



University of  
Stavanger

**FACULTY OF SCIENCE AND TECHNOLOGY**

# MASTER'S THESIS

<b>Study programme/specialisation:</b> Industrial Economics/Petroleum Engineering and Project Management	<b>Spring semester,</b>  <b>Open</b>
<b>Author:</b> Kristian Grimsmo Haug	<i>Kristian G. Haug</i>  (signature of author)
<b>Faculty supervisor:</b> Kjell Kåre Fjelde	
<b>Title of master's thesis:</b> Early kick detection and development of a simulator for generating data for machine learning purposes	
<b>Credits (ECTS): 30</b>	
<b>Keywords:</b> <ul style="list-style-type: none"><li>- Machine Learning</li><li>- Big Data</li><li>- Python</li><li>- Simulator development</li><li>- Kick</li><li>- Drilling</li></ul>	Number of pages: 87 + enclosure: 13  Stavanger, June 14 <sup>th</sup> 2021 date/year

Front page of master thesis

Faculty of Science and Technology

# Abstract

One of the most harmful incidents with regards to safety, economic losses, and environmental damages that can happen during drilling is a blow-out resulting from a kick obtained by an uncontrolled influx of formation fluid/gas. The blow-out is a result of the influx building up over a substantial amount of time without closing the BOP, where the driller completely loses control of the well. A crucial factor to mitigate this problem is therefore to identify the influx at an early stage, and start procedures to circulate the influx out of the well. Although there is substantial knowledge of several drilling parameter's behavior before and during obtaining a kick, this is not easy to detect in all incidents for a driller or by the alarm. Several factors can contribute to hide this kick, such as dissolution into oil based mud, transient flows, and poor quality of sensor equipment. However, the rapid advancements of machine learning technology used in the oil and gas industry shows a great potential of applying artificial intelligence for early influx and loss detection. The technology shows promising results in several cases, but is currently at the early stages of development.

In this thesis a simulator was developed for generating data of random influxes into a well in order to generate training and testing sets for machine learning algorithms. Here, the aim was to generate random data for each simulation as the volume and timing of the influx is drawn from a random distribution. Furthermore, this will generate a unique result for each run, making it fitting for generating several data sets used for both training and testing of machine learning models. The simulator is based on typical physical and mathematical aspects of how influxes impacts several surface parameters. It is still however limited to some assumptions for simplification and the lack of realistic noise levels. These implementations are seen as necessary before applying the data sets to machine learning algorithms.

The results show that the simulator is effective in generating a large quantity of data, which is necessary in training machine learning models. Additionally, the simulator proves to be flexible through two presented case studies, which in both cases provide results coinciding with the expected results based on the physical aspects of the event.

# Acknowledgement

First of all, I would like to thank my supervisor Kjell Kåre Fjelde for exceptional guidance and an endless supply of literature within kick detection and machine learning. His input on the development of the simulator has also been extremely valuable. I would also like to thank my co-supervisor, Mesfin Belayneh Agonafir, for his contributions towards providing relevant literature as well as very helpful feedback on my work. Lastly, I would like to thank Tim Robinson and Dalila Gomes from Exebenus for their great input on how the simulator could be applied for industry purposes.

Finally, I would like to thank my family and friends for their understanding, patience, and support throughout this period.

# Table of content

Abstract .....	II
Acknowledgement.....	III
List of figures .....	VI
List of Tables.....	VII
Nomenclature .....	VIII
1. Introduction.....	1
1.1 Background and motivation.....	1
1.2 Problem formulation and research questions.....	2
1.3 Scope and Objective of this thesis.....	3
1.4 Structure of thesis .....	3
2. Literature review .....	5
2.1 Kick and kick detection .....	5
2.1.1 Formation of kicks and influx management procedures .....	6
2.1.2 Obtaining influx/kick during different drilling processes.....	7
2.1.3 Kick detection parameters and measuring equipment.....	9
2.1.4 Oil based mud vs. Water based mud .....	11
2.2 Kick detection using Machine Learning.....	13
2.2.1 Understanding machine learning.....	14
2.2.2 Training the system and pre-processing data .....	16
2.2.3 Machine learning models researched for kick detection today .....	19
2.2.4 Decision Tree and Random Forests.....	20
2.2.5 K-Nearest Neighbors.....	21
2.2.6 Artificial Neural Network .....	23
2.2.7 Support Vector Machine .....	25
2.2.8 Machine learning models for kick detection - comparison .....	26
2.3 Potential economic aspects of an early kick detection .....	28
2.3.1 Cost reduction through digitalization and machine learning.....	28
2.3.2 Cost reduction through reduced non-productive time and early prediction.....	30
2.3.3 Costs of the Deepwater Horizon blow-out – an extreme case scenario .....	32
3. Random Influx Simulator .....	34
3.1 Basic assumption of the Simulator .....	35
3.2 Simulator structure.....	36
3.3 Impact of kick on drilling parameters.....	37

3.3.1	Pump Pressure .....	37
3.3.2	Bottom Hole Pressure.....	40
3.3.3	Equivalent Circulating Density .....	41
3.3.4	Rate of Penetration .....	42
3.3.5	Flowrate Out.....	44
3.3.6	Connection gas .....	45
3.3.7	Pit gain.....	48
3.3.8	Hook load .....	49
3.4	Application of model .....	52
3.4.1	Simulator base case .....	52
3.4.2	Case 1 - base case simulation .....	54
3.4.3	Case 2 – more connections, longer event period, and larger influx .....	60
4.	Discussion .....	67
5.	Conclusion .....	72
5.1	Further Work .....	73
	References .....	74
	Appendix .....	80
A.1)	Structure of the complete simulation model.....	80
A.2)	Structure of sub-systems: Drilling and Drilling with Kick.....	81
A.3)	Structure of sub-system: Connection.....	82
B.1)	Simulator base-code: .....	83

# List of figures

Figure 1: Macando Well Blow-out and its impacts on Fauna and Flora [2] .....	1
Figure 2: Kick with size of 4 bbl taken in 9 7/8" casing, with circulation rate 300 gpm [28] .....	12
Figure 3: How artificial intelligence, machine learning, deep learning and big data are related [30] .....	14
Figure 4: Illustration of a simple Decision Tree [38].....	20
Figure 5: Example of ANN used for kick detection.....	23
Figure 6: Importance of different input parameters [5].....	24
Figure 7: Visualization of a SVM model [6].....	25
Figure 8: Structure of plug and play machine learning system [48] .....	30
Figure 9: Cost savings by using preventive machine learning tools for early kick detection .....	32
Figure 10: Visualization of single bubble approximated kick trapped in drilling fluid .....	35
Figure 11: Code calculating the measured depth/TVD during drilling.....	37
Figure 12: Importance of different input parameters for ML kick detection [42] .....	43
Figure 13: Code for adding a sudden increase in ROP before kick is taken.....	44
Figure 14: Iterations to find time index of delayed connection gas indicator .....	46
Figure 15: Creating an array to store delayed connection gas at correct time index .....	47
Figure 16: Example of how connection gas is presented in a 5 connection simulation.....	47
Figure 17: Example of drilled depth during a 5 connection simulation .....	47
Figure 18: Comparison of raw pit volume data (left) vs. filtered pit volume data (right) [29].....	48
Figure 19: Comparison of pump pressure data with and without noise case 1 .....	55
Figure 20: Pump pressure last drilled section with and without noise .....	56
Figure 21: Comparison of bottom hole pressure data with and without noise case 1 (influx marked with red circle).....	56
Figure 22: ECD data from simulation without noise case 1 .....	57
Figure 23: Comparison of ROP data with and without noise case 1 (influx marked with red circle) .....	57
Figure 24: Comparison of flowrate out data with and without noise case 1 .....	58
Figure 25: Comparison of pit gain data with and without noise case 1 .....	58
Figure 26: Comparison of hook load data with and without noise case 1 .....	59

