gradual decrease in SPP. The lighter density fluid reduces the hydrostatic head in the annulus, leading to lower density and consequently to decreased frictional pressure loss. Simultaneously the return flowrate increases, showing increased flow-out measurements at the surface. Similar behaviour is noticeable for an ongoing loss event. However, the difference between fluid loss and a kick is that when losses occur, the flow-out decreases, with the corresponding reduction in surface pressure (Elmgerbi & Thonhauser, 2022). Drillstring washout can either be indicated through a sudden drop of SPP (several hundreds of psi) (Ambrus et al., 2018), but the washout can also develop slowly over hours, exhibiting a gradual decreasing trend of SPP and is hence even harder to detect (Kuesters et al., 2020). A case presented by Zhao et al. (2019) shows the behaviour of SPP and flowrate during an ongoing washout event as can be seen in Figure 10 (Zhao et al., 2019).
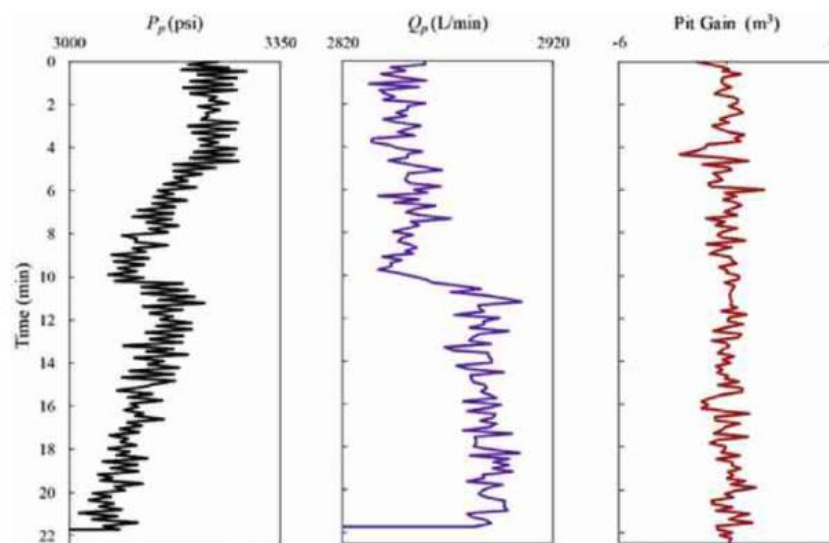


*Figure 10 – Real-time sensor measurements of SPP (black), flowrate (purple), and pit gain (red) versus time during drill string washout (Zhao et al., 2019).*

In this case, a washout caused a reduction of SPP at the minute of 4, with the flowrate and all other operational parameters kept constant. The drilling crew increased the pump rate at minute 12, which should theoretically result in higher frictional pressure losses (SPP). However, the SPP trend gradually decreased again, with no significant change in pit gain, and finally, the washout was confirmed at minute 22 followed by a stop of the operation. A pump failure or partial degradation usually shows similar trend behaviours as a drillstring washout and is

indicated by a decrease in standpipe pressure, but with a deviation between expected flowrate into the well and the actual. Flow out monitoring is essential to distinguish a pump failure from a drillstring washout. Pump failure results in decreased flow out, while drillstring washouts do not (Ambrus et al., 2018).

Monitoring and trend analysis of SPP are vital tools for linking drilling symptoms to potential downhole issues. Several studies investigate on utilizing machine learning approaches to identify deviations in the hydraulic trend, thus enabling a methodology for early downhole problem detection. Ambrus et al. (2018) propose a machine learning approach for early drill string washout and pump failure detection. As well as a study conducted by Zhao et al. (2019) shows the successful application of machine learning assisted washout event detection in real-time. Youcefi et al. (2022) propose a new model for SPP prediction using Group Method of Data Handling (GMDH) method. The proposed approach utilizes drilling parameters and mud properties as an input for the model to predict SPP in real time. The model was trained with data from two wells and validated on a third well. Results of this study show high accuracy in SPP prediction. The developed model represents the normal trend of SPP, and the difference between measured and predicted SPP provide means for early detection of several downhole issues such as drillstring washout, pump failure and bit balling. Elmgerbi and Thonhauser (2022) introduced a sophisticated algorithm capable of autonomously analysing real time drilling data and thus enabling detection of most common downhole drilling problems. The approach facilitates the construction of a risk predictive window through a stochastic process and combining it with a data-driven model. The approach doesn´t solely focus on detection of a specific downhole issue, moreover it provides means to detect the most common downhole problems utilizing various real time drilling parameters. To detect abnormal downhole behaviour real-time sensor measurements are compared with the risk predictive window of the corresponding measurement, and consecutive datapoints outside the predictive window, exceeding a certain threshold, are triggering an alarm. Results of this study showcase the successful application of this method for detecting most common down drilling problems.

The literature review presented in the previous chapters, covering various subtopics related to the research task, has been necessary to establish a fundamental understanding of the problem and the accompanying intricacies, laying the groundwork for a nuanced understanding of the critical factors influencing the parameter of interest. As a next step, the following chapters present the applied methodology for building the predictive model.

# Chapter 3

# Developed Methodology Overview

## 3.1 Chapter Overview

The subsequent sections provide information about the developed methodology comprising of concept overview, general description of the available data, and required data preparation and preprocessing steps for building the predictive SPP model based on GPR.

## 3.2 Methodology Background

The main objective of the thesis is to develop a machine learning model to predict standpipe pressure (SPP) of one stand or the next based on the data recorded in the same stand. The idea is to use Gaussian process regression to predict the SPP (regression target value) with controllable drilling surface parameters such as WOB, flowrate, and RPM (as features), which can be adjusted by the driller at the rig. The application of GPR also provides a confidence interval for the predictions. This should enable a methodology for identifying eventual downhole abnormalities when comparing the actual value derived from the sensor measurement versus the predicted value calculated by the model. Therefore, for the model setup, it is vital that only clean and trouble-free drilling data is used for training the machine learning model, and the way of preparing the data for the model and the applied procedure to derive trouble-free drilling data is described in the following paragraphs. Figure 11 schematically represents the required internal data preparation process. Beginning with the extraction and identification of actual real time drilling data, and ultimately yielding trouble free drilling data for each individual drilled stand.
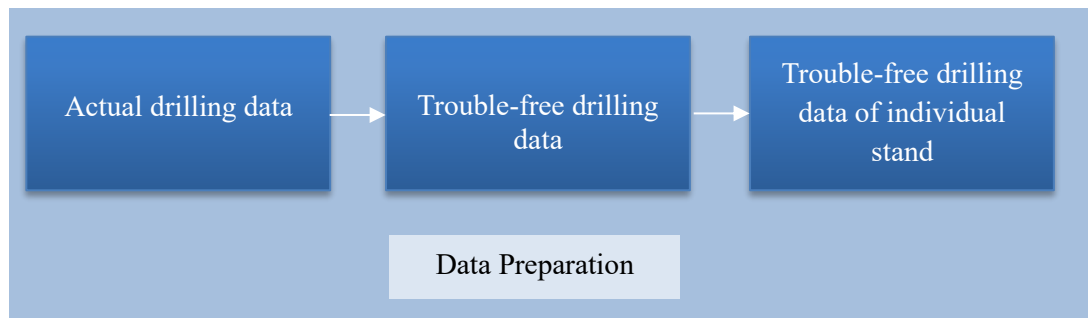
*Figure 11 – Required data preparation.*

The model setup starts once the data set is considered to be trouble-free and clean. The predictive model is trained with 1-2 hours of trouble-free recorded data of one stand, and predictions of SPP are made for the following 1-2 hours of drilling for the same stand, with controllable surface drilling parameters as inputs. Figure 12 depicts the concept behind the predictive model, where it is trained on 1-2 hours of real-time drilling data from an individual drilled stand and then used to forecast standpipe pressure values for the next 1-2 hours for the same stand, based on currently applied operational parameters.
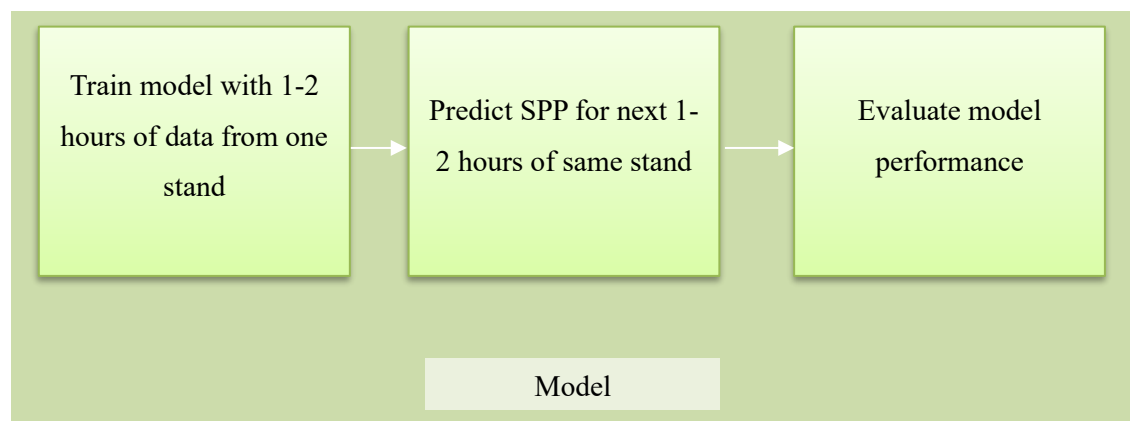


*Figure 12 – Methodology for the predictive model.*

The idea behind the proposed method is that fitting the model with trouble-free drilling data should enable the model to represent the normal behaviour of the target variable according to set operational parameters used during drilling (Flowrate, RPM and WOB) and the actual borehole condition. Consequently, the model should be representative for a "healthy" SPP trend, thus enabling potential abnormality identification by comparing the modelled value with the actual sensor measurement. Furthermore, providing information about resulting standpipe pressure readings due to change in operational parameters.

Figure 13 presents a flowchart that shows the sequential steps of the developed and applied methodology.
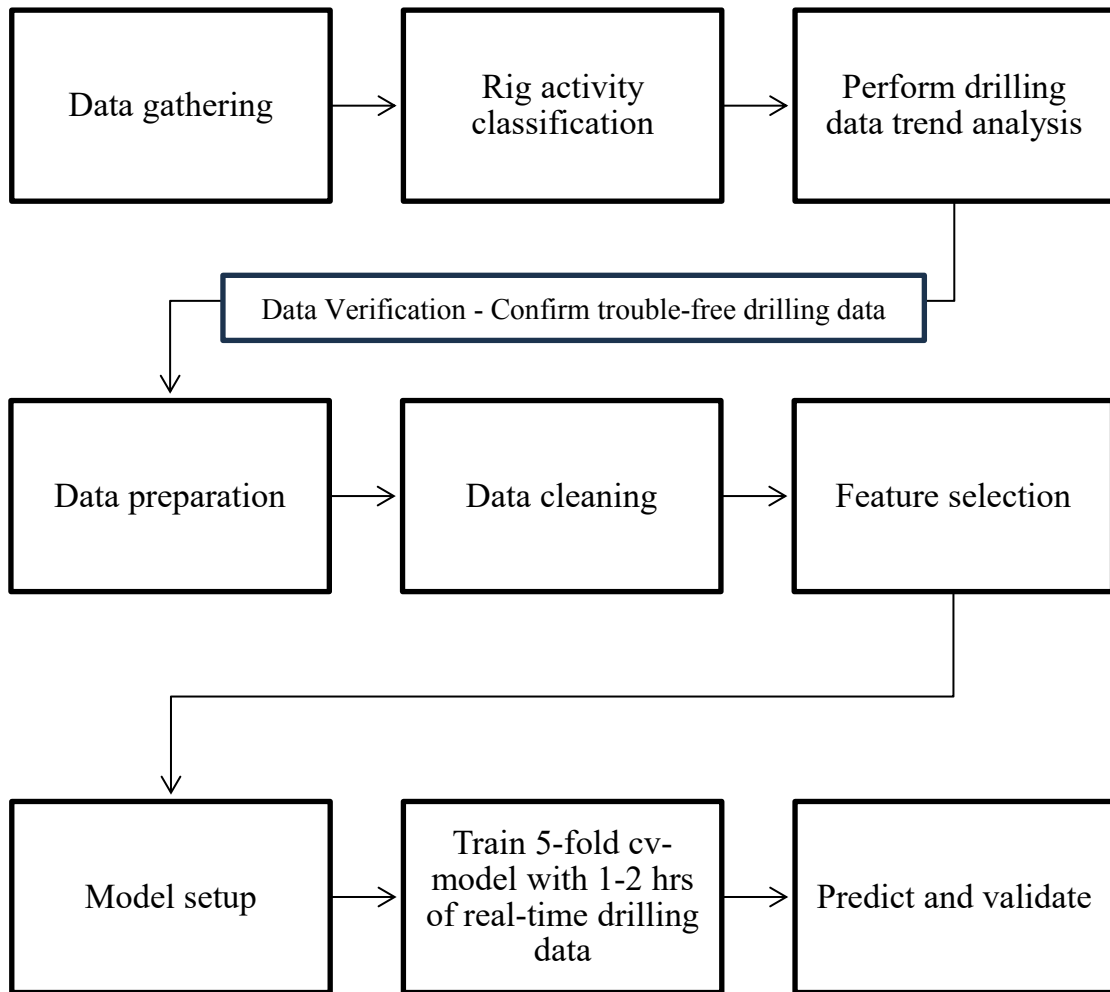
```
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│                 │     │                 │     │                 │
│  Data gathering │ ──> │  Rig activity   │ ──> │ Perform drilling│
│                 │     │  classification │     │ data trend      │
│                 │     │                 │     │ analysis        │
└─────────────────┘     └─────────────────┘     └─────────────────┘
```

Data Verification - Confirm trouble-free drilling data

```
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│                 │     │                 │     │                 │
│ Data preparation│ ──> │  Data cleaning  │ ──> │Feature selection│
│                 │     │                 │     │                 │
└─────────────────┘     └─────────────────┘     └─────────────────┘
```

```
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│                 │     │ Train 5-fold cv-│     │                 │
│   Model setup   │ ──> │ model with 1-2  │ ──> │Predict and      │
│                 │     │ hrs of real-time│     │ validate        │
│                 │     │ drilling data   │     │                 │
└─────────────────┘     └─────────────────┘     └─────────────────┘
```

*Figure 13 – Flowchart representing the workflow and methodology.*

## 3.3 Developed Model – Core Operational Principles

The applied methodology and the internal data process are discussed, which are necessary for building the predictive model, describing it in a step-by-step manner. As mentioned above, the aim is to train the predictive model with trouble-free drilling data, and the methodology to prepare and analyse a raw drilling dataset is presented below. All required steps to build the predictive model, starting from raw drilling data import, data preparation, data cleaning, and to setup the predictive model are performed utilizing MATLAB environment. MATLAB is selected because it provides many useful built-in functionalities, the ability to handle large amount of data, computational efficiency, as well as a variety of machine learning algorithms for model selection. From this variety of algorithms, the MATLAB built-in Gaussian process regression is used as a model to predict standpipe pressure based on controllable surface parameters.