Figure 27: Hook load last drilled section with and without noise case 1 (influx section marked with red circle).....59

Figure 28: Comparison of pump pressure data with and without noise case 2 .....61

Figure 29: Pump pressure last drilled section with and without noise .....61

Figure 30: Comparison of bottom hole pressure data with and without noise case 2 (influx section marked with red circle).....62

Figure 31: ECD data from simulation without noise case 2 .....62

Figure 32: Comparison of ROP data with and without noise case 2.....63

Figure 33: Comparison of flowrate out data with and without noise case 2 .....63

Figure 34: Connection gas in an extended simulated number of connections for case 2.....64

Figure 35: Comparison of pit gain data with and without noise case 2 .....64

Figure 36: Comparison of hook load data with and without noise case 2 .....65

Figure 37: Hook load last drilled section with and without noise case 2 (influx section marked with red circle).....65

**List of Tables**

Table 1: Sensor description with corresponding noise mean and standard deviation [20] ..... 18

Table 2: Comparison of machine learning models used for kick detection .....27

Table 3: Initial input parameters base case simulator run .....54

Table 4: Overview of standard deviation in noise data.....55

Table 5: Initial input parameters case 2 simulator run .....60

# Nomenclature

ANN - Artificial Neural Network

BHA – Bottom Hole Assembly

DP – Drill Pipe

ECD – Equivalent Circulating Density

EDR - Electronic Drilling Recorder

HPHT - High Pressure and High Temperature

IDAPS - Influx Detection at Pumps Stop

KNN - K-Nearest Neighbors

LSTM - Long Short-Term Memory

MD – Measured Depth

MPD – Managed Pressure Drilling

NCS – Norwegian Continental Shelf

ROP – Rate of Penetration

SVM - Support Vector Machine

TVD – True Vertical Depth

WITMSL - Wellsite Information Transfer Standard Markup Language

WOB – Weight on Bit

# 1. Introduction

This thesis presents theory of conventional kick detection and the advancements of smarter kick detection methods by using machine learning for detecting anomalies in surface parameters signaling an incoming kick. Here, the main objective is to use this theory combined with the physics of a well's behavior during regular drilling and influx incidents to build a simulator for data generation. This data is intended to train and test machine learning algorithms for early kick detection by randomly generating kick data based on a probabilistic approach of random kick volume and occurrence time. All the data sets will resultingly be unique with different kick data, allowing for extensive training and testing to obtain a high detection accuracy in the machine learning algorithms when applied in real environments.

## 1.1 Background and motivation

When performing any drilling operation, there is always a possibility of unpredicted incidents such as obtaining unwanted influxes of formation fluid to the wellbore. This is commonly known as a kick and it should be controlled properly. If a kick is not controlled, it might in a worst-case scenario result in a surface blow-out condition. For instance, in the North Sea, the Piper Alpha blow out incidence in 1988 cost 167 human lives and several of serious injuries in addition to costs of billions of dollars [1]. Another recent blow-out accident, at the Deepwater Horizon platform in the Gulf of Mexico in 2010, caused 11 casualties and billions of USD in costs due to the oil spill resulting in negative environmental impacts as shown in Figure 1 [2]. Thus, early kick detection has great value in order to control the well safely and mitigate or avoid kick related negative impacts such as mentioned earlier.



Figure 1: Macando Well Blow-out and its impacts on Fauna and Flora [2]

Smaller influxes can be hard to detect from a driller's point of view, as the measurement data are containing a large amount of noise. Here, it is hard to detect abnormalities at a rather small scale, thus the driller will in several cases not detect the kick until it has reached a rather large volume, further complicating the situation. Another factor is that the ability to observe a kick is based on each driller's individual ability to interpret data. Hence, reaction time might vary to a large extent from location to location.

To counter this problem, the industry has in the recent years started researching the use of machine learning algorithms for early kick detection [3]. The algorithms are able to make predictions of drilling parameters, thus detecting abnormalities signifying that an influx is incoming. These kick detections are to be made at a relatively earlier stage compared to what the driller would be able to, meaning preventive actions can be initiated early in order to mitigate the problem. However, a full implementation of the system is yet to be tested, and the real-time input parameters still need a substantial amount of pre-processing and filtering for the machine learning algorithms to make accurate predictions. The tested models are also tested with experimental data or comprehensive pre-processed data, meaning results does not necessarily represent reality. Another challenge is that the data sets will vary vastly from well to well, thus training a model for detection in a new well might require data that is hard to acquire. This represents the need for a data generation simulator, that can generate random kicks based on some defined well settings, and use this as training data for the machine learning algorithms.

## **1.2 Problem formulation and research questions**

Several research papers present preliminary stages of testing machine learning algorithms for early kick detection [4]. However, these are limited by using either simulated data lacking realistic noise, or real well data with too much noise. Consequently, the machine learning algorithms will either make predictions with a higher accuracy due to no noise or make predictions with a low accuracy due to many outliers in the noisy data. This thesis will therefore address issues such as:

- Poor availability of kick data for machine learning purposes.
- Substantial need for pre-processing of real well data for noise removal.

- Difficulties in detecting kick at an early stage using conventional methods, without supportive tools.
- The predictive power of machine learning based modelling and forecasting of early kick occurrence.

### **1.3 Scope and Objective of this thesis**

The main objective of the thesis is to build a simulator for random kick generation with probabilistic volume and time of incident, where the effects of the kick are shown on different drilling parameters. Here, the purpose is to generate data that later can be applied in machine learning models for training and testing towards early kick detection purposes. This is done by answering the issues addressed in section 1.2, with following activities:

- a. To review kick detection in conventional methods.
- b. To review machine learning models tested for early kick detection.
- c. Present economic aspects related to kick consequences and the application of machine learning in minimizing undesired cost related to non-productive time and blow-outs.
- d. To develop a simulator for random generation of drilling- and kick data using probabilistic volumes and kick generation time.
- e. Testing the simulator in two case studies to show functionality and adaptability of the simulator in generation of kick data.

### **1.4 Structure of thesis**

The thesis is structured into two main parts: a literature review, and a developed kick generation simulator.

Firstly, the literature study (chapter 2) aims to provide theory of how kick occurs, are detected, and how it behaves in different conditions. Then a general overview of different machine learning models typically researched for early kick detection purposes are presented, along with some case studies for each method. Lastly the literature review represents the potential economic aspects of obtaining and mitigating kicks.

In chapter 3, the structure of the developed simulator is presented along with the physics theory of each calculated parameter in the simulations. Lastly, two case studies are conducted to present the functionality and adaptability of the simulator.

## 2. Literature review

This chapter presents the literature review of the kick phenomenon, the application of machine learning modelling for kick detection and its implication of economic aspects of reducing cost directly and indirectly to kick related issues.

### 2.1 Kick and kick detection

As the world evolves towards drilling in deeper waters the drilling operations face many challenges [5]. These are challenges such as poor borehole stability, high wellbore temperature and pressure control requirements, narrow drilling fluid density safety window, and several other challenges. The deep-water environments make the well much harder to control, thus increasing the drilling risk sharply. The two most common downhole complexities affecting drilling construction safety are influx (kick) and loss circulation [6]. These incidents can cause serious safety issues as well as unnecessary increased costs for exploration and development. In a worst-case scenario the influx can cause an out of control well blow-out, resulting in huge damages with large economic losses and potential casualties [5]. This will also result in major damages to the offshore ecological environment. Taking this into account, early kick detection is an extremely important focus area in order to prevent loss of well control [7]. As a matter of fact, analysis of the Bureau of Safety and Environmental Enforcement's incident database shows that around 50% of drilling related loss of well control incidents could have been prevented by early kick detection. This is under the definition that loss of well control includes:

- Formation fluid or other fluids flowing uncontrolled, either to an exposed formation or at the surface.
- Flow through a diverter.
- Failed procedures or surface equipment leading to an uncontrolled flow.

In order to understand how kicks occur, behaves, and are detected it is important to understand associated drilling operations, measuring equipment, impact of kick in chosen drilling fluid, and conventional detection methods. This will be described in the following sub-sections related to this sub-chapter (2.1).

### **2.1.1 Formation of kicks and influx management procedures**

A kick can be described as an unintentional flow of formation fluids into the wellbore whilst performing drilling operations [8]. The occurrence of a kick is often due to the pressure in the well falling below the formation pore pressure [7]. However, a kick will only occur under certain conditions [9]:

- The formation pressure must exceed the hydrostatic pressure in the well.
- The formation has to be porous and permeable for the fluid to flow.
- The formation must have sufficient reservoir fluids such as oil, water, or gas.

When obtaining an influx/unintentional flow of formation fluids, it may be in the form of water, oil, gas or a combination of fluids [8]. The formation fluids will be lighter in density than the drilling fluid, causing it to travel rapidly towards the top of the well, while the bulk density of the drilling fluid will be reduced, which in turn reduces the bottom hole pressure in the wellbore. If kick is not detected, the influx of formation fluids might continue for some time, further reducing the bottomhole pressure by reducing the annulus fluid density, resulting in an accelerated loss of well control.

When the drilling system indicates the sign of a kick, an alarm is raised to warn the driller. However, these are not always accurate and might raise a substantial amount of false alarms during for example transient periods [10]. This might evidently result in the driller not having confidence in the system. Consequently, the driller will perform a flow-check in order to identify if there is in fact a kick occurring. In the case of kick, standard procedure is to shut-in the well in order to prevent further influx. The formation fluid/kick is then circulated out of the well by introducing a new mud-weight (kill mud) after the well has been shut-in and the pore pressure is known to be below the bottomhole pressure [11]. The conventional methods to circulate the well is using slow circulation rates in order to minimize the annular friction loss to a level where it can be ignored, thus allowing for simpler calculations [12]. Furthermore, shut-in procedures can be divided into soft shut-in and hard shut-in. Here the hard shut-in is performed by firstly executing a flow check from the annulus with stopped main pumps for maximum 15 minutes, then close the blow-out preventor (BOP) simultaneously as the choke line remains closed if flow is observed [11]. A soft shut-in is very

similar regarding pump stop and flow check, but the choke line valve is opened before the BOP is closed, then the choke line valve is closed again. Consequently, the soft shut-in will experience less pressure peaks in the choke line system and the annulus. It is however important to shut-in the well as soon as possible when obtaining a kick in order to minimize the kick volume gained to easily circulate the kick out of the well and not lose well control, thus the hard shut-in is often preferred as it shuts down the well faster.

### **2.1.2 Obtaining influx/kick during different drilling processes**

Managed Pressure Drilling (MPD) is a widely used method on the Norwegian Continental Shelf (NCS). This is seen as an adaptive drilling process that allows the operator to accurately manage the pressure profile of the annulus throughout the well [13]. Here, the pressure is controlled by managing back pressure, circulating friction, and drilling fluid properties. MPD are shown to be successful in reducing non-productive time compared to conventional methods in wells where lost circulation, kicks, ballooning, differential sticking, and/or wellbore instability are common issues. However, this seems risky knowing that the main mechanism of the MPD method is to operate with a downhole pressure just above the pore pressure [11]. The pressure might in some cases accidentally fall below the pore pressure due to potential local variations in the pore pressure when drilling. As a result, an influx of formation fluids might be taken into the wellbore possibly causing a kick, depending on the influx volume. The MPD method does on the other hand offer a crucial benefit over the conventional methods in kick detection ability and well control at minimum kick size until it threatens well integrity [14]. Here, MPD allows to circulate the kick out of the well at normal circulation rates without a shut-in procedure by increasing the back pressure until the return flow eventually matches the flow in while keeping a constant pump rate. Another advantage provided by the MPD method is the ability to detect kicks at an earlier stage. As most drilling operations use a flow-paddle to measure the flowrate out of the well, the vast majority of MPD services use a Coriolis flowmeter [15]. The Coriolis flowmeter provides much more accurate measurements than the flow-paddle, thus allowing for earlier kick detection. This is because it is easier to interpret when the flowrate out deviates from the flowrate in for these measurements. The differences and functionality of these two types of equipment are described in 2.1.3 below.

Kicks can occur in several operating events, such as during P&A, well completion, tripping in/out, connection, or while drilling with circulation. However, this thesis is focused on kick events that can occur during drilling events, meaning the latter three events mentioned above. Now, looking at well-construction operations from a kick-detection perspective, the operations can be placed into one of the following categories with a specific purpose for the driller to apply different kick-detection methods [16]:

1. Tripping in or out.
2. Drilling or circulating.
3. Making connections.

Tripping in/out and connection events are especially challenging with regards to kick detection. This is mainly due to the transient nature (transient/unstable flow in the well) of these events as well as pumps are shut off, meaning there is no flowrate in/out which is one of the main parameters used for kick detection. However, the well can experience changes in e.g. temperature during the period when pumps are shut off which can lead to return flow due to thermal expansion, which can be difficult to distinguish from an influx event. Furthermore, pressure surges and pressure swabs might occur while tripping in or out of the well, respectively [17]. When a surge or swab is induced during tripping procedures, a sudden pressure increase or decrease will occur [18]. A surge event will generate a pressure increase, while a swab event will generate a pressure decrease. Consequently, the increased or reduced pressure might increase the bottomhole pressure above the fracture pressure, leading to a loss, or decrease the bottomhole pressure below the pore pressure, leading to an influx. To avoid this problem, drillers need to take account for these types of events before performing such operations. Additionally, the shut off pumps during both tripping in/out and connection events will result in a reduced bottomhole pressure as the frictional pressure is removed, meaning the possibility of well pressure falling below the pore pressure increases. As a matter of fact, at least 25% of all influx events occur while making connection in deep-water wells [19]. On the other hand, only a small amount of deep-water rig contractors uses kick-detection alarms during connection in order to alert the driller. This is mainly because the kick detection alarms are hard to automate efficiently for connections due to difficult interpretations of well data as a consequence of the transient flow. Because of this issue in detecting kick during connection, the development of an Influx Detection at Pumps Stop (IDAPS) software were initiated [19]. The

purpose of this system is to detect irregular flowback conditions during connection. Here, the software logs flowrate out adjusted for flowrate in, and pit gain data for the previous 4 connections establishing data of what is considered “normal” patterns during a connection. This is of course adjustable for other influencing factors, such as increased drilled depth. When making a new connection, the system will match this with the “normal” pattern flagging any abnormality found.