### 3.3.1 Data Gathering

The raw drilling dataset contains sensor measurements from various data channels for monitoring the drilling operation in real time. The measured parameters in the provided dataset comprise of several common surface-measured drilling parameters. An overview of the available data channels is shown in Table 3.

*Table 3 – Overview of different drilling data channels.*

| Measurement | Unit | Measurement | Unit |
|---|---|---|---|
| Time | s | RPM | rpm |
| Total Depth | ft | TORQUE | lb*ft |
| Bit Depth | ft | SPP | psi |
| ROP | mn/ft | FlowPmps | gpm |
| ROPav | ft/hr | MW IN | ppg |
| HookHght | ft | T Rev | kRev |
| WOH | klbs | Pump Time | h |
| WOB | klbs | T Gas Main | ppm |

### 3.3.2 Rig Activity Classification

Correct data preparation is one of the most important steps because the performance of the ML model relies highly on the quality of the data. As part of the objective, clean trouble-free drilling data is required to set up and train the predictive model. Therefore, as a first step, the identification of the actual operation or rig activity is necessary. This allows not only to extract and allocate all datapoints related to drilling activity, but also provides means to perform drilling data analysis across various rig activities. During a well construction process, several

routine drilling activities are carried out, which can be characterized based on the combined readings of different sensor measurements at a certain timestamp. These routine rig activities comprise of actual drilling (making hole), making connections, reaming up/down, tripping in/out, circulating, etc. Figure 14 shows an outcrop example of real time drilling data streams of six key surface parameters including Hook Hght (Block position), Weight on Hook (WOH), SPP, FlowPmps (Flowrate), RPM, and torque from the available dataset. Patterns in these data channels can indicate certain routine rig activities. To differentiate between drilling (making hole) and non-making hole rig activities, a MATLAB script has been developed that assigns labels to each timestamp based on the sensor reading at that time. The assigned labels represent the current rig activity, providing means for accessing and extracting important desired datapoints.

The developed script to classify the datapoints and assign these to the current rig activity is based on simple logic, case specific threshold limits for operational parameters, and conditional statements. Therefore, characteristic behaviours of the sensor measurements are utilized to identify the rig activity and rig state. To make proper classification of the activity possible, new columns needed to be created and added to the existing table. Amongst these new columns are for example a column representing the Block Movement, Hole Depth Change, and a column representing the Bit off Bottom Distance. Block Movement column displays the vertical displacement of the block/hook, and is calculated by subtracting two consecutive HookHght measurements, and the resulting sign convention indicates either upward or downward movement of the block/hook at the current timestamp. The same principle is applied for Hole Depth Change, where two consecutive TotalDepth readings are subtracted and resulting in relative hole depth change. In addition to the snapshot information provided by these two columns, another column is required considering a sequence of measurements over 50 seconds (which is equal to 5 measurements based on 0.1 Hz frequency) and labelled as "ZeroHoleDepthChange". The "ZeroHoleDepthChange" column employs boolean values to indicate whether the hole depth remained unchanged over the past 50 seconds. With this newly extracted information and other related key surface measurements (WOB, WOH, RPM, flowrate, SPP) different rig activities can be classified accordingly based on multiple conditions which are characteristic for the actual rig state.
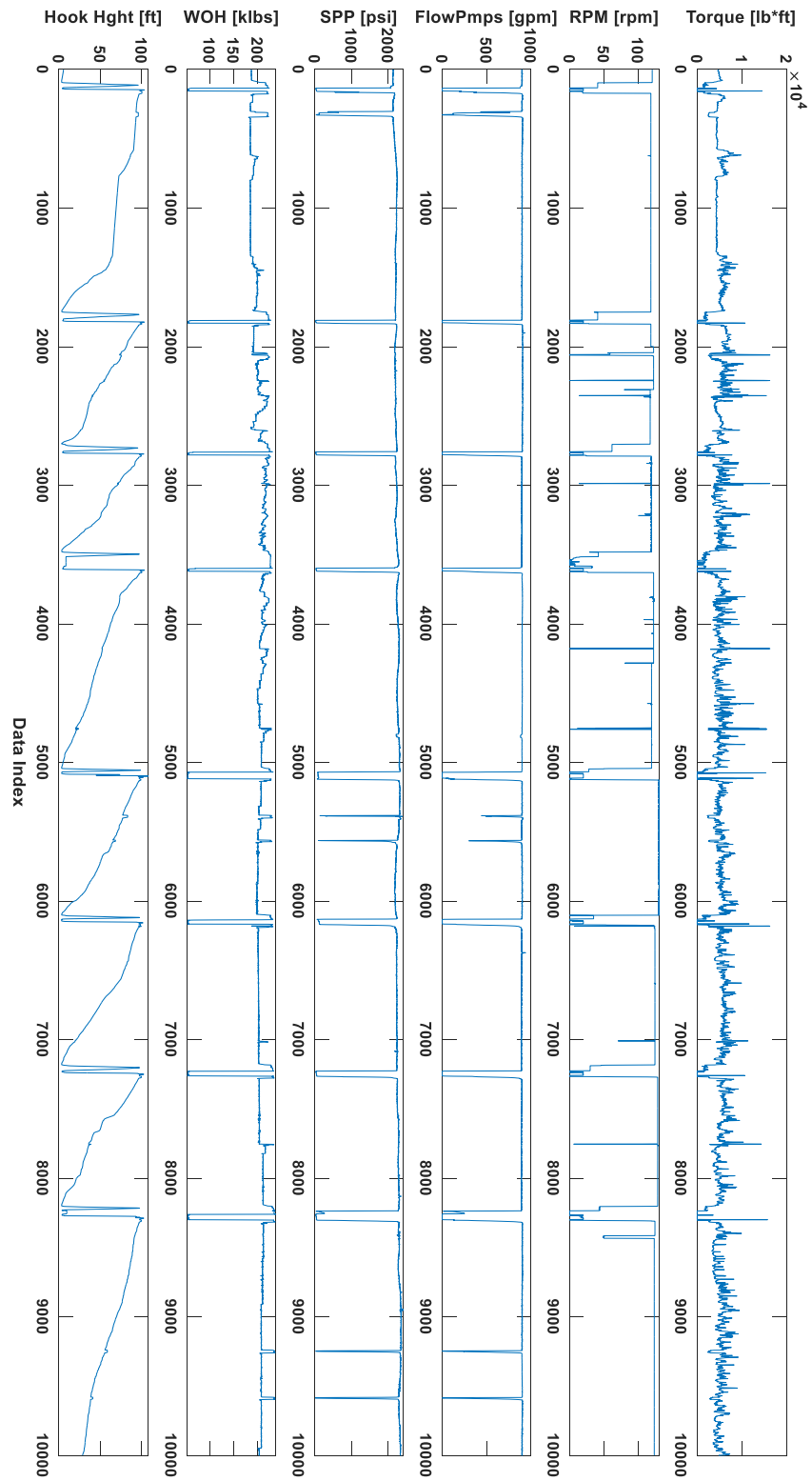
*Figure 14 – Outcrop of real time drilling data streams from the provided dataset.*

Making a connection is a good example to demonstrate and visualize different rig activities which are consecutively carried out during the well construction process. The connection process itself is a crucial step in the well construction process, as it is essential for extending the length of the drillstring by adding a so-called drilling stand, and thus enabling to progress deeper into the wellbore. The corresponding sensor readings show particular behaviours before, during, and after making a connection, enabling the rig activity recognition at the corresponding timestamp. Figure 15 shows the HookHght (Block position) in blue and the corresponding WOH (Weight on hook) in orange for a selected interval where a connection is made. With the use of other operational parameters (e.g. flowrate, RPM, movement of the block, etc.), a clear breakdown of rig activities is possible, allowing the assignment of the current rig activity label to each datapoint.
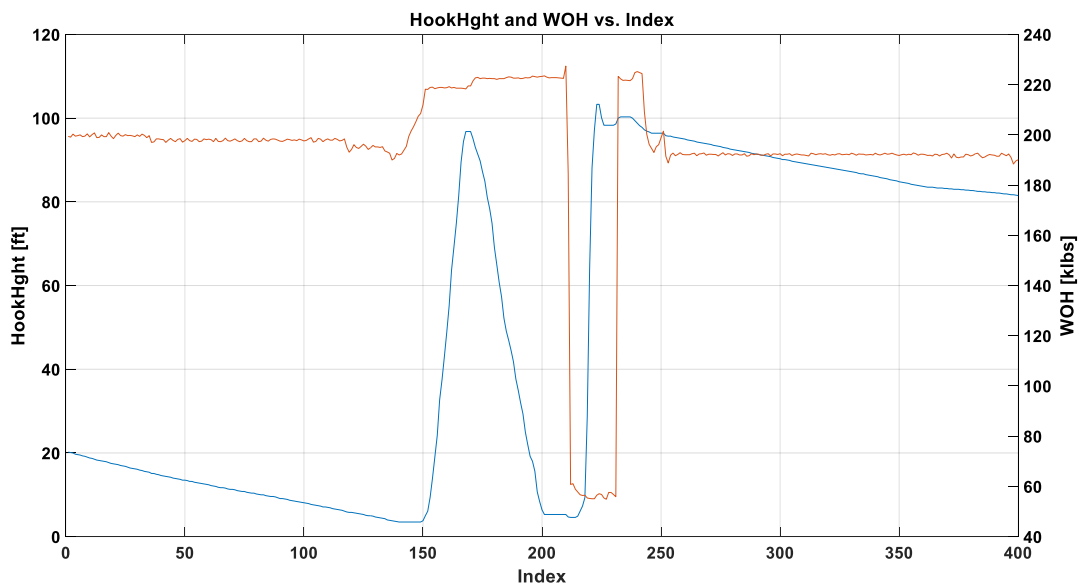


*Figure 15 – Real time data trends of HookHght (blue) and WOH (orange) vs Index.*

Figure 16 shows the classified datapoints for the same interval as shown in Figure 15, where the HookHght vs Index is plotted. It can be seen that several different routine rig activities are classified according to the current rig activity, with each rig activity label highlighted in different colours. According to this principle, the coded rig activity classifier can assign a label (the current operation/rig activity) to the corresponding datapoints. This is done for the whole dataset. This procedure is necessary to conduct further data analysis on various rig activities and build the predictive model with data related to drilling activity. In addition to assigning each datapoint in time to a specific routine drilling operation, datapoints could be assigned to individual drilling stands, resulting in 11 individual stands. This encoding allowed the access of each individual drilled stand and rig activity (Figure 17) according to the label of the operation and the number of the current stand. For simplicity, the labels of the stand numbering

start with 1, representing the first drilled stand in the available dataset, and is ending with the last drilled stand, stand number 11. This numbering should not be mistaken with the total number of stands equipped in the drill string.
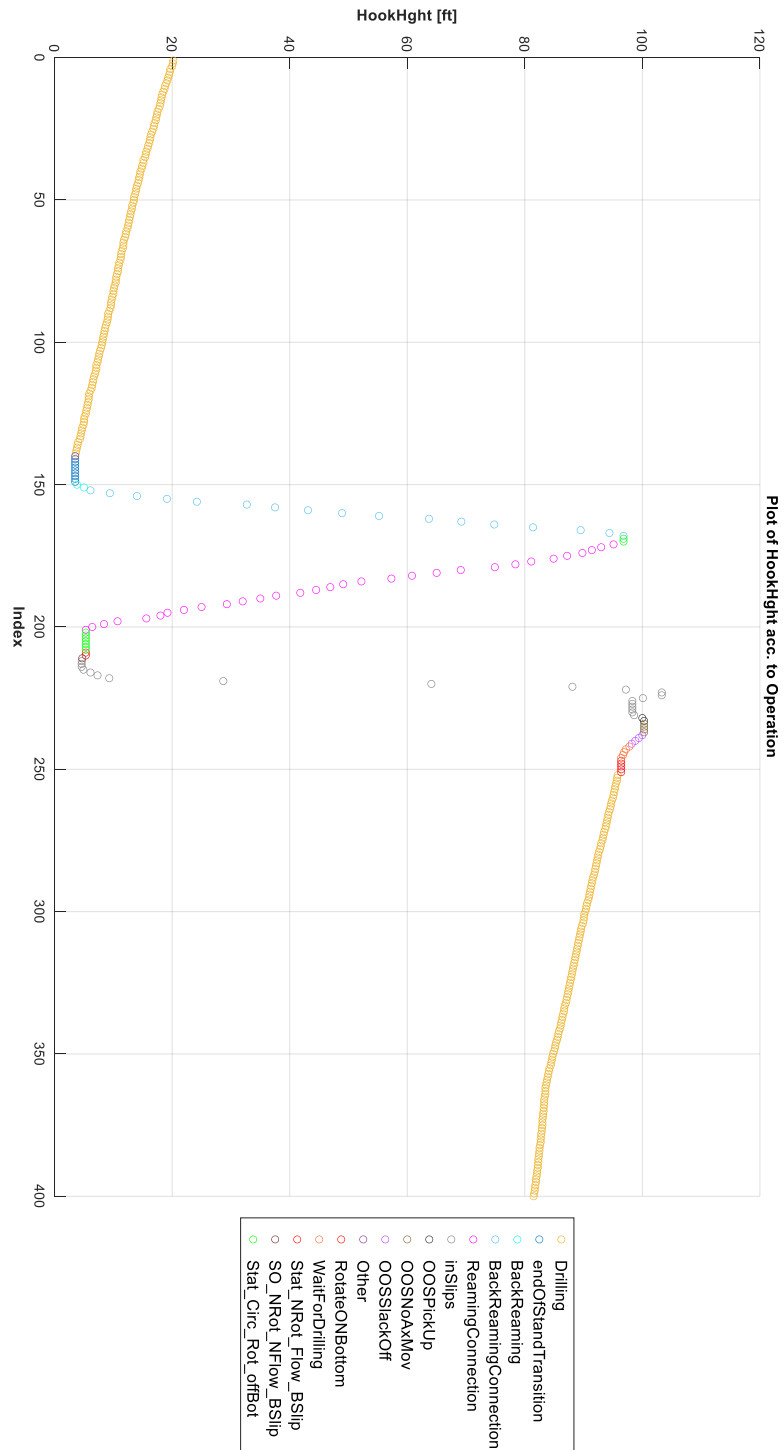