### **2.1.3 Kick detection parameters and measuring equipment**

Today, kick detection methods mainly include the monitoring of surface flow data (flow in and out), pit gains, pressure variations at surface and down-hole, and the outputs from models based on physics [20]. Here an increased flowrate out, higher than the injection rate, might signal that there has been an influx of formation fluid and/or gas expansion in the well [8]. Another parameter that can indicate a kick is to measure the density of the drilling fluid coming out of the well. Here a decreased density signals the presence of formation fluids mixed into the drilling fluid. However, solids from the formation might be suspended in the fluid, resulting in changes in the drilling fluid without presence of a kick. These parameters might vary without a kick being present, but a significant change might give a stronger indication that a kick is forming in the well. It is also worth to mention that different parameters are used as primary kick indicators depending on whether the current drilling operation is drilling with circulation, making a connection, or tripping in/tripping out. For instance, drilling with circulation will use increase in mud flowrate-out (compared to flowrate-in) and pit gain as primary kick indicators [16]. The primary kick indicator when making connections is pit gain and continuous mud return flow with pumps off, whereas for tripping in or out will use return flow and trip-tank gain as primary kick indicators.

Drillers go through extensive well-control training to learn how to detect a kick or any form of undesirable influx of formation fluids into the well [16]. The training consists of comparing real values with the expected, or planned, normal values for the specific well-construction operations. As a result, the kick detection can vary a lot as it depends on the driller’s individual skills and abilities to interpret these drilling parameters, model predictions and symptoms [20]. However, this is also highly dependent on the size of the kick as previously stated, whereas smaller kicks are harder to detect.

The driller is heavily dependent on measuring equipment used on-site. As a matter of fact, a large majority of the kick metrics used today are incomplete [7]. Several of the measuring methods are associated with time lags and poor accuracy, resulting in a most likely even bigger influx to the well at detection time. For instance, one of the most important parameters to track for kick detection, the flowrate out of the well, can be measured by either indirect measurements from variations in pit volume or direct measurement of the flow [21]. Here, the most typical way of direct measurements is by using a flow paddle placed in the return flowline, whereas a few operators use a Coriolis vibrating tube flowmeter. It is worth to mention that these flowrate out measurements are qualitative measurements and not “actionable” by themselves, meaning that the driller cannot positively identify a kick solely based on the flow out measurements [7]. A flow-paddle will only give gross indications of the flowrate out of the well, and has several limitations [21]. For example, the flow-paddle depends on a minimum flowrate to be able to provide readings. The flow-paddle might also get stuck in a particular position due to accumulated dirt on the hinges, making it unable to detect changes in flowrate out of the well. A Coriolis mass flowmeter can be installed as a substitute to the flow-paddle. This instrument is a modified densiometer, to be able to measure both mass flowrate and density of a substance. These measurements are seen as trustworthy and with a high accuracy.

The Coriolis flowmeter, with a known internal volume, measures the mass flowrate by measuring changes in the natural resonance frequency of the tube as the mass varies [21]. As the Coriolis flowmeter is structured with an entry bend in one direction and an exit bend in opposite direction, the Coriolis effect will cause forces on each of these bends due to the circulation of fluid in the tube. As the entry and exit is bent in opposite directions, a torque with a phase shift between the two ends of the tube will occur. The mass flowrate will be directly related to this phase shift; thus, the volumetric flowrate can be determined by combining the mass flowrate with the measured density. The instrument does, however, require a minimum pressure to make the return fluid pass through the instrument [7]. Even though this pressure is not that particularly high (around 3-5 psi), it is still larger than what a majority of open return systems can provide. It is also limited by other factors, such as the need to be constantly filled with drilling fluid, as well as the possibility of cuttings to accumulating in the tube, resulting in falsified measurements [21]. Another problem is that mud might coat the inside of the tube, resulting in systematic measurement errors. The

instrument does therefore require regularly cleaning of sensors and removal of mud on tubing walls and accumulated cuttings in the tubing.

#### **2.1.4 Oil based mud vs. Water based mud**

As one of the key concerns for safe drilling and efficient operations is to reduce the severity of possible kick situations, one has to understand the interaction between natural gas and drilling fluid at relevant conditions [22]. It is important to differentiate how a gas kick behaves in a well with oil based mud and in a well with water based mud, especially under High Pressure and High Temperature (HPHT) conditions. When using oil-based drilling fluids there is a major problem associated with detecting gas entering the borehole and dissolving into the drilling fluid [23]. If leaving an overbalanced drilled well without circulation for a longer period of time, formation gas will start to accumulate in the drilling fluid [24]. I.e. the well do not always need to be underbalanced to take an influx. The solubility of the gas is however pressure and temperature dependent [25]. An increased pressure and decreased temperature will result in an increased solubility. When looking at drilling under HPHT conditions, methane will be infinitely soluble in the oil-based mud, meaning substantial volumes of gas will dissolve in the mud [24] [26]. Another factor to consider is that a kick does not necessarily have to be methane gas in all cases. The solubility of gas into the oil-based mud will differ based on the specific gravity of the hydrocarbon gas, whereas the solubility of the gas in the oil-based mud increases as the specific gravity of the hydrocarbon gas increases [26]. In other words, the methane gas, having the lowest specific gravity of the carbon gases, will be the least soluble hydrocarbon gas in oil-based drilling fluid.

When suspecting a kick, the driller often performs a flow check. This is a process where the pumps are shut down and the rotation and drilling stopped to check if the well is static or not [27]. The time for a flow check is highly dependent on how long it takes to confirm the status of the well, where it might take a few minutes or around 30 minutes varying with well situation. However, this procedure will reduce the bottom hole pressure, potentially allowing substantial amounts of formation gas to flow into the well. This is often the case in HPHT-wells. As the gas dissolves into the oil-based drilling fluid, the kick will be harder to detect from a driller's point of view. The regular surface responses used to detect kick, such as flowrate out compared to flowrate in, and pit gain shows a significantly lower increase when using oil-based mud compared to using water-

based mud [23]. This is of course due to the gas dissolving at a much higher rate into the oil-based mud than into the water-based mud, whereas the solubility of methane in water-based mud is 8 kg/m<sup>3</sup> compared to 164 kg/m<sup>3</sup> in oil-based mud at well conditions with a temperature of 145°C and pressure of 724 bar [24]. The gas that is dissolved into the drilling fluid will travel with the drilling fluid up the annulus until reaching a certain pressure that is low enough for the gas to liberate from the fluid [25]. The kick may be undetected all the way up to this point, however, the gas liberation might happen at a height where there is little or no time to react and initiate counteracting actions. As a consequence, the sudden release will lead to a rapid volume expansion of the free gas in the riser causing drilling fluid to be pushed out by the expanding gas. In Figure 2, Rommetveit et al., shows how a 4 barrel kick dissolved into the oil-based mud will affect the pit gain when it expands as free gas after being circulated to a certain pressure where the gas is released from the mud [28]. This shows how severe an initially rather small kick will impact the well when hidden into the mud and allowed a sudden expansion at certain pressure close to the top of the well, whereas this leads to a pit gain of around 50 bbl.

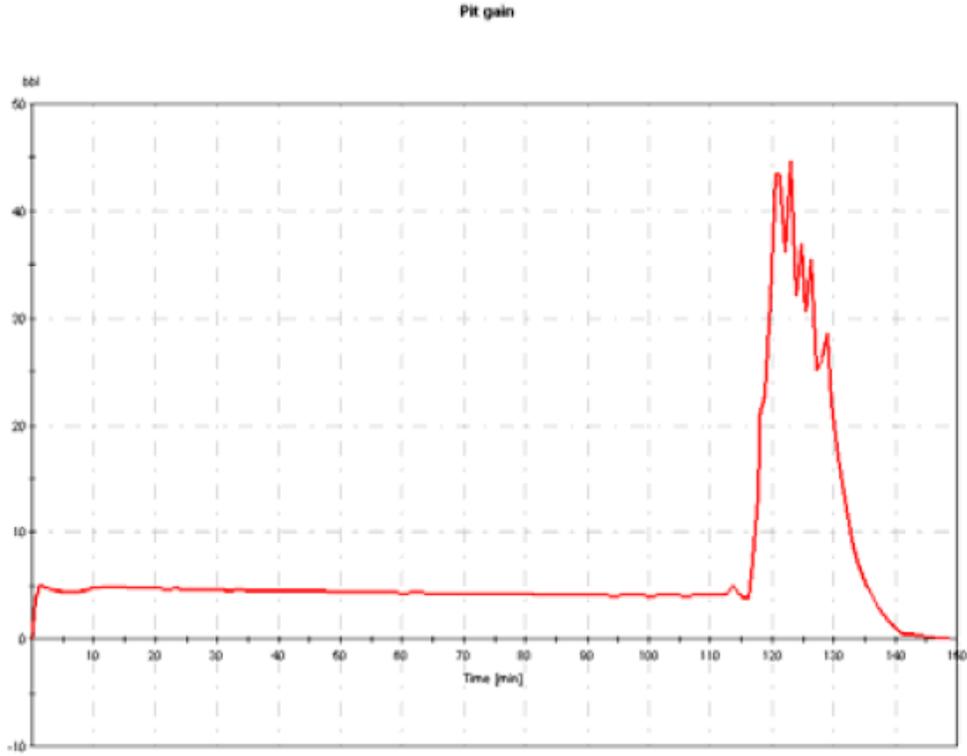


Figure 2: Kick with size of 4 bbl taken in 9 7/8" casing, with circulation rate 300 gpm [28]

This kind of impact might cause the riser to collapse, since the hydrostatic pressure in the well might be drastically reduced, leading to the liberation of dissolved gas further down the annulus as well [25]. As an additional consequence, the pressure difference between the pore pressure and the wellbore might increase, thus accelerating the influx of gas into the well and thereby increasing the severity of the situation.

Now looking at the diffusion of gas into drilling fluid from a water-based mud perspective there would still be diffusion of gas into the mud, but to a significantly smaller degree [24]. This is due to the gas having low solubility in brines. As a matter of fact, the methane gas solubility in oil-based drilling fluids will also decrease at an increasing volume of brine and emulsifiers in the drilling fluid [26]. Thus, when using water-based drilling fluids, naturally containing large volumes of brines, only a small amount of gas will dissolve into the drilling fluid, whereas the residual gas will continue up the wellbore as free gas [25]. As a result, the gas can be detected early by an immediate detectable volumetric impact. Another aspect that makes the kick easier to detect in this case is the expansion of the gas as the hydrostatic pressure decreases when the gas travels towards the top of the well. This will give a better confirmation of the presence of gas as a more noticeable volumetric impact.

## **2.2 Kick detection using Machine Learning**

The recent years has had substantial advances in sensor and data transfer technology, along with simulation tools with increasing precision and power [29]. This advancement has opened up new possibilities for kick detection, possibly integrated with and enhanced by machine learning methods. Machine learning has several possible application areas within the oil and gas industry; however, the focus in this thesis lies on early kick detection for better well control. Here the machine learning algorithms can be used to for example reduce number of false alarms during drilling operations [3]. These alarms are originally set with defined ranges for acceptable changes in flow rates and mud levels, such that they can alarm the crew when the levels move outside this range. A general problem with this is that the flow rates and mud levels vary a lot during normal drilling operations, meaning even a correctly calibrated and functioning system might generate several false alarms for lost circulation or influx taken. As a consequence, drillers might become insensitive to alarms, reduce the alarm sensitivity, or completely shut off the alarms for periods.

The current gas kick monitoring methods are divided into three parts: seawater section monitoring, wellhead section monitoring, and downhole section monitoring [5]. However, there are some limitations and problems by the conventional detection methods, including: several influencing factors, only experienced technicians are suited for detection in many cases, poor real-time measurements, hard to evaluate degree of severity, late discovery time or poor accuracy in bottom hole pressure measurements. The down hole section monitoring in particular is associated with higher costs as well as larger risk of measurement instrument failure [6]. Artificial intelligence technology and information technology has been researched for application in influx and loss detection, where they provide substantially more accurate detection results. However, this monitoring method is considerably harder to model as well as promotion is harder.

### 2.2.1 Understanding machine learning

To understand how machine learning algorithms can improve influx and loss detection methods, it is important to understand the basics of how these kind of algorithms operate and improve themselves. Machine learning is a segment within artificial intelligence, meaning that these kind of algorithms can make their own choices based on previous training. That being said, a machine learning algorithm will always focus on improving its self-performance through calculations [6].

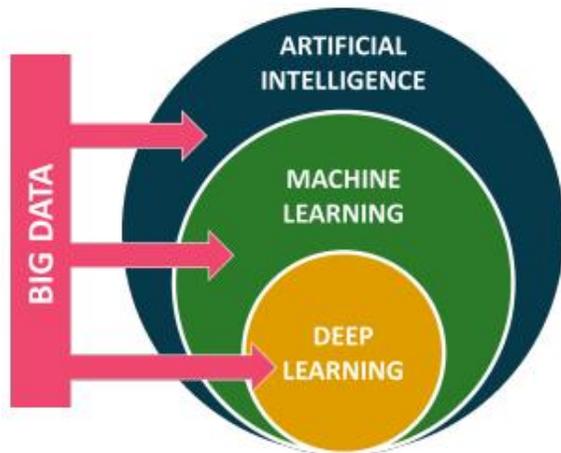


Figure 3: How artificial intelligence, machine learning, deep learning and big data are related [30]

Figure 3 shows how artificial intelligence, machine learning, deep learning and big data all are connected [30]. This clearly depicts the relationship between the different types of algorithms, where deep learning is a subset of machine learning and machine learning is a subset of artificial intelligence. Big data is large and complex data sets that requires advanced algorithms such as in machine learning and deep learning to analyze them. Deep learning mainly focuses on using multi-layered neural networks to

interpret and analyze data. Artificial intelligence and machine learning are often used interchangeably, but they are however not necessarily the same [31]. To this extent, artificial

intelligence is associated with a computer's or a machine's ability to make decisions and behave like humans. This means that artificial intelligence does not necessarily depend on learning to be able to perform certain actions, whereas it simply can be triggered by for example a sensor to autonomously perform a given action. Now looking at machine learning, these types of algorithms rely on training and learning in order to identify the appropriate decisions to make. Machine learning algorithms are able to receive several unique inputs to learn from, without the need to know their inter-relationships prior to the calculations [30]. The algorithms will then make predictions and thereby be able to forecast future events and performance.