*Figure 16 – Plot of HookHght vs Index. Datapoints are classified according to current rig activity.*

| Time | TotalDepth | BitDepth | ROP | HookHght | WOH | WOB | RPM | Torque | SPP | FlowPmps | Operation | Current Stand Number |
|------|-----------|----------|-----|----------|-----|-----|-----|--------|-----|----------|-----------|---------------------|
| 00:00:10 | 4731.7 | 4731.7 | 7.1 | 5.5 | 187.3 | 36.3 | 121 | 4728 | 2153 | 902 | "Drilling" | 1 |
| 00:00:20 | 4731.7 | 4731.7 | 7.1 | 5.5 | 187.1 | 36.5 | 121 | 4734 | 2151 | 899 | "Drilling" | 1 |
| 00:00:30 | 4731.7 | 4731.7 | 7.1 | 5.5 | 187.2 | 36.5 | 121 | 4731 | 2140 | 905 | "Drilling" | 1 |
| 00:00:40 | 4731.7 | 4731.7 | 7.1 | 5.5 | 187 | 36.6 | 121 | 4808 | 2149 | 901 | "Drilling" | 1 |
| 00:00:50 | 4731.7 | 4731.7 | 7.1 | 5.4 | 187.4 | 36.3 | 121 | 4691 | 2141 | 900 | "Drilling" | 1 |
| 00:01:00 | 4731.8 | 4731.8 | 7.1 | 5.4 | 186.9 | 36.8 | 121 | 4674 | 2142 | 901 | "Drilling" | 1 |
| 00:01:10 | 4731.8 | 4731.8 | 7.1 | 5.4 | 187.2 | 36.5 | 121 | 4656 | 2147 | 902 | "Drilling" | 1 |
| 00:01:20 | 4731.8 | 4731.8 | 7.1 | 5.4 | 187.2 | 36.4 | 121 | 4632 | 2148 | 905 | "Drilling" | 1 |

*Figure 17 – Outcrop of the applied labelling scheme. This enables easy filtering of the data, to further prepare and extract subsets of data for analysis, preprocessing, and training the predictive model. The two headers "Operation" and "Current Stand Number" refer to the classified rig activity and the number of the drilled stand, respectively.*

### 3.3.3 Drilling Data Trend Analysis

Analysis of the drilling data needs to be performed to confirm trouble-free drilling data which consequently can be used for the predictive model. Therefore, trend analysis is performed to identify eventual deviations from a certain trend or changes in surface drilling sensor measurements for comparable routine rig activities. Due to the absence of modelled or calculated ideal values for comparison, it was opted to analyse different trends and trend development of several key surface drilling parameters across the available dataset, and across drilled individual stands to identify abnormal downhole behaviour or changes in the downhole condition. This approach seemed to be suitable given the lack of ideal values.

This approach compares key surface drilling parameters such as SPP, WOH, and torque during comparable routine rig activities or operations. Therefore, it is essential to select rig activities that are carried out with similar operational parameters (flowrate, RPM, time, same mud, etc.) to make a comprehensible comparison possible. However, there is a certain difficulty in locating operations that are conducted with similar operational parameters and continuity when working with real world data. Routine drilling rig operations may be conducted differently and depend on the company's requirements or guidelines or how the different rig crews perform them. Identifying comparable activities may also become an issue when working with historical data because, what is common practice now, may not was common practice several years ago.

Therefore, trends are analysed for two operations/cases. First, the hydraulic trend of SPP is examined with other related measurements such as flowrate, RPM, mud weight (MW IN) during actual drilling activity. The sensor trends during actual drilling are selected and analysed to get an overview of the development of the SPP trend during actual drilling over time/depth/drilled stands. Analysing the trend of SPP across the drilled depth interval can provide valuable insights of the current downhole situation. Secondly, the WOH (Hook load) trend during backream (POOH) and ream down (RIH) at the end of each drilled stand before

the next connection. This operational sequence was selected because it was carried with similar operational parameters (flowrate, RPM, time, etc.) throughout the available dataset and thus also provides insights about the current downhole conditions or trend development. Although not mentioned in detail in this thesis, monitoring the hook load trend and corresponding torque during various rig activities is common practice to monitor the current downhole condition.

### 3.3.4  Data Preparation

Some more data preparation steps are necessary following the verification of trouble-free drilling data. This step comprises of the extraction of data related to actual drilling activity and partition the dataset in individual stands ending up with different subsets of data equal the amount of individual drilled stands. This provides following advantages: further investigate drilling data related to individual drilled stands, easier data cleaning, and the individual datasets at this point only contain data related to the actual drilling activity.

### 3.3.5  Data Cleaning

After data analysis and the acquisition of trouble-free drilling data, remaining artifacts, misclassified drilling datapoints, or numerical outliers had to be assessed, analysed, and removed if necessary. Therefore, the drilling data of each stand is examined individually, allowing a more accurate approach for outlier identification and removal. Selection of the proper outlier identification and removal technique is crucial. Therefore, different drilling parameters need to be assessed individually. Applied outlier detection technique is based on the behaviour and nature of the corresponding parameter. Hence, for parameters such as RPM, which normally do not vary much during drilling process, the z-score method is applied. For parameters which usually show more dynamic behaviour during drilling, such as weight on bit or standpipe pressure, moving mean and moving median filters are selected.

### 3.3.6  Feature Selection

Feature selection is a very important part of building a machine learning model. Feature selection in this specific case is based on multiple factors. First, to align with the research objective of the thesis and predict SPP based on controllable key surface parameters. Hence, domain knowledge played a crucial role in the feature selection process, which is proven to have an influence on the target variable. However, features were also tested and validated using F-Test feature ranking. This involved statistical analysis and other relevant techniques for feature identification to confirm that selected features are significant for the modelling task.

As a result, flowrate, RPM, and WOB are the features selected for building the predictive model.

Another important point in the selection of only 3 features to model the corresponding target variable is due the nature of the GPR algorithm. Increasing the number of features would introduce computational complexity making the model more resource intensive and slower due to the increased dimensions of the input space, which is also known as the curse of dimensionality. Also, there is a strong belief based on expert knowledge, that adding of more features may result in deterioration. Therefore, it was decided to focus on the aforementioned features and thus keep the model simple.

### 3.3.7 Predictive Model Setup

The preparation steps and data analysis are necessary to acquire a clean and trouble-free drilling dataset to build and train the predictive model. Training the model with clean trouble-free drilling data should enable the model to learn the underlying function and relation of the features and the target. The predictive models are trained and evaluated on several individual stands, and the best-performing models are discussed in more detail in the following next chapter.

The general workflow for building the model is to train the predictive model with 1-2 hours of actual drilling data of a stand and make predictions of SPP for the following 1-2 hours of the same stand, to provide an estimate of SPP reading based on currently applied operational parameters from the operator (the input features). This approach results in a training set size of 360-720 observations based on a 0.1 Hz frequency. Different training set sizes were investigated based on data availability to ensure sufficient datapoints of the same drilling stand are reserved for model testing and evaluation. The number of available datapoints depends on the real drilling ROP. Therefore, when drilling commenced with a higher average ROP, smaller datasets were generally used to train the predictive model (360 observations which is equal to 1 hour of actual drilling), and on the other hand, while drilling with a lower average ROP, more observations were available for training and still ensured enough datapoints were reserved for testing the models on unseen data of the same stand. Features are standardized before model training to ensure consistent scales. When using selected input features to predict the target variable SPP, models are, in theory, valid for the current stand to be drilled. However, if performance was extraordinary, approaches are taken to make prediction for the next stand to be drilled.

A useful MATLAB built-in function when building a predictive model based on GPR algorithm is that it allows for simultaneous training of the model and hyperparameter optimization, respectively. The model is trained while the hyperparameters of the model are optimized accordingly. After training and optimization of hyperparameters the model's performance is evaluated using 5-fold cross validation. The final model is then tested on the reserved data.

The predictive Gaussian process regression model in this approach has two outputs, the predicted mean, and a 95% confidence interval. The predicted mean value represents the most likely value at the input location for a given set of features. The variance in the predictions is expressed by the predicted 95% confidence interval, comprising of a lower and upper limit. The confidence interval of a GPR addresses the uncertainty of the predictions. That said, for a given set of features, the predicted confidence interval indicates that the true value of the target lies between the lower and upper limit with a 95% confidence, providing a range of probable values.

Each model is then evaluated using common performance metrics for machine learning models applied to regression problems. The performance metrics used comprise of Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-Squared ($R^2$ – also known as the coefficient of determination). These performance metrics derived from the implemented machine learning model, not only allow for interpretation of the predictions but also provide a reasonable evaluation of effectiveness of the general modelling approach. A visual representation of the predictions is also provided to further assist in interpretation of the results. Figure 18 provides an example of the model outputs, showing the result of one of the investigated case studies (Case #1), which will be discussed later in the thesis. As seen in this figure, the blue trendline represents the standpipe pressure measurements in the training dataset, then the regression starts, and predictions are made with the input features. The predicted mean is indicated by the green trendline, and the green shaded region indicates the confidence interval of the predictions.
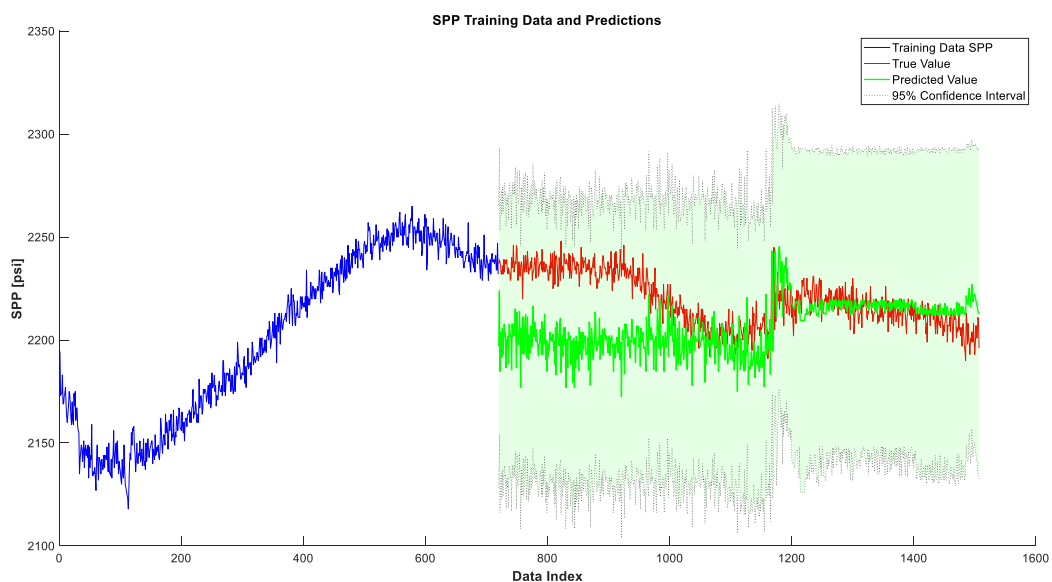


*Figure 18 – SPP training and test results from Case 1. The blue trendline represents the training data, red trend represents the actual sensor measurement (true value), green trendline indicates the model predictions, and green shaded region is representing the confidence interval of the predictions.*

# Chapter 4

# Case Study

## 4.1 Chapter Overview

This chapter provides background information of the dataset used in this study, where general well information as well as all necessary information extracted from the corresponding Daily Drilling Reports (DDRs) is shown. Also, the above described way of analysing the drilling data is shown. Finally, the predictive modelling results are presented and evaluated on three distinct cases, where the findings and limitations of the taken approach are discussed and presented.

## 4.2 Case Study Background

The historical data provided for analysis and for building the predictive model comes from a real-time recorded onshore drilling operation. The full raw dataset consists of two consecutive days (48-hour period) of a standard well construction process with two corresponding Daily Drilling Reports (DDRs). During these two days, a 17 ½" hole section with a starting depth of 4731.7 ft MD and an ending depth of 5644.4 ft MD was drilled. A summary of general well information is provided in Table 4. The drilling crew encountered a problem during drilling, resulting in the decision to Pulling Out Of Hole (POOH) the complete drillstring, including the BHA and to inspect the used BHA. A summary of the event and the actions taken according to the DDR is described below.

*Table 4 – General well information.*

| General Information | |
|---|---|
| Location | Onshore |
| Hole Size | 17.5" |
| Depth Interval | 4731.7 ft – 5644.4 ft |

| Inclination of Interval | vertical |
| LWD / MWD / special BHA components | Pendulum Assembly |
| LWD / MWD / special BHA components | none |
| Data Resolution | 0.1 hertz frequency (every 10 seconds) |

## 4.2.1  DDR Summary

On the first day, drilling started at a depth of 4731.7 ft to 5403.1 ft MD, where a total footage of 672.4 ft was drilled with a reported average ROP of 31ft/hr within a 24-hour period. A formation change at 4775 ft was confirmed. Return flow (flow-out) showed an increased density of 9.2 ppg. Running centrifuges and dilution were carried out to keep the mud weight below 9.2 ppg. Except for the increased density of the return flow, no special remarks are noted in the DDR.

The next day, the crew drilled a total footage of 241.3 ft to 5644.4 ft MD with an average ROP of 14.5 ft/hr. Again, centrifuges ran and continuously diluted the returning mud to keep mud weight below 9.2 ppg. At 18:00, the rig crew started pumping 100 bbl of a High-Viscous (HIVIS)/LCM pill, reciprocating and rotating with reduced flowrate and RPM to clean the hole. When the pill was at the surface, the crew observed overflowing shakers, indicating the unloading of the wellbore. With the pill at the surface and the well under static conditions, the crew started to wash and ream back and trip out the drillstring. While POOH from 5644 ft to 5415 ft, overpulls of 30-40 klbs were reported.

Figure 19 represents the timeline for the 48-hour period of the drilling operation and progress according to the real-time data and the DDR. As described in the summary above, the time and current depth are highlighted where the drilling progress stopped and pumping of the HIVIS pill started. The dataset and DDRs provided for this two-day period does not extend beyond this timeframe. Therefore, the exact nature of the encountered problem and the exact time it began is unknown and needs to be verified with comprehensive data analysis and the help of expert knowledge. However, the DDR summary leads to the assumption that the crew is facing hole cleaning issues during drilling.

# Drilling Operation Timeline
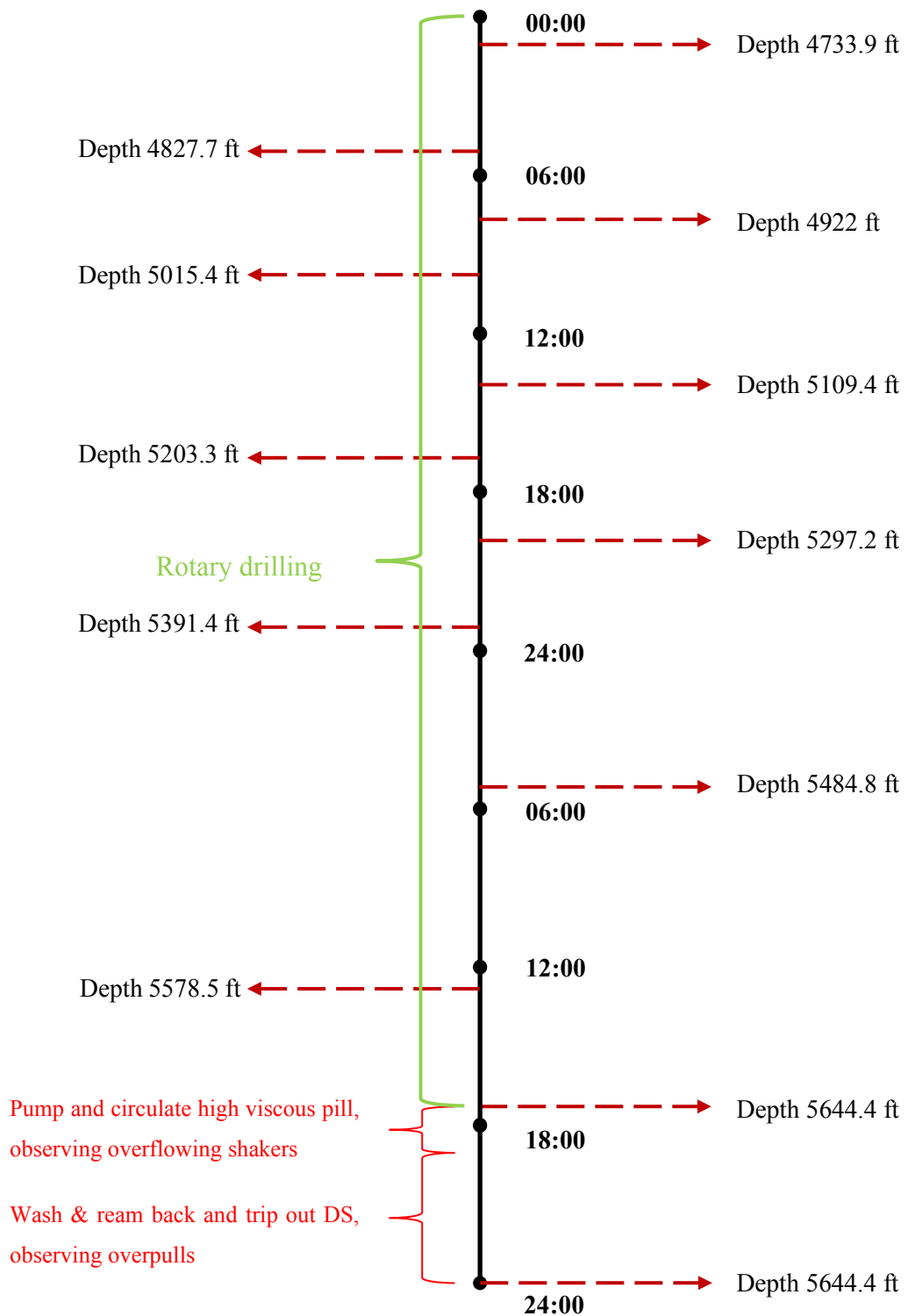


*Figure 19 – Timeline of drilling progress and operation history. Red-dotted arrows indicate current hole depth and are representing the corresponding end depth for each individual drilled stand. Around 18:00 the drilling process stopped and pumping of HIVIS pill started, followed by wash and ream back, and trip out drill string.*

## 4.2.2 Data Analysis and Verification

In response to the detected problem, remediate actions were implemented beginning at 18:00 on the second day. Most likely, the problem causing the rig crew´s decision to trip out is stemming from a borehole cleaning issue. To guarantee the integrity of the data employed for model training, data analysis is performed. Therefore, trends of different key surface drilling sensor measurements such as SPP, weight on hook (WOH), torque, etc. during various routine rig activities (drilling, pick/up, slack-off, …) are compared. A comprehensive analysis of these measurements should provide information about the current downhole situation. Therefore, the data trends are analysed to identify any recognizable trends, deviations, or patterns that could indicate the onset of the encountered problem. This is essential because the issue may have been present and undetected for several minutes or even hours before the necessary trip out (starting at 18:00 of the second day). Any data related to poor hole cleaning should therefore not be included for training the predictive model.

### 4.2.2.1 Hydraulic Trend Analysis During Actual Drilling

To identify potential trend development across the drilled depth interval, the hydraulic trends during the actual drilling operation are reviewed. This includes analysing the trends of sensor measurements stemming from SPP, flowrate, RPM, and mud weight respectively, during drilling of individual stands. For representation box plots are chosen to visualize the distribution of these measurements, because they provide a good statistical summary of the parameters of interest. The bottom and the top of the blue box indicate the 25th and 75th percentiles of the sample data respectively. The height of the box represents the Interquartile Range (IQR). The median of the data is indicated with the red line inside the box. The black lines extending from the top and the bottom of the blue box plot indicate the range or spread of the data, excluding identified outliers. The red plus sign represents the flagged outliers, where a datapoint is considered to be an outlier when 1.5 times the IQR away from the bottom or the top of the blue box (25th and 75th percentile) (MathWorks, 2024).

Figure 20 offers a statistical overview of the SPP measurements at different depths of the borehole, specifically depicting the standpipe pressure values of each individual stand drilled. Each box plot corresponds to a specific stand. What can be noticed is that for drilled stands in interval depth interval 4733.9 - 5109.4 ft, SPP showed a normal trend, with an acceptable range. Interval 5109.4 - 5203.3 ft shows the highest IQR and whiskers bottom and top, indicating a large spread. SPP in the interval 5203.3 - 5297.2 ft shows a very small box (IQR), and short whiskers, indicating a low spread and relatively consistent SPP reading. Afterwards, for the following intervals/stands, SPP is increasing continuously as indicated by the median and box position of the box plots. Figure 21 - Figure 23 represent flowrate, RPM and MWIN,

respectively and have been included to offer supplemental details and enhance the understanding of the presented data.
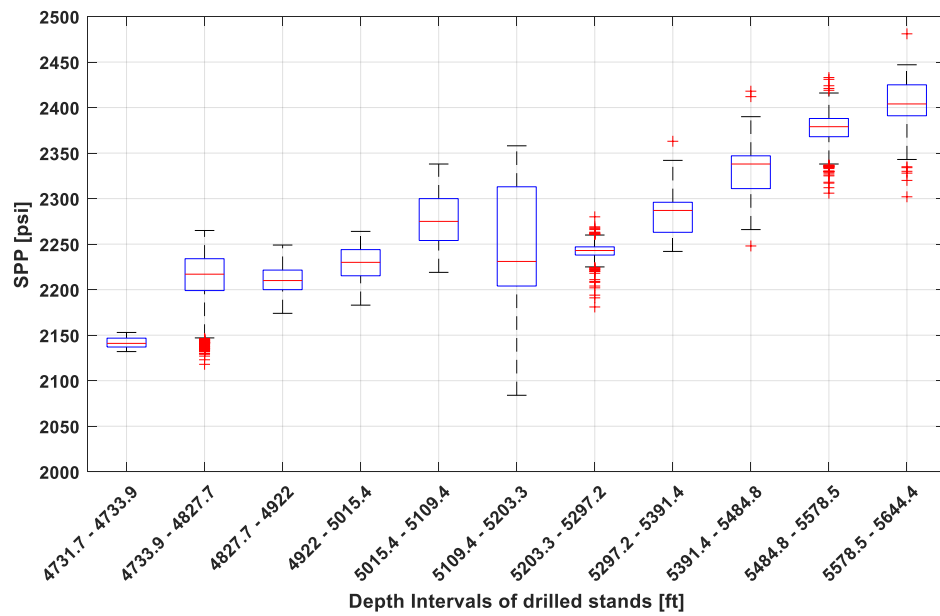


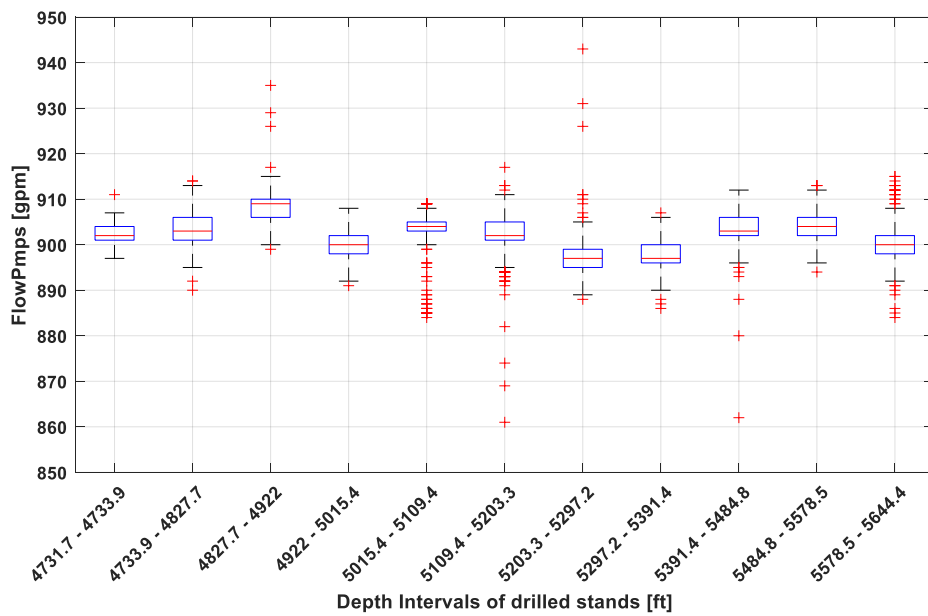*Figure 20 – Box plots of SPP during drilling of individual stands.*



*Figure 21 – Box plots of FlowPmps (Flowrate) during drilling of individual stands.*
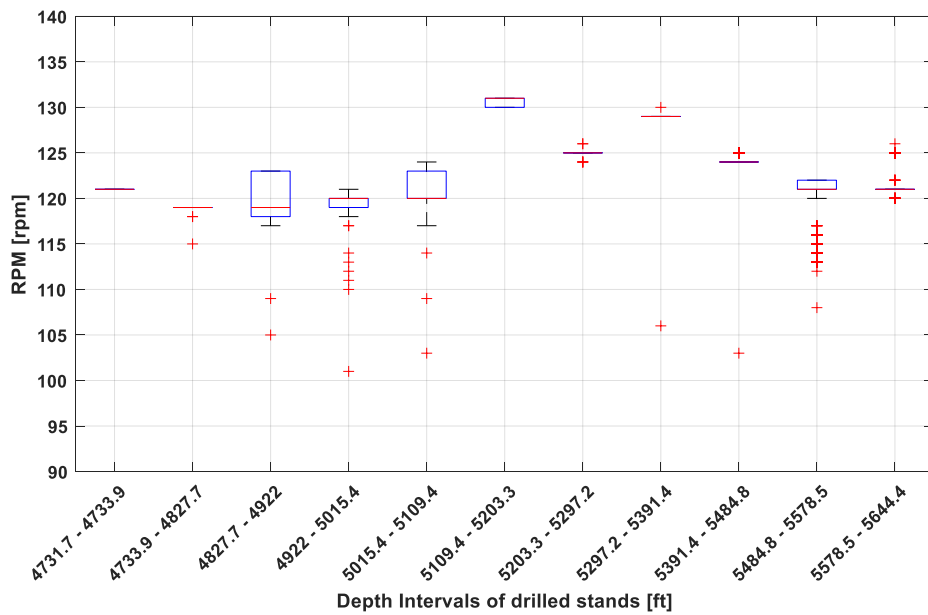
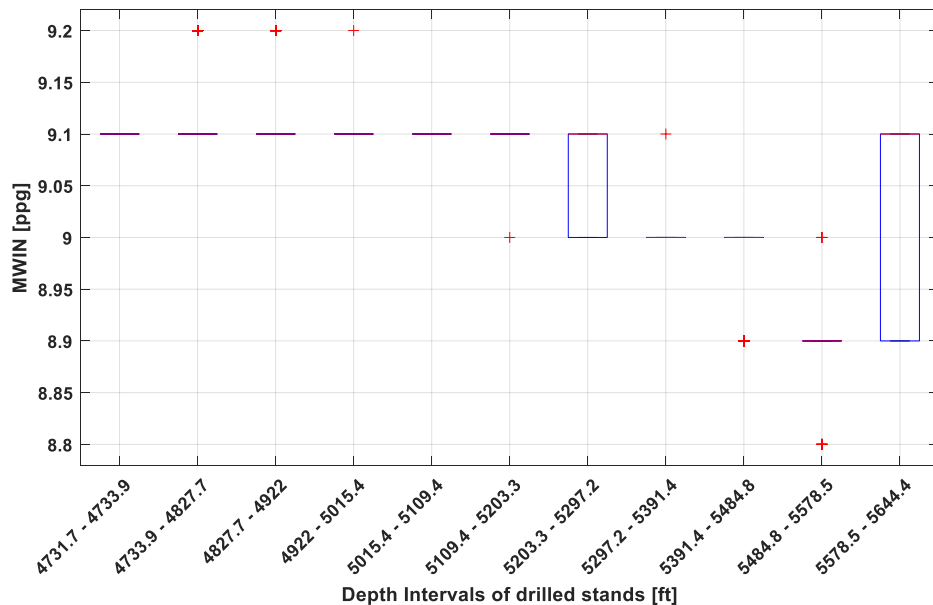*Figure 22 – Box plots of RPM during drilling of individual stands.*