As previously stated, the machine learning algorithms depend on training and learning in order to make intelligent decisions. That being said, the learning processes can be divided into unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning. Unsupervised learning is the process of learning through feeding unspecified data (no defined input or output data) to the system [31]. The system will then try to identify any plausible relationship between the data in the data set. This method is often used to detect hidden or underlying patterns in the data set, that is hard to detect by the interpreter [30]. Supervised learning on the other hand rely on some defined input and corresponding output parameters in order for the system to build relationships between the inputs and outputs [31]. This means that the system should be trained to predict certain outputs for input data that were not originally used in the training data sets. Furthermore, the semi-supervised learning is a combination of supervised and unsupervised learning [32]. This learning method is used for data sets with an uneven combination of unlabeled and labeled data, where the unlabeled data is represented to a significantly larger extent [31]. The last learning model, the reinforcement learning, learns through experience. In other words, the system interacts with a virtual environment through several simulations in order to learn based on several consequences of its actions. The system will continuously look to maximize their reward for every iteration by learning from previous iterations and use this experience to further improve the processes [30]. The four learning methods will have different levels of robustness in terms of learning accuracy [31]. However, other factors such as ease of interpreting results, speed of implementation, nature of supervised machine learning task, amount of input and output data, and complexity play a major part when considering which type of learning model to apply. It is therefore extremely important to understand the data set and important factors to be used in order to apply the best learning model for that specific application.

### 2.2.2 Training the system and pre-processing data

As previously stated, the machine learning models require a large amount of data to learn and train to optimize processes and make intelligent decisions. The Electronic Drilling Recorder (EDR) on drilling rigs can provide enough data for the machine learning model to detect anomalies in the wells behavior [3]. By learning the relationship between surface parameters such as block position, pump rate (SPM), mud levels in every tank, and flow out of well (using either a standard flow paddle or Coriolis flowmeter as described in part 2.1.3), the machine learning model can detect any abnormal activity in the circulation system. These abnormalities are found as the machine learning model predicts values and compares them to the measured real values. Any value outside of a specific range will then be tagged as an abnormality and alert the driller. It is worth to mention that several of these parameters are only effective for these purposes during drilling with circulation, as the pumps are turned off during for example a connection. During a connection event, the flow out and pit-volume signatures will be transient by nature due to pressure and temperature changes in the well as previously stated [19]. As a result, this is one of the most challenging processes to automate kick detection for as the real parameters will give several abnormal measurements compared to the predicted parameters of the machine learning model.

Data gathering during drilling related operations will generate a lot of noise in the data sets. As a matter of fact, machine learning algorithms cannot differentiate this noise from informative data in the initial data [33]. Consequently, preprocessing of the initial data is crucial for these types of algorithms. A critical element for the preprocessing is to understand and identify different events in the data, along with several events in the well. Here, a certain change in parameters over time will define a well event. Another important aspect in preprocessing and treatment of data is to understand the quality of the data. To this aspect, most research agrees with following six parameters defined by the DAMA UK Working Group: Completeness, Uniqueness, Timeliness, Validity, Accuracy and Consistency [34]. Each of these parameters can be described as:

- **Completeness:** How much of the original data is available, seen up to a potential of “100% complete” data set. The aim here is to get as close to 100% as possible as this represents that all data is present in the data set.
- **Uniqueness:** The same thing should not be recorded more than one time, to ensure that duplicate values do not disturb the complete data set.

- **Timeliness:** The data should be as up to date as possible to represent reality to a higher extent.
- **Validity:** Recorded data should be within specified ranges, meaning that data recorded outside of this range will be seen as invalid.
- **Accuracy:** To what extent the data accurately describes the real event or object being described.
- **Consistency:** Recorded data across two or more datasets should be as close to equally as possible represent the object being recorded.

Well logging can be complex, and the most common problems related to the well log deliverables are inaccuracy, incompleteness, and inconsistency [35]. The common issues for each of these parameters can be described as following:

- **Inaccuracy:** Loss of accuracy can be due to incorrect sampling rate, mixing of wellbore names, or poor labeling (meaning it is hard to differentiate between the data sets and tag them to a specific operation).
- **Incompleteness:** Lack of completeness in the data set can be a result of missing sections or files, missing curves, segmented curve representation, or missing essential records.
- **Inconsistency:** The data sets can become inconsistent due to not all printed curves being available as digital information, printed curves or headers might differentiate from their representative digital information, or updated digital information might differ from previous versions of itself.

Different sensor equipment and measuring intervals used in logging will result in a mixture of data with several different dimensions and formats. The raw data can often be “dirty”, meaning there is a need to quality control the raw data and perform data filtering actions [36]. If data quality is not sufficiently controlled, the machine learning model might homogenize and neglect relevant information. Another aspect that might “confuse” the machine learning model is the use of irrelevant data. This might result in loss of accuracy and efficiency in the predictive models. To avoid these problems in the data set, the best way would be to avoid them from occurring at all [30]. However, dealing with these types of data or missing data is not always possible due to all equipment not working perfectly all the time, borehole environment is not sound, data quality is not excellent, or data processing is not accurate. If there is missing data (often tagged as -999, -

9999, etc.) in the data set after all, the regular way of dealing with these values is to exclude them from the data set to avoid the machine learning model treating them as real values. The data cleansing process is the most time consuming process related to building machine learning models, normally consuming more than half of the development time [37]. As a matter of fact, this also requires substantial IT knowledge in order to create components to identify defected data, as well as great mathematical and statistical skills in order to find the optimal way to fix/clean these data.

Generation of training and testing data will also be challenging with respect to the vast variation in noise possibilities. Many of the tested machine learning models today are trained and tested using clean synthetic data, meaning data without noise. Geekiyanage et al. (2019) presented a study on a simulated MPD system set up in a laboratory rig to generate data sets for training machine learning models in kick detection [20]. Here, sensors were used with a defined amount of real-like noise. The mean and standard deviation of the noise in the sensor data is presented in Table 1 below. Although this is a controlled laboratory experiment the standard deviation is pretty huge as seen in the table below. They still found comprehensive data pre-processing necessary to filter out outliers and reduce noise even in this controlled laboratory environment. However, they noticed that filters might create a delay for the kick detection time, as early kick indicating symptoms was removed in some instances.

Sensor	Description	Recorded Data Range (min-max)	Mean ( $\mu$ )	Standard deviation ( $\sigma$ )	Units
RP 401	Pump pressure	0.07-4.05	2.14	0.94	Bar
RP 403	Stand pipe pressure (SPP)	(-0.42)-2.64	1.13	0.57	Bar
RP 404	Bottom-hole pressure (BHP)	0.01-2.58	1.24	0.57	Bar
RP 406	Upstream choke pressure	(-0.14)-0.78	0.28	0.15	Bar
RP 407	Downstream choke pressure	(-0.02)-0.51	0.15	0.08	Bar
RF 401	Flow out rate (Coriolis)	(-0.26)-161.83	79.58	40.02	l/min
JP 401	Pump flow rate	(-0.61)-125.40	86.80	34.20	l/min

Table 1: Sensor description with corresponding noise mean and standard deviation [20]

When a machine learning model is developed and tested, a confusion matrix is often applied to test the overall accuracy of the model [20]. This is most of the times tested by using following formula:

$$A = \frac{TN+TP}{TN+TP+FN+FP} \quad (1)$$

Where TN represents the True Negative, TP the True Positive, FN the False Negative, and FP the False Positive. These values denote whether the model made the actual negative and positive predictions, i.e. the true negative and positive, or if the model made false predictions, i.e. the false negative and positive. By using equation (1) one will find the overall percentage accuracy of the model predictions, where closer to 100% means more accurate predictions.