*Figure 23 – Box plots of MW IN (Mud Weight In) during drilling of individual stands.*

## 4.2.2.2 Back Reaming – Reaming Hook Load Trend Before Connection

The second trend analysis was performed for back reaming / down reaming at the end of a drilled stand. This standard rig activity was selected because it was carried out continuously throughout the 48-hour period of available data and, most importantly, with similar operational parameters (RPM, flowrate, mud weight, time). At the end of each drilled stand, the rig crew reamed back and down the drilled stand, with reduced RPM and flowrate. This is a common routine rig activity in a well construction process attempting to clean out and smoothen the

wellbore before the next connection is made. It was decided to analyse the hook load reading for back ream (POOH) and ream down (RIH), because a deviation from a clear trend can indicate a change in current downhole condition. As mentioned earlier, that the problem is most probably related to hole cleaning issue, this analysis should aid in confirming the changing downhole condition.

Figure 24 shows the hook load (WOH) trend for back reaming (POOH) at the end of each drilled stand. The WOH is plotted versus the current total depth, corresponding to where the connections are made. As indicated by the orange dotted line, the hook load at the depth of 5578.5 ft slightly deviates from the trend when compared with previous readings. This depth 5578.5 ft is at the same time the depth of the last connection before the trip out. This possibly indicates a changing downhole condition for the drilled interval 5484.8 - 5578.5 ft until the drilling operation stopped at a MD of 5644 ft. Figure 25 shows the corresponding trend of WOH for ream down (RIH).
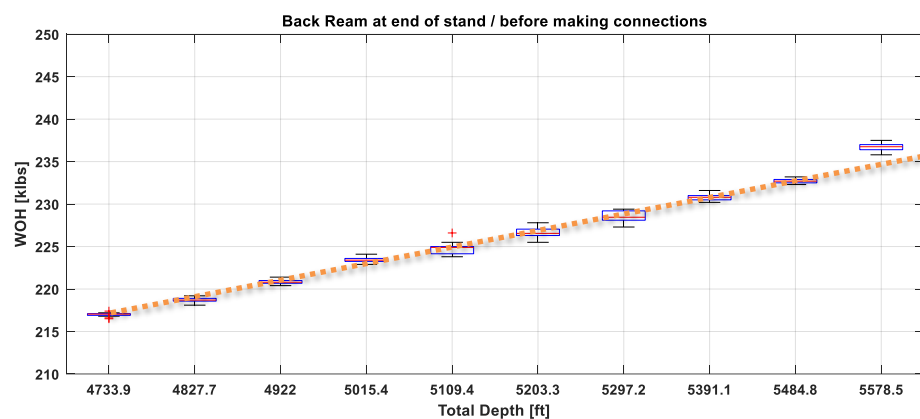


*Figure 24 – Box plots of WOH during backreaming at end of a stand, before next connection.*
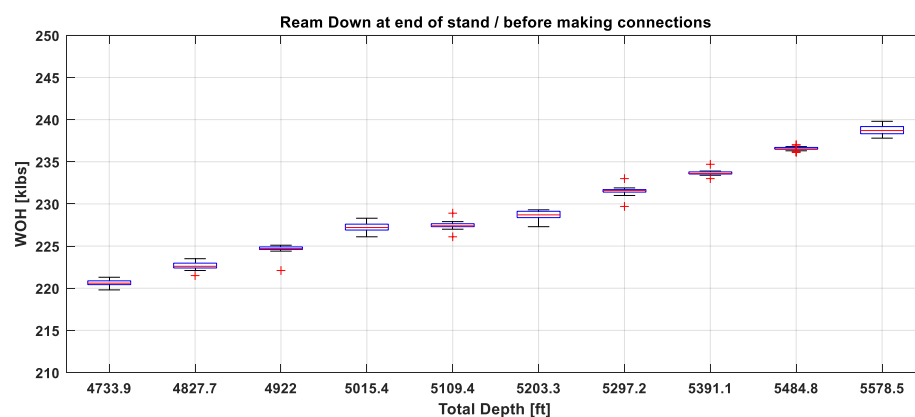


*Figure 25 – Box plots of WOH during reaming at end of a stand, before next connection.*

Also, what needs to be noted is the behaviour of the hook load encountered during backreaming (POOH) and reaming down (RIH) at the end of each stand, as standard practice carried out for keeping the borehole in a good condition. As shown in  Figure 26, where the hook load (Weight

on Hook) is plotted against depth for this rig activity. The depth indicates the depth at the end of a drilled stand, when the crew reamed up/down for hole conditioning purposes and preparing for the upcoming connection. Down reaming created higher friction, resulting in higher hook loads (Weight on Hook) readings than reaming up. This may be caused either due to a tight hole problem or a general hole condition problem.
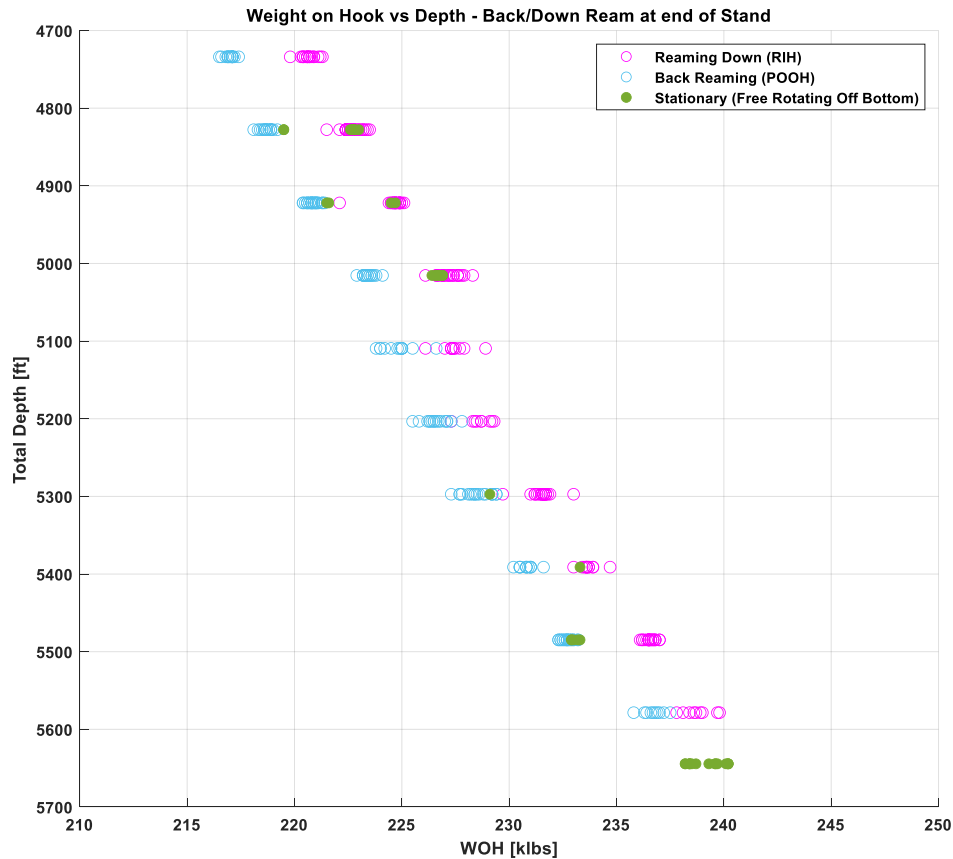


*Figure 26 – WOH vs Depth during backreaming, reaming at the end of a stand, before next connection.*

As stated earlier, during drilling the crew accounted, most probably a hole cleaning problem. Therefore, it was vital do identify the approximate time where this hole cleaning problem started or become visible in the drilling data channels, and to not include data containing hole cleaning problems in when building the predictive model.

### 4.2.2.3 Drilling Data Analysis Results

As seen in the box plots (Figure 20 - Figure 23) representing the hydraulic parameters during drilling and the box plots (Figure 24 and Figure 25) representing the hook load during reaming back and down at the end of each stand drilled, it is decided to exclude any data from depth intervals 5484.8 - 5578.5 ft and 5578.5 - 5644.4 ft respectively, since between drilling these

two stands, an increase in hook load during backream/ream is noticed, deviating from the trend, when compared to readings stemming from previous depth intervals.

## 4.3　Results and Discussion

This section contains the results of the modelling approach. At this point in the workflow the data has been analysed, verified, numerically cleaned, and drilling related datapoints are divided into subsets of data according to the drilled stand. This incident-free and clean drilling data is now used to train the predictive models. Models trained and tested on three different stands are chosen for representation and result discussion (Table 5).

*Table 5 – Overview of chosen cases for the predictive models.*

| Case Name | Stand Number | Depth Interval [ft] | Time [Day] |
|-----------|--------------|---------------------|------------|
| Case 1 | 2 | 4733.9 – 4827.7 ft | Day 1 |
| Case 2 | 7 | 5203.3 – 5297.2 ft | Day 1 |
| Case 3 | 9 | 5391.4 – 5484.8 ft | Day 2 |

### 4.3.1　Case 1

Case 1 presents the predictive model results from the drilled depth interval starting from 4733.9 – 4827. ft. Table 6 contains general information related to the selected subset of data. The number of total observations represents datapoints corresponding to actual drilling activity. Also, the selected training and reserved test set sizes are shown below.

*Table 6 – Case 1: Chosen subset of data for training and testing the predictive model for the selected depth interval.*

| | |
|---|---|
| Stand Number in Data Set | # 2 |
| Depth Interval | 4733.9 - 4827.7 ft |
| Total Observations | n = 1508 (~ 4.19 hours) |
| Training Data Points | n = 720 (2 hours) |
| Testing Data Points | n = 788 (~ 2.19 hours) |

Table 7 presents a statistical summary of target variable (SPP) and the selected features (flowrate, RPM and WOB) of the selected training set.

*Table 7 – Case 1: Statistical summary of the training dataset for case 1. SPP is the target variable.*
*Flowrate, RPM and WOB are the selected features.*

| Parameter | SPP [psi] | Flowrate [gpm] | RPM [rpm] | WOB [klbs] |
|---|---|---|---|---|
| Min | 2118 | 896 | 119 | 23.7 |
| Max | 2265 | 912 | 119 | 41.6 |
| Data range | 147 | 16 | 0 | 17.9 |
| Mean | 2201 | 904.43 | 119 | 37.915 |
| Std | 41.41 | 3.3196 | 0 | 3.648 |
| n | 720 | 720 | 720 | 720 |

The training dataset presented in Table 7 is used to train the predictive model, where in this specific case, 720 observations are used for training, corresponding to 2 hours of actual drilling based on 0.1 Hz frequency of the related measurements. The performance of the model is evaluated using 5-fold cross-validation on the training dataset, as outlined in section 2.2.4 Machine Learning Workflow. The optimization of the hyperparameters is performed during training of the model. The resulting model parameters are shown in Table 8. The averaged performance metrics of the model are calculated for both the validation and testing datasets and displayed in Table 9. Model performance metrics chosen to evaluate and interpret model predictions are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and R-Squared ($R^2$). The results show that R-Squared value for the cross validation of the validation dataset is 0.27, however R-Squared for the test set has a completely different value of -2.2043. On the other hand, other metrics such as MAE and RMSE achieved very good results with a MAE and RMSE of 17.557 psi and 22.868 psi, respectively, for the test set.