### **2.2.3 Machine learning models researched for kick detection today**

The substantial improvement of sensors and data gathering equipment in the oil and gas industry has naturally made Big Data an integral part of the industry [38]. This has opened up for the use of machine learning models to interpret raw data. However, this is still in the early stages for several applications in the industry. Several research papers present work on the development of different machine learning models used for kick detection today. Nevertheless, most of these are used with filtered data sets, meaning substantial amounts of noise is removed. This will give an accuracy that is far greater than what could be obtained with data containing noise in a real-working environment. Another issue is that these models are trained towards very specific data sets, possibly making them unfit for applications in other fields.

Machine learning models can use two approaches for prediction of kicks: a data mining approach or a physics approach [39]. Both of these methods use historical data to learn and use this to predict future kick occurrences. As previously stated, machine learning algorithms highly depend on sufficient and qualitative data. Consequently, it is very important to choose the correct input parameters for the machine learning models in order to provide good enough data. Data are either collected at surface or at the subsurface, whereas the surface data provides a much higher data collection frequency than the subsurface data [39]. Another aspect to consider is that the surface data is always available, compared to the subsurface data that needs a continuous mud column in order to transfer data to the surface. As a result, the surface data is much more favorable to use for machine learning applications and kick detection. Typical input parameters here are block position,

rate of penetration, weight on bit, hook load, pump pressure, torque, flowrate in, flowrate out and pit gain [40]. However, the parameters used as input data for the machine learning model might vary depending on either the model or the quality of the parameter-data at that particular location. Due to the larger difference in available and quality of data from operator to operator several machine learning models have been developed and tested in experiments and described in papers. Here, some of the most frequent used machine learning models include Decision Trees, Random Forest, K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Support Vector Machine (SVM). They all have in common that they intend to predict a value/data based on some input parameters, whereas for kick detection this can be used to detect abnormal values that differ from the predicted values of the machine learning model, signaling that a kick is about to occur or is occurring. It is important to understand the basics of how these models work in order to understand the limitations and advantages for each of these models when looking at kick detection applications. A brief introduction to each model will therefore be given before presenting experiments on applicability in different kick detection scenarios in subsection 2.2.4 - 2.2.7. What is important to study is the prediction accuracy and how it fits the intended data set, as the models perform differently depending on available training data. This is a key factor when considering application areas for the machine learning model.

## 2.2.4 Decision Tree and Random Forests

**Decision trees** provide an easy overview of input data and is therefore used extensively [38]. Figure

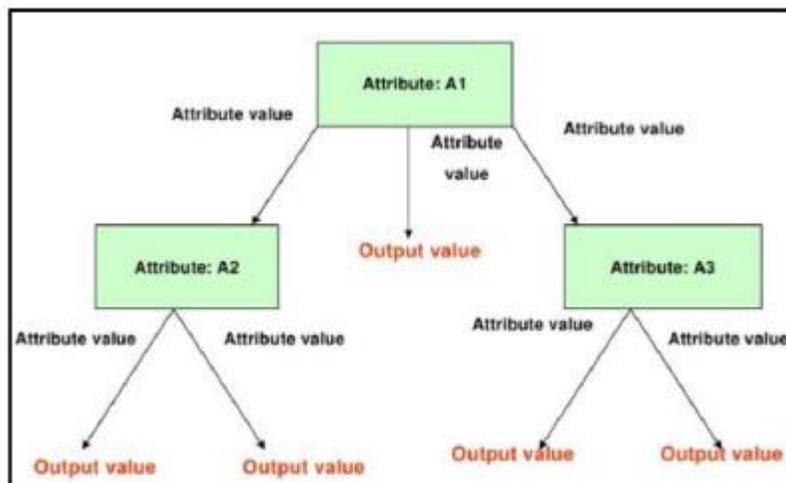


Figure 4: Illustration of a simple Decision Tree [38]

4 shows a simple overview of how a typical decision tree is built. This model takes the input data in the uppermost node, also called the root node marked as Attribute: A1 in the figure. Certain criteria in the model will then split this root node into multiple child nodes (Attribute: A2 and Attribute: A3) with a

goal to keep splitting the data until reaching a specific class. The splitting choices are made by if-else decision making [39]. Consequently, the predicted results will be based on a reasonable and justifiable logic, as the choices are made solely on if-else analysis aiming to give the best attribute with a major impact on the output. However, the decision tree might struggle when obtaining a large number of attributes as well as needing a substantial amount of records in order to obtain a great quality result. It is also worth to mention that this model can be volatile when changes are made in the training data [38]. Nevertheless, this problem can be avoided by using the **Random Forests** model, meaning the use of multiple decision trees growing in parallel based on different sample data generated from the same data set. These models are associated with high prediction accuracy, as well as great tolerance to noise and outliers (data points outside the ordinary standard deviation) [6]. They are also associated as easy to train, based on the low amount of required hyper-parameters (parameters other nodes in the model are derived from) as well as functioning well with a small amount of data [41]. The overall principle of the model is to generate a defined number of decision trees, then voting for each of the generated outputs to determine the final output. Shi et al. (2019) presented a study on the performance of random forests and SVM with well data from four wells in Tarimu oil field [6]. The two models were tested using a data set of 18,720 data points consisting of surface parameter measurements such as flowrate in and out, and pit gain, individually tagged as influx, loss or normal status. Here, both models performed to a high degree of accuracy where the SVM method provided an accuracy of 93.72% and the random forests provided an accuracy of 92.23%. Although the accuracies are pretty similar, the SVM model was in this case deemed the best based on less false kick predictions than the random forests model.

### **2.2.5 K-Nearest Neighbors**

The **K-Nearest Neighbors** model combines regression and classification algorithms based on calculating the distance between data points [39]. However, the model scans the whole data set each time it makes a prediction, making it slow with large data sets. Another problem is that it often overfits to the training data, meaning the performance will drop drastically when new data is introduced because the model is perfectly formed to match the training data. Nonetheless, the model makes prediction with a very high accuracy for each prediction. As a matter of fact, Alouhali et al. (2018) presented a study comparing the performance of kick detection between decision tree,

KNN, SMO (SMV based model), ANN and Bayesian Network, where the KNN proved the highest prediction accuracy [39]. The goal set for the machine learning model was to predict kicks using surface parameters. Data was collected from wells with kick incidents with an initial data set volume of over one million instances of drilling measurements. These instances were then cleaned and labeled to tag kick incidents with “kick” and non-kick incidents with “no-kick”, reducing the training data set to around 122,000 instances. The KNN showed a very high accuracy of 99.2% followed by the decision tree and the ANN models, also with an accuracy over the 90<sup>th</sup> percentile. In another study on KNN used for early kick detection using a series of 108 field experiments, Yin et al. (2020) showed a prediction accuracy of 80.9% for the KNN model [42]. Here, the input parameters used were surface parameters (ROP, SPP, WOB, flowrate in and out, etc.). The model made some false predictions, where it predicted a kick when no kick was occurring for several instances (around 20). Although the KNN model shows great results in some experiments, it might be too time and capacity consuming for real operating environments and early kick detection. This is because, as previously stated, the model has to run through the complete data set each time a prediction is made. Data produced in an operating environment can also contain a lot of noise as well as the data set is extremely large, which will slow down the system further and potentially reduce accuracy.