*Table 8 – Case 1: Model parameters.*

| Model Parameters | |
|---|---|
| Kernel | ARD Squared Exponential |
| Basis Function | Constant |
| Standardize | true |

*Table 9 – Case 1: Model performance metrics.*

|  | MAE [psi] | MSE [psi] $^2$ | RMSE [psi] | MAPE [%] | $R^2$ |
|---|---|---|---|---|---|
| Validation | 27.868 | 1240 | 35.213 | 1.2693 | 0.27574 |
| Test | 17.557 | 522.95 | 22.868 | 0.78844 | -2.2043 |

A visualization of the results is shown in Figure 27, where the blue trend represents the datapoints used to train the predictive model, the red trend is the actual sensor measurement (true value) and the green trendline represents the predicted values of SPP, with the green shaded region depicting the 95% confidence interval of the predictions. The confidence interval around the mean prediction indicates the range in which the true value is likely to lie. As shown in Figure 27, the first 720 datapoints (equal to 2 hours) of drilling stand #2 are used to train the predictive model, whereas the remaining (future) datapoints of the corresponding stand are reserved to make predictions. The SPP values (target) used for model training show an increasing trend during this two-hour period. At the point where the regression starts, the predictions from the model are lower than the true value of SPP, seemingly like an underestimation, according to the given input features (flowrate, RPM, WOB). However, around Data Index 1000 in the plot, the true values and modelled values converge. From that point on, the difference of the predicted value and the true value decreases, showing a better fit in this time interval. The green shaded region represents the 95% confidence interval, providing a range of values where the true value is likely to be. In this case, the predicted range is approximately +/- 75 psi, which correspond approximately to the 2 standard deviation distance to the mean. Consequently, an acceptable range of values is provided by the predicted confidence interval at the input locations. Considering the practical applicability, the confidence interval and mean predictions can aid in identifying abnormal downhole behaviour.
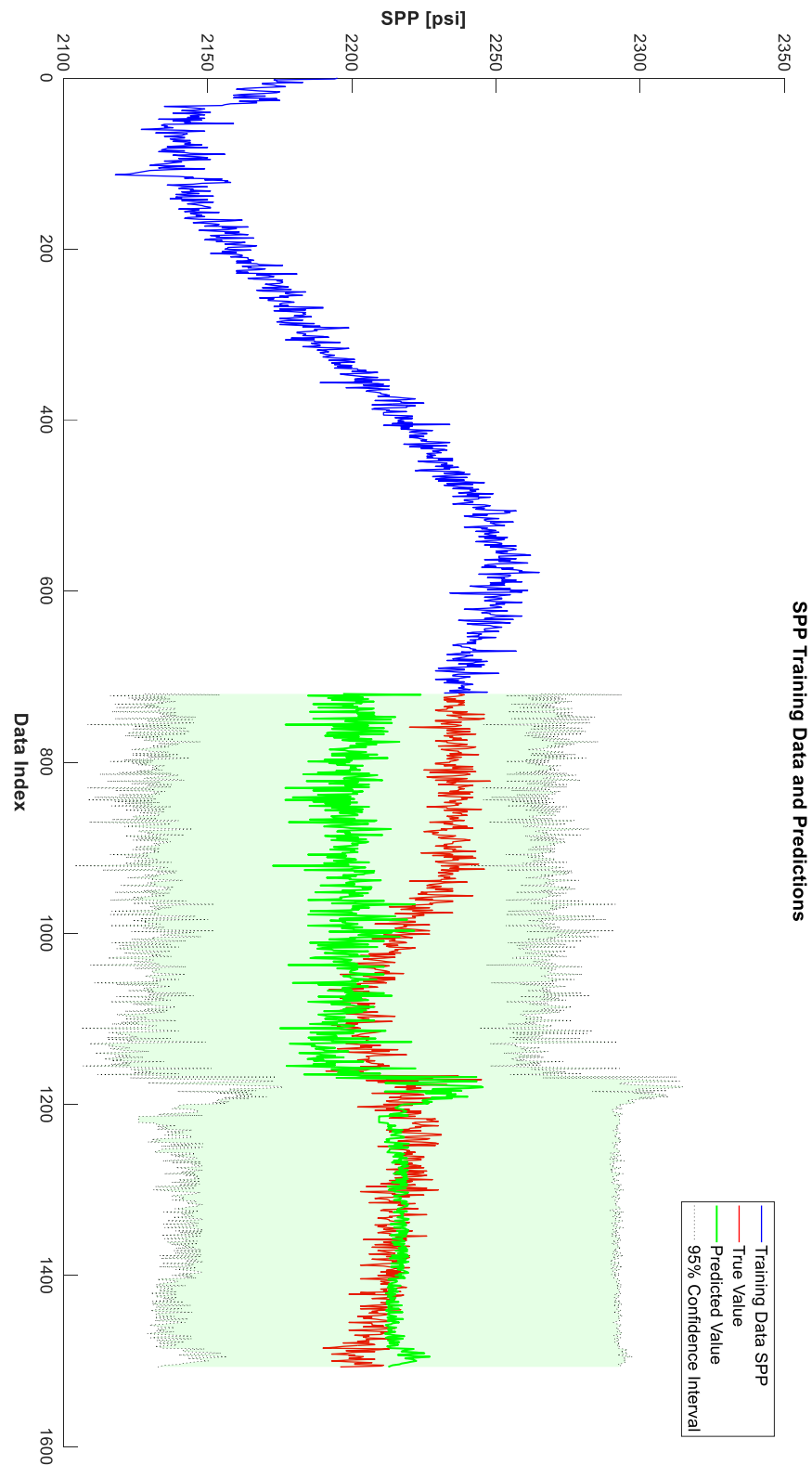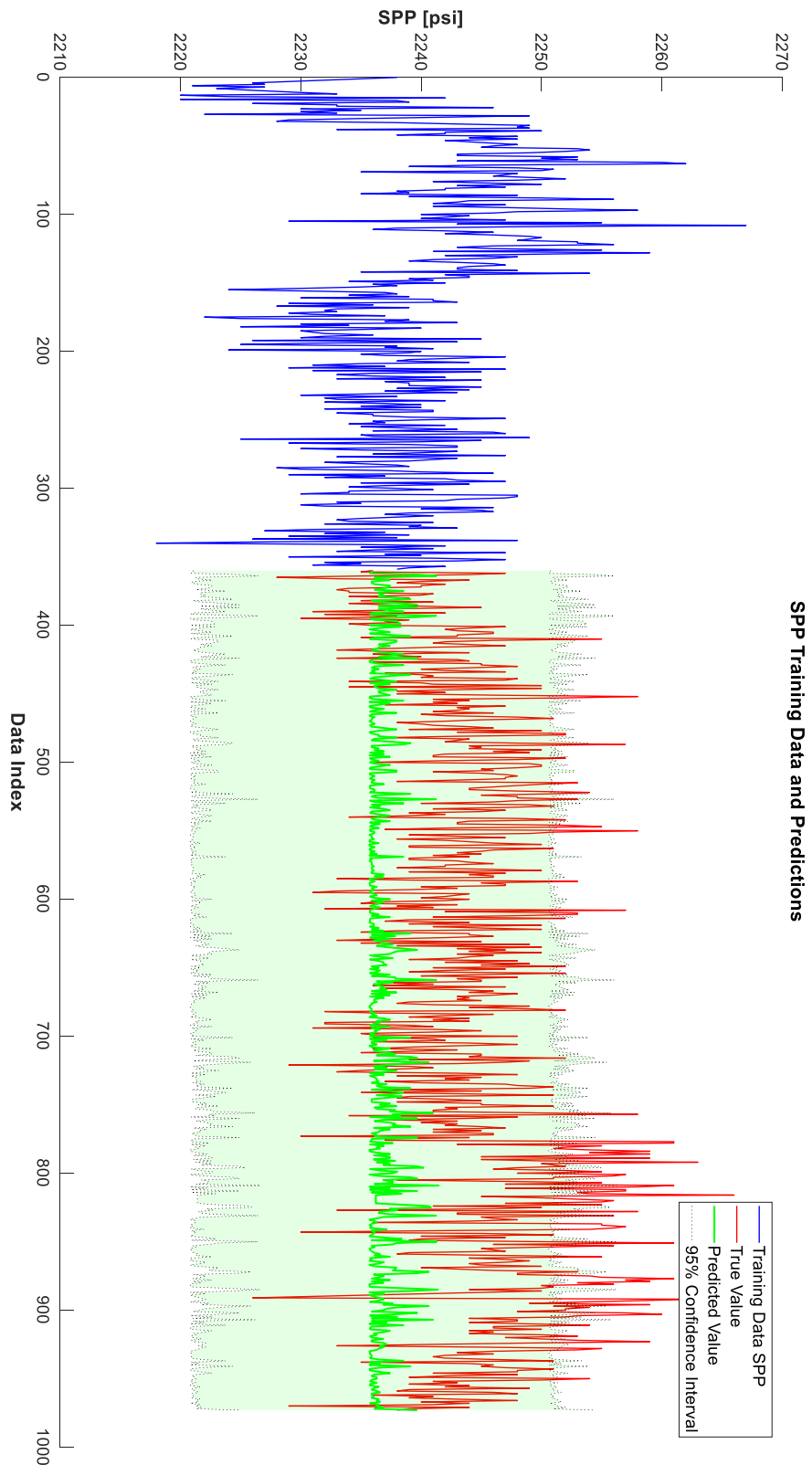
*Figure 27 – Case 1: Plot of SPP trend vs Data Index during drilling of stand #2. Blue trendline represents trend of the data used for training the model. Green trend is the predicted value, the green shaded region indicates the predicted 95%-confidence interval, and the red trend represents the actual sensor measurement (true value).*

## 4.3.2 Case 2

Another case chosen for presentation is Case Number 2. In this particular example, the methodology remained consistent, but the emphasis of the investigation shifted towards applying the approach to another subset of data with a smaller dataset size. Another important fact to mention is that the trend of SPP during drilling this stand showed a relatively steady behaviour. The dataset chosen represents the drilled interval from 5203.3 – 5297.2 ft. Table 10 provides general information about the training and testing data. Since drilling this stand was conducted in less time, fewer datapoints for training the model were available.

*Table 10 – Case 2: Chosen subset of data for training and testing the predictive model for the selected depth interval.*

| | |
|---|---|
| Stand Number in Data Set | # 7 |
| Depth Interval | 5203.3 – 5297.2 ft |
| Total Observations | n = 974 (~2.71 hours) |
| Training Data Points | n = 360 (1 hour) |
| Testing Data Points | n = 614 (~ 1.71hours) |

Table 11 contains a statistical summary of the target variable and the selected features of the training dataset. The SPP reading during drilling this stand has a low data range and standard deviation, meaning that during drilling SPP showed a relatively constant trend.

*Table 11 – Case 2: Statistical summary of the training dataset for case 2. SPP is the target variable. Flowrate, RPM and WOB are the selected features.*

| Parameter | SPP [psi] | Flowrate [gpm] | RPM [rpm] | WOB [klbs] |
|---|---|---|---|---|
| Min | 2218 | 890 | 124 | 31.3 |
| Max | 2267 | 907 | 125 | 36.7 |
| Data range | 49 | 17 | 1 | 5.4 |
| Mean | 2239 | 898.4 | 124.9 | 34.3 |
| Std | 7.8 | 2.99 | 0.07 | 0.5 |
| n | 360 | 360 | 360 | 360 |

The parameters of the model are shown in Table 12. Additionally, Table 13 presents the error measures obtained for the validation and test dataset used in this case. Again, R-Squared value shows unsatisfying results. However, considering the MAE and RMSE for the test set, the resulting model has an error of 8.471 psi and 10.171 psi, respectively, when predicting future values with current input parameters.

*Table 12 – Case 2: Model parameters.*

| Model Parameters | |
|---|---|
| Kernel | ARD Squared Exponential |
| Basis Function | Constant |
| Standardize | true |

*Table 13 – Case 2: Model performance metrics.*

| | MAE [psi] | MSE [psi] $^2$ | RMSE [psi] | MAPE [%] | $R^2$ |
|---|---|---|---|---|---|
| Validation | 6.076 | 57.863 | 7.607 | 0.27 | 0.041 |
| Test | 8.471 | 103.46 | 10.171 | 0.377 | -1.4348 |

Figure 28 showcases a graphical depiction of the model´s output for drilled stand #7. The first 360 datapoints (1 hour of drilling activity), indicated in blue, are used for training the predictive model and predictions are made for the remaining datapoints of the corresponding stand (test). The red trend represents the true value, the actual sensor measurement. The green trend and green shaded region depict the predicted mean and the predicted confidence interval respectively. As can be seen, the trend of SPP used for model training shows no significant increasing nor decreasing trend. Predicting future values of SPP based on given input features provides also acceptable results of the predicted mean. The predicted confidence interval in this section is narrow, where approximately 15 psi represent two standard deviations. Comparing the predicted confidence interval of stand #7 with the confidence interval predicted in the previous case, from stand #2, it can be seen, that depending on the distribution of the training data set (target and corresponding features) the predicted confidence interval can vary significantly. The predicted narrow confidence interval in this case indicates a lower uncertainty related to the predictions, providing a lower range of possible values according to the input features. Also, some datapoints are outside of the predicted confidence interval, especially from Data Index 800 to 1000, in theory indicating anomalous behaviour. However, in practice, a change of 15 psi of the standpipe pressure reading is in most of the times not significant. While the predicted confidence interval can't be directly used to identify anomalous downhole behaviour, it still provides a good range of possible values that can be helpful for assessing the overall drilling condition.
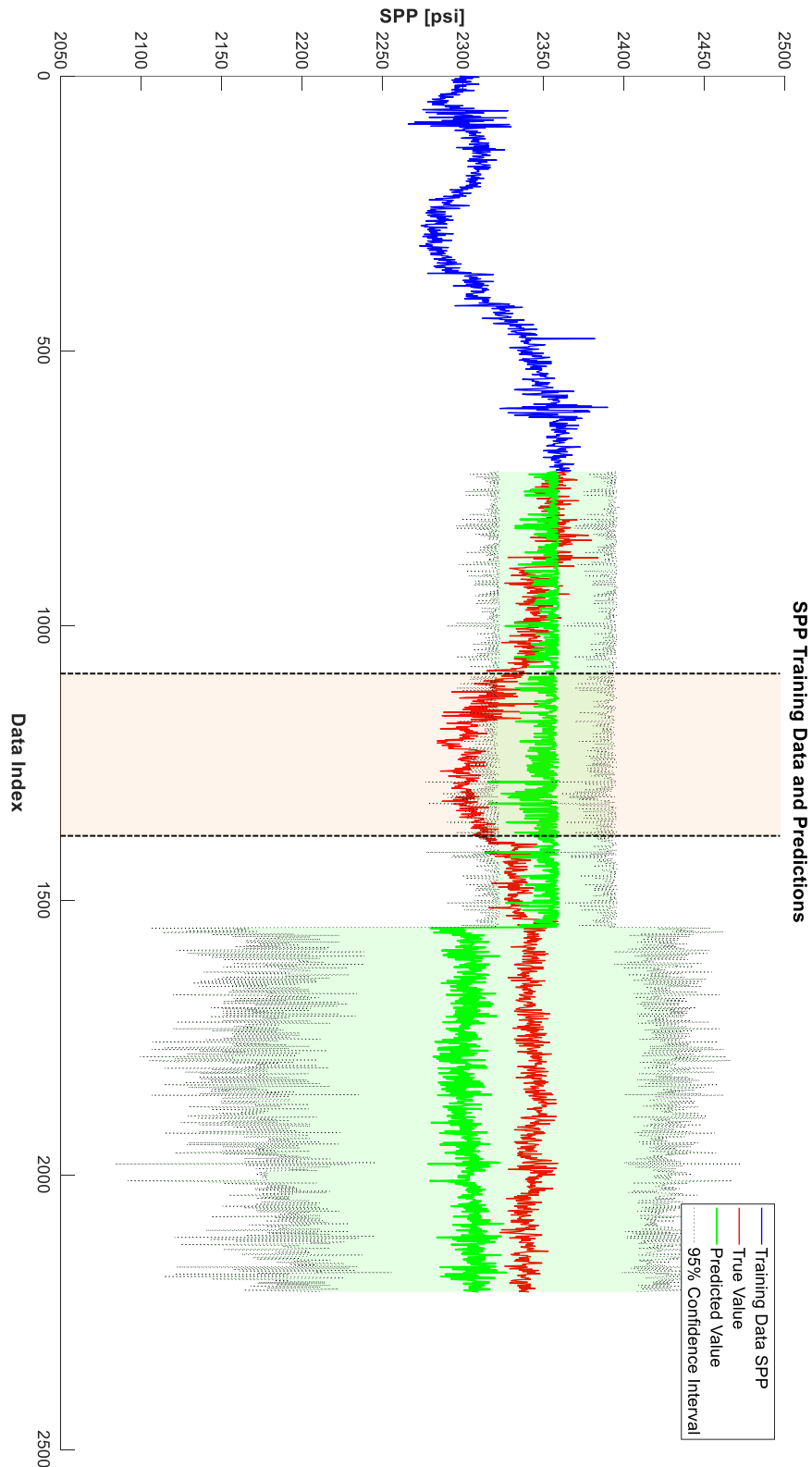
*Figure 28 – Case 2: Plot of SPP trend vs Data Index during drilling of stand #7. Blue trendline represents trend of the data used for training the model. Green trend is the predicted value, the green shaded region indicates the predicted 95%-confidence interval, and the red trend represents the actual sensor measurement (true value).*

### 4.3.3 Case 3

Last case selected for presentation is Case Number 3. This particular case refers to the drilled depth interval of 5391.4 – 5484.8 ft and represents the final stand to be eligible to be used for training, according to the performed data analysis. General overview of the selected dataset is provided in Table 14. In this case, a training set size of 720 datapoints was chosen to train the predictive model.

*Table 14 – Case 3: Chosen subset of data for training and testing the predictive model for the selected depth interval.*

| | |
|---|---|
| Stand Number in Data Set | # 9 |
| Depth Interval | 5391.4 – 5484.8 ft |
| Total Observations | n = 2234 (~6.21 hours) |
| Training Data Points | n = 720 (2 hours) |
| Testing Data Points | n = 1514 (~4.21 hours) |

A statistical summary of the datapoints used for training the predictive model are shown in Table 15.

*Table 15 – Case 3: Statistical summary of the training dataset for case 3. SPP is the target variable. Flowrate, RPM and WOB are the selected features.*

| Parameter | SPP [psi] | Flowrate [gpm] | RPM [rpm] | WOB [klbs] |
|---|---|---|---|---|
| Min | 2266 | 897 | 124 | 24.5 |
| Max | 2390 | 912 | 125 | 33.2 |
| Data range | 124 | 15 | 1 | 8.7 |
| Mean | 2319.3 | 904.76 | 124.12 | 28.585 |
| Std | 28.38 | 2.720 | 0.31 | 2.7204 |
| n | 720 | 720 | 720 | 720 |

Model parameters are shown in Table 16 and resulting model performance metrics are displayed in Table 17 for the validation and test set respectively. Optimisation of the model parameters, resulted in a different kernel, compared to the previous presented models. The resulting R-squared score for the validation set is 0.569, which is significantly higher than for previous models. The performance metrics for the test set, in particular MAE and RMSE which are 30.917 psi and 35.381 psi respectively, are acceptable.

Figure 29 shows a visual representation of the modelling results for drilled stand #9. The blue trendline corresponds to the data used for training (2 hours of actual drilling activity). The red

trendline represents the actual sensor measurement. The green trendline and the green shaded region represent the predicted value and the predicted confidence interval. As can be seen in this figure, standpipe pressure measurement of the training data exhibits a certain increasing trend. Then it is tried to predict and regress on the remaining datapoints of this drilled stand based on the input features. When the regression starts at Data Index 721 in this figure, the predicted mean seemingly fit the true value. The model shows two significantly different predicted confidence intervals, where around Data Index 1500, the predicted confidence interval changes notably. Mean predictions and predicted confidence interval of SPP from 721-1500 seem to be good with a predicted range of possible values of +/- 50 psi. It is noteworthy that a few datapoints fall outside the predicted range, indicated by the orange shaded region. Consultation with drilling experts and crosschecking other related drilling parameters at this corresponding interval revealed, that there might have been a borehole washout scenario. Afterwards, the true value and modelled value approach again until Data Index 1500. At this point, the predicted confidence interval changed, resulting in an increased range of possible values for SPP (+/- 150 psi). The explanation for this increase in model uncertainty can be found in the feature space of the test set. Drilling operational parameters changed. And the features used for model training, didn't cover this domain. Therefore, the model predictions get more uncertain, represented by the wider range of the predicted confidence interval. So, the model does also quantify the uncertainty related to the input parameters. Not only is the confidence interval wider, but also there is a greater discrepancy between the predicted and actual values.

*Table 16 – Case 3: Model parameters.*

| Model Parameters | |
|---|---|
| Kernel | ARD Matern 3/2 |
| Basis Function | None |
| Standardize | true |

*Table 17 – Case 3: Model performance metrics.*

| | MAE [psi] | MSE [psi]$^2$ | RMSE [psi] | MAPE [%] | $R^2$ |
|---|---|---|---|---|---|
| Validation | 14.088 | 346.05 | 18.602 | 0.6081 | 0.569 |
| Test | 30.917 | 1251.8 | 35.381 | 1.326 | -3.1928 |

*Figure 29 – Case 3: Plot of SPP trend vs. Data Index during drilling of stand #9. Blue trendline represents trend of the data used for training the model. Green trend is the predicted value, the green shaded region indicates the predicted 95%-confidence interval, and the red trend represents the actual sensor measurement (true value). The orange shaded region indicates possible borehole washout.*

## 4.4   Main Findings of the Studied Cases

The case study shows the results of the predictive models and the applied methodology. The main findings of this study show, the possible practical applicability as well as the constraints of the approach. The models rely on trouble-free drilling data from the currently drilled wellbore, eliminating the need for historical data, and required preprocessing steps are uncomplicated. The parameter of interest, could be predicted and modelled quite accurate using the developed methodology, as showcased by the three case studies. Low MAE, MSE, and RMSE were achieved with the use of only three controllable surface parameters to make the predictions. The predicted confidence interval does provide a range of possible values for the target variable according to the given set of features, thus providing means for analysis and SPP trend identification. But the predicted confidence interval cannot be directly interpreted as a safe operational window, according to the author´s opinion. The two model outputs, the predicted confidence interval as well as the predicted mean, highly rely on the used training dataset e.g. in Case 1 & Case 3, during actual drilling SPP showed a certain increasing trend and hence the training dataset included this information, resulting in wider confidence interval of the predictions – a higher range of possible values of SPP according to the set operational parameters. In Case 2 on the other hand, the SPP trend used for model training was more continuous, resulting in a narrow confidence interval of the predictions, where an increase of 15 psi would have been marked as anomalous. Also, as only data related to actual drilling is used to train the predictive models, the variance in the feature dimensions (flowrate, RPM and WOB) is reduced, while the standpipe pressure can exhibit a certain trend during drilling.

## 4.5   Model Limitations

Training the predictive models with data only related to drilling activity also has a limitation. The selected features may not fully be able to describe the behaviour of the target variable in some cases and could be the reason for bad performance in some intervals. Also, as SPP is dependent on various factors, including depth/length of the drillstring, the model doesn't exactly account for this. Therefore, it is possible, that the model could be optimized by adding features representative for wellbore/drillstring elongation. In addition to that, maybe more sophisticated sensor signal preprocessing techniques can result in better model performance than the simple approach tried in this thesis. Also, the predicted confidence interval cannot be directly seen as safe operational window, and it highly relies on the provided training data. However, the provided range of possible values can maybe be coupled or compared with other sophisticated methods which create a safe operational window and assess the different results. Besides that, as the model requires trouble-free drilling data, there is the possibility, that corrupted data may have been used for model training, although drilling data analysis was

performed carefully and thoughtfully. Therefore, it would be beneficial to test a similar approach with the use of another dataset, which can be considered 100% trouble free. This could result in better predictions and better representation of the trouble-free standpipe pressure trend during drilling.

# Chapter 5

# Conclusion and Future Work

## 5.1 Summary

The thesis provides a sophisticated approach and methodology to build a predictive model based on Gaussian process regression for monitoring and modelling SPP. The general approach of using 1-2 hours of trouble-free actual drilling data to make predictions of the hydraulic trend for the next 1-2 hours of drilling with only three controllable drilling surface parameters as input features, namely flowrate, WOB and RPM, is achieved with acceptable results, still needs to be critically reviewed. The methodology and the models were tested on different dataset sizes and for different available depth intervals, whereas the results for 3 cases are presented in this thesis. The choice and selection of training data, as well as the size of the dataset were found to have a significant influence on the model´s performance and outputs. The predicted mean with the predicted 95% confidence interval of the target variable could give, in some instances, acceptable results, given the complex behaviour of the target variable and surrounding environment, and can therefore also aid in decision making when drilling a stand. However, no good generalization of the models could be achieved. Also, the models do not require complex data preprocessing steps, they only require a limited amount of data stemming from the current wellbore to be drilled and that the data is considered to be trouble-free.

## 5.2 Challenges and Evaluation

The thesis aimed to predict the target value using a straightforward approach, limiting the model to only three controllable surface parameters to make predictions. However, the outcomes were only acceptable, and potential reasons for this will be tried to evaluate in the following paragraphs.

One possible problem can be related to the overall dataset, its quality, and the general challenge when working with real-time data. Due to the fact that the model relies on trouble-free drilling data, it still can be possible that there are bad intervals present in the data, which are not completely trouble-free, and which were not recognized during data analysis nor could be identified with cross checking with the DDRs.

Data cleaning and outlier identification was also accompanied by challenges. It was particularly difficult to distinguish between numerical outliers originating from a sensor measurement error/inaccuracy or operational issues. In real-time data analysis it can be difficult to distinguish between normal and abnormal behaviour, due to the fact that information is limited.

One limitation may also be caused by feature selection and dataset selection, which was primarily driven by expert knowledge and tailored to fulfil the overarching objective of the thesis, namely, to predict SPP and provide a possible range of values, with controllable surface parameters during actual drilling process. Reducing available data to datapoints only related to the actual drilling process, resulted in a significant reduction of variance in the data, especially in the feature dimensions. This becomes an issue if in certain drilled intervals the target variable shows an increasing or decreasing trend which cannot be described with the given features (e.g.: cuttings loading, mud alteration while drilling with high ROP). Therefore, the model will produce incorrect estimates. In this specific case, the addition of extra features (Total Depth, ROP) did not lead to improved model performance. On the contrary, it resulted in poorer predictions and significant increase in training time.

Likewise, there is a difficulty associated with the selection of appropriate training set sizes. Therefore, in this thesis two different training set sizes were tested (360 observations representing 1 hour of actual drilling, and 720 observations representing 2 hours of actual drilling) to build the predictive model. Model predictive power and performance was highly relying on the dataset size and the data selection in general. That said, if the target variable exhibits a certain increasing or decreasing trend, dataset size selection must be performed carefully. This in fact has a great influence in the predicted mean and predicted confidence interval.

Considering the mediocre results of the model performance metric, especially the value of R-squared was irritating. This may suggest that the selected features might not be able to fully explain the variance in the target variable, and this could be attributed to the fact that training of the models was performed utilizing datapoints limited to actual drilling operation (making hole). Also, while $R^2$ is commonly reported for various machine learning models in the literature to indicate the performance and generalization, the underlying assumption of $R^2$ is a linear relationship between the predictor and the target, which might not be totally valid in this

case, where the relation between target and features is believed to be non-linear. On the other hand, assessment of alternative model performance metrics such as MAE and RMSE, revealed better results. Low values of MAE and RMSE are observed, suggesting that the chosen approach was not entirely misguided. Also, the visualization of the predicted mean and the predicted confidence interval indicated quite positive results.

## 5.3 Future Work

The main research objective to predict the hydraulic trend with three controllable surface parameters with Gaussian process regression was only partially successful. Future work could be done on GPR and explore the application of GPR for modelling the hydraulic trend in more detail. Because up to now, there is limited literature available on the application of this specific algorithm applied for monitoring and modelling the hydraulic trend. Since in a machine learning workflow there is no universal solution from type "one size fits all", investigations should be performed also on the applied methodology. An adaption or improvement of the applied methodology or workflow to build the predictive model can result in better model performance. The preprocessing of the sensor signals could be done more sophisticatedly since the current approach only uses a minimum of preprocessing and data transformation techniques to reduce model complexity.

Furthermore, the same or an adapted methodology should be applied to a different dataset. Therefore, to achieve best possible results it would be beneficial to use a labelled or well-known drilling dataset, where there is proof of complete trouble-free drilling data to train the predictive model. This would make the data analysis part less exhaustive, and the focus could be shifted towards building and optimization of the predictive model. Also with a new dataset, it could be investigated if additional features like ROP or Total Depth would increase model performance. Whilst in this approach datapoints solely related to the actual drilling operation are used to build the predictive model, another approach of incorporating more routing drilling operations, specifically transient rig states, which may could have a positive effect.

Finally, to further evaluate the performance of the proposed model, it is essential to compare it with other, already existing models. In this regard, not only the mean predictions, but also the provided confidence interval of the proposed model should be compared and assessed, to evaluate the capability of the model.

# References

Alkinani, H. H., Al-Hameedi, A. T. T., & Dunn-Norman, S. (2020). Artificial neural network models to predict lost circulation in natural and induced fractures. *SN Applied Sciences*, *2*(12). https://doi.org/10.1007/s42452-020-03827-3

Alshaikh, A., Albassam, M. K., Al Gharbi, S. H., & Al-Yami, A. (2018). Detection of Stuck Pipe Early Signs and the Way Toward Automation, Article SPE-192975-MS.

Alshaikh, A., Magana-Mora, A., Al Gharbi, S., & Al-Yami, A. (2019). Machine Learning for Detecting Stuck Pipe Incidents: Data Analytics and Models Evaluation. *International Petroleum Technology Conference*, *2019*(IPTC-19394-MS).

Ambrus, A., Ashok, P., Ramos, D., Chintapalli, A., Susich, A., Thetford, T., Nelson, B., Shahri, M., McNab, J., & Behounek, M. (2018). Self-Learning Probabilistic Detection and Alerting of Drillstring Washout and Pump Failure Incidents During Drilling Operations. *IADC/SPE Drilling Conference and Exhibition*, *2018*, Article IADC/SPE-189700-MS.

Bachoc, F., Suvorikova, A., Ginsbourger, D., Loubes, J.-M., & Spokoiny, V. (2020). Gaussian processes with multidimensional distribution inputs via optimal transport and Hilbertian embedding. *Electronic Journal of Statistics*, *14*(2). https://doi.org/10.1214/20-EJS1725

Bishop, C. M. (2006). *Pattern recognition and machine learning. Information science and statistics*. Springer.

Bourgoyne, Jr, A T, Millheim, K. K., Chenevert, M. E., & Young, Jr, F S. (1991). *Applied drilling engineering*. Richardson, TX (United States); Society of Petroleum Engineers. https://www.osti.gov/biblio/5050497

Bousquet, O., Luxburg, U. von, & Rätsch, G. (Eds.). (2004). *Tutorial: Vol. 3176. Advanced lectures on machine learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003 [and] Tübingen, Germany, August 4-16, 2003 revised lectures*. Springer.

Brankovic, A., Matteucci, M., Restelli, M., Ferrarini, L., Piroddi, L., Spelta, A., & Zausa, F. (2021). Data-driven indicators for the detection and prediction of stuck-pipe events in oil&gas drilling operations. *Upstream Oil and Gas Technology*, *7*, 100043. https://doi.org/10.1016/j.upstre.2021.100043

Brownlee, J. (2019). *How to use Data Scaling Improve Deep Learning Model Stability and Performance*. https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/

Chowdhury, D., Skalle, P., & Rahman, M. M. (2009a). Prediction of Stand Pipe Pressure Using Conventional Approach. *Chemical Engineering Research Bulletin*, *13*(1). https://doi.org/10.3329/cerb.v13i1.2703

Chowdhury, D., Skalle, P., & Rahman, M. (2009b). Prediction of Standpipe Pressure Using Real Time Data. *International Conference of Mechanical Engineering*, Article ICME09-FM-02.

Duvenaud, D. K. (2014). *Automatic Model Construction with Gaussian Processes*. University of Cambridge.

Elmgerbi, A., Chuykov, E., Thonhauser, G., & Nascimento, A. (2022). Machine Learning Techniques Application for Real-Time Drilling Hydraulic Optimization. In *Day 1 Mon, February 21, 2022*. IPTC. https://doi.org/10.2523/IPTC-22662-MS

Elmgerbi, A., & Thonhauser, G. (2022). Holistic autonomous model for early detection of downhole drilling problems in real-time. *Process Safety and Environmental Protection*, *164*, 418–434. https://doi.org/10.1016/j.psep.2022.06.035

Erge, O., & van Oort, E. (2020). Combining Physics-Based and Data-Driven Modeling for Pressure Prediction in Well Construction. In *Proceedings of the Python in Science Conference, Proceedings of the 19th Python in Science Conference* (pp. 125–131). SciPy. https://doi.org/10.25080/Majora-342d178e-011

Erge, O., & van Oort, E. (2021). Hybrid Physics-Based and Data-Driven Modeling for Improved Standpipe Pressure Prediction. *SPE/IADC*, *2021*, Article SPE/IADC-204094-MS.

Erge, O., & van Oort, E. (2022). Combining physics-based and data-driven modeling in well construction: Hybrid fluid dynamics modeling. *Journal of Natural Gas Science and Engineering*, *97*, 104348. https://doi.org/10.1016/j.jngse.2021.104348

Forshaw, M., Becker, G., Jena, S., Linke, C., & Hummes, O. (2020). Automated Hole Cleaning Monitoring: A Modern Holistic Approach for NPT Reduction, Article IPTC-19639-Abstract.

Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2021). *Introduction to Statistical Learning: with Application in R*. Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Islam, R., Khan, F., & Venkatesan, R [Ramchandran] (2017). Real time risk analysis of kick detection: Testing and validation. *Reliability Engineering & System Safety*, *161*, 25–37. https://doi.org/10.1016/j.ress.2016.12.014

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jeong, C., Yu, Y., Mansour, D., Vesslinov, V., & Meehan, R. (2020). A Physics Model Embedded Hybrid Deep Neural Network for Drillstring Washout Detection. *IADC/SPE International Drilling Conference and Exhibition*, *2020*, Article IADC/SPE-199629-MS.

Karthik, B., Minou, R., Vural, S., Celal, H. C., Zinyat, A., Suleyman, T., Ummugul, B., & Cenk, T. (2018). Status of Data-Driven Methods and their Applications in Oil and Gas Industry, Article SPE-190812-MS.

Kehinde, S. A., Ajayi, O. I., Akpan, U., & Odesa, D. E. (2023). Bit Balling: Causes, Effects and Mitigation Using Bunmi-01 Well in Niger Delta as a Case Study. In *Day 3 Wed, August 02, 2023*. SPE. https://doi.org/10.2118/217094-MS

Khan, J. A., Irfan, M., Irawan, S., Yao, F. K., Abdul Rahaman, M. S., Shahari, A. R., Glowacz, A., & Zeb, N. (2020). Comparison of Machine Learning Classifiers for Accurate Prediction of Real-Time Stuck Pipe Incidents. *Energies*, *13*(14), 3683. https://doi.org/10.3390/en13143683

King, G. (2023, December 11). *9.2.3: The Circulation System | PNG 301: Introduction to Petroleum and Natural Gas Engineering*. Penn State. https://www.e-education.psu.edu/png301/node/726

Kuesters, A., Mason, C., Gomes, P., Cockburn, C., & Lodhi, H. (2020). Drillstring Failure Prevention - A Data Driven Approach to Early Washout Detection. *IADC/SPE International Drilling Conference and Exhibition*, *2020*, Article IADC/SPE-199610-MS.

Kumar, S. (2020). *Supervised vs Unsupervised vs Reinforcement - AITUDE*. AITUDE. https://www.aitude.com/supervised-vs-unsupervised-vs-reinforcement/

Lafond, A., Ringer, M., Le Blay, F., Liu, J., Millan, E., Ba, S., & Chao, M. (2021). Detecting Pressure Anomalies While Drilling Using a Machine Learning Hybrid Approach, Article SPE/IADC-204035-MS.

Leco, M., & Kadirkamanathan, V. (2021). A perturbation signal based data-driven Gaussian process regression model for in-process part quality prediction in robotic

countersinking operations. *Robotics and Computer-Integrated Manufacturing, 71,* 102105. https://doi.org/10.1016/j.rcim.2020.102105

Lyashenko, V., & Jha, A. (2022, July 21). *Cross-Validation in Machine Learning: How to Do It Right*. https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right

Mahesh, B. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, Article ISSN: 2319-7064.

MathWorks. *Introducing Machine Learning*. https://doi.org/10.1007/978-3-030-67626-1_8

MathWorks. (2024, January 25). *Visualize summary statistics with box plot - MATLAB boxplot - MathWorks United Kingdom*. https://uk.mathworks.com/help/stats/boxplot.html

Mitchell, R. F., & Miska, S. Z. (Eds.). (2011). *SPE textbook series: Vol. 12. Fundamentals of Drilling Engineering*. Society of Petroleum Engineers.

Mopuri, K. R., Bilen, H., Tsuchihashi, N., Wada, R., Inoue, T., Kusanagi, K., Nishiyama, T., & Tamamura, H. (2022). Early sign detection for the stuck pipe scenarios using unsupervised deep learning. *Journal of Petroleum Science and Engineering, 208,* 109489. https://doi.org/10.1016/j.petrol.2021.109489

Muojeke, S., Venkatesan, R [Ramachandran], & Khan, F. (2020). Supervised data-driven approach to early kick detection during drilling operation. *Journal of Petroleum Science and Engineering, 192,* 107324. https://doi.org/10.1016/j.petrol.2020.107324

Muqeem, M. A., Weekse, A. E., & Al-Hajji, A. A. (2012). Stuck Pipe Best Practices – A Challenging Approach to Reducing Stuck Pipe Costs, Article SPE 160845.

Nhat, D. M., Venkatesan, R [Ramachandran], & Khan, F. (2020). Data-driven Bayesian network model for early kick detection in industrial drilling process. *Process Safety and Environmental Protection, 138,* 130–138. https://doi.org/10.1016/j.psep.2020.03.017

Quinonero-Candela, J., & Rasmussen, C. E [Carl Edward] (2005). A Unifying View of Sparse Approximate Gaussian Process Regression. *Journal of Machine Learning Research, 6,* 1939–1959.

Rasmussen, C. E [C. E.], & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Schlumberger (2016). The Defining Series: Rheology.

scikit-learn. (2023, December 11). *3.1. Cross-validation: evaluating estimator performance*. https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation

Skalle, P., Aamodt, A., & Gundersen, O. E. (2013). Detection of Symptoms for Revealing Causes Leading to Drilling Failures. *Human Resources for Health*, *21*. https://doi.org/10.1186/s12960-023-00834-4

Sule, I. O., Khan, F., & Butt, S. (2018). Experimental investigation of gas kick effects on dynamic drilling parameters. *Journal of Petroleum Exploration and Production Technology*, *9*(1), 605–616. https://doi.org/10.1007/s13202-018-0510-z

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. *Adaptive computation and machine learning*. MIT Press.

Tang, H., Zhang, S., Zhang, F., & Venugopal, S. (2019). Time series data analysis for automatic flow influx detection during drilling. *Journal of Petroleum Science and Engineering*, *172*, 1103–1111. https://doi.org/10.1016/j.petrol.2018.09.018

Varadarajan, P. A., Roguin, G., Abolins, N., & Ringer, M. (2021). A Digital Twin for Real-Time Drilling Hydraulics Simulation Using a Hybrid Approach of Physics and Machine Learning. *Offshore Technology Conference*, Article OTC-31278-MS.

Wang, J. (2022). *An Intuitive Tutorial to Gaussian Processes Regression*. http://arxiv.org/pdf/2009.10862v4

Yin, Q., Yang, J., Tyagi, M., Zhou, X., Hou, X., & Cao, B. (2021). Field data analysis and risk assessment of gas kick during industrial deepwater drilling process based on supervised learning algorithm. *Process Safety and Environmental Protection*, *146*, 312–328. https://doi.org/10.1016/j.psep.2020.08.012

Youcefi, M. R., Hadjadj, A., & Boukredera, F. S. (2022). New model for standpipe pressure prediction while drilling using Group Method of Data Handling. *Petroleum*, *8*(2), 210–218. https://doi.org/10.1016/j.petlm.2021.04.003

Zhao, Y., Liu, S., Wang, Z., Ren, M., & Sun, B. (2019). An adaptive pattern recognition method for early diagnosis of drillstring washout based on dynamic hydraulic model. *Journal of Natural Gas Science and Engineering*, *70*, 102947. https://doi.org/10.1016/j.jngse.2019.102947

Zhong, R., Salehi, C., & Johnson, R. (2022). Machine learning for drilling applications: A review. *Journal of Natural Gas Science and Engineering*, *108*, 104807. https://doi.org/10.1016/j.jngse.2022.104807

# List of Figures

# List of Tables

# Nomenclature

| | |
|---|---|
| $X$ | Features |
| $Y$ | Target |
| $n$ | Observations |
| $X_p$ | observed value for p-example |
| $P$ | Pressure |
| $\Delta P$ | Pressure Differential |
| $\tau$ | Shear Stress |
| $\mu$ | Dynamic Fluid Viscosity |
| $\gamma$ | Shear Rate |
| $\mu_p$ | Plastic Viscosity |
| $\tau_y$ | Yield Point |
| $K$ | Consistency Index |
| $m$ | Flow Behaviour |
| $v$ | Velocity |
| $Q$ | Flowrate |
| $A$ | Area |
| $N_{Re}$ | Reynolds Number |
| $\rho$ | Density |
| $d$ | diameter |
| $E$ | Equipment Coefficient |
| $C$ | Discharge Coefficient |
| cP | Centipoise |

# Abbreviations

| | |
|---|---|
| BHA | Bottomhole Assembly |
| BHP | Bottomhole Pressure |
| CH | Cased Hole |
| CV | Cross-Validation |
| DC | Drill Collars |
| DDR | Daily Drilling Report |
| DP | Drill Pipe |
| ECD | Equivalent Circulating Density |
| GP | Gaussian Process |
| GPR | Gaussian Process Regression |
| HIVIS | High Viscous (Pill) |
| HL | Hook Load |
| IQR | Interquartile Range |
| LCM | Lost Circulation Material |
| LWD | Logging while Drilling |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MD | Measured Depth |
| ML | Machine Learning |
| MSE | Mean Squared Error |
| NPT | Non-Productive Time |
| OH | Open Hole |
| PCA | Principal Component Analysis |
| POOH | Pull out of Hole |
| RIH | Run into Hole |

RMSE                      Root Mean Squared Error

ROP                        Rate of Penetration

RPM                       Rotations per Minute

SPP                        Standpipe Pressure

WOB                      Weight on Bit

WOH                      Weight on Hook