## 2.2.6 Artificial Neural Network

The **Artificial Neural Network** is based on the same principles as biological neural networks, in a simplified manner [43]. This model is able to make associations, transformations, and mapping amongst data. ANNs can handle complex functions and non-linear relations in the data set, making it quite suitable for dealing with data sets with missing or incomplete data and for problems where mathematical modeling fails. Additionally, the ANN does not require any system explanation

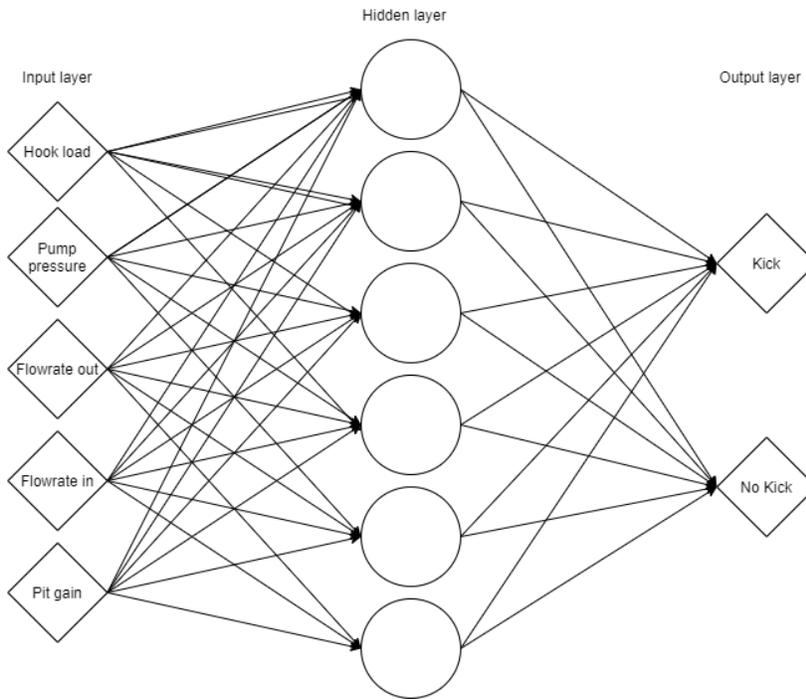


Figure 5: Example of ANN used for kick detection

through physical phenomenon to make relationships between output and input parameters. The ANN can be divided into three parts, the input layer, the hidden layer, and the output layer as shown in Figure 5. Here, the hidden layer can be seen as a group of independent neurons [38]. These do not have any direct connection to anything else than the input and output layers [39]. Their role is to make computations by received data from the input

layer, and provide output data to the “outside world”. The neural network will learn by following steps [38]:

1. Identifying key parameters and defining problem.
2. Definition of network structure for solving the problem, i.e. input data format, network structure, and learning algorithms.
3. Generation of training data set as input data with known output values.
4. Input and output data are put into the model for parametric identification.
5. The model will learn through reiterating towards a specific order of maximum total error, or steady state is achieved.

Muojeke et al. (2020) presented a research work on the application of ANN for kick detection, using data sets generated from a Small Scale Drilling Simulator (SDS) and a Large-scale Drilling

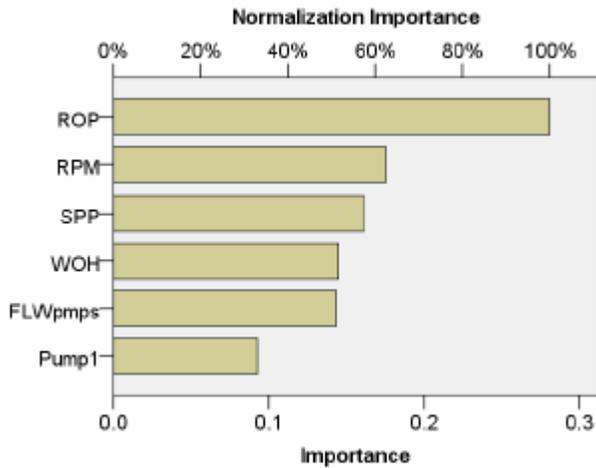


Figure 6: Importance of different input parameters [5]

Simulator (LDS) [4]. The input parameters used for kick detection was downhole pressure, mud flow-out rate, density, and conductivity. Using a supervised learning model, the ANN showed a 100% prediction accuracy on both data sets. It is however worth to mention that the data sets contain a limited amount of data noise compared to real world data, and the model used is only a simplified replica of a real-life situation. Here, only a few

input parameters were used, whereas ROP, WOB, and torque were kept constant to simplify the experiment. Now looking at the experiment on kick detection performance among decision tree, KNN, SMO, ANN, and Bayesian Network provided by Alouhali et al. (2018) as presented earlier, it is clearly seen that the ANN has a high accuracy even for large data sets (over 122,000 real well data points) [39]. The accuracy of the ANN was measured to 98.8% for this experiment, where it was beaten by the KNN and the decision tree model by a small margin. Although these are good results, the model might need some adjustments to be applied in a real working environment as the training and testing sets are filtered data to a large extent, meaning a lot of noise is removed. This is much harder to do in a working environment in real time, thus risking poor predictions by the model used. It is also important to study which parameters are most effective for early predictions of kick, as this should identify the kick at a reasonably earlier time than the driller. Jin Yang et al. (2019) showed in a study using ANN on a data set containing thousands of real well data the importance of the different input parameters [5]. The prediction model consisted of five neurons in one hidden layer, and showed an accuracy of 96.2% using the testing data set. Studying Figure 6 it is clearly seen that the ROP is the most important parameter to use for early predictions. This coincides with the theory that the ROP has a sudden increase before kick is obtained. It is worth to mention that this might be due to difference in the rock formation, however, the model should be able to not make false predictions when combining the prediction with other input parameters.

## 2.2.7 Support Vector Machine

The **Support Vector Machine** model is generally applied in regression problems demanding a high degree of accuracy [38]. This model is advantageous for linear data sets containing substantial volumes of noise and non-linear data sets. However, a substantial amount of noise in the data set

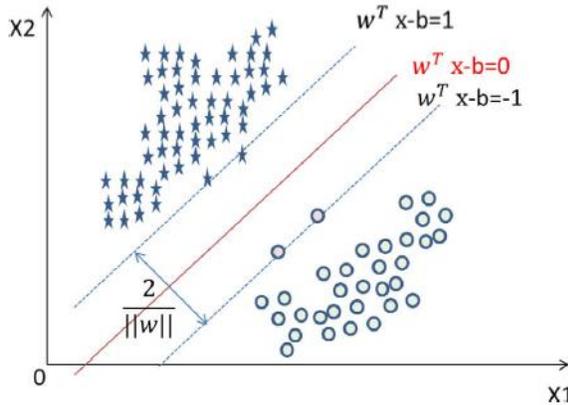


Figure 7: Visualization of a SVM model [6]

might also result in high processing time. Advantages such as global optimization, strong adaptability, rigorous theory, good generalization performance and high training efficiency are strong drivers for its advantageous applicability in these kinds of data sets [6]. The classification of data is done by the SVM finding the optimal hyperplane (red line in Figure 7), separating all data points belonging to a class from data points

belonging to another class [44]. The optimal hyperplane is defined as the one with the largest width between the classes. This is used to classify the data, whereas data points above the hyperplane are categorized with a +1 value and the data points below the hyperplane is categorized with a -1 value [38]. In the case of non-linear data sets, the SVM transforms the non-linear hyperplane into a linear hyperplane existing in higher-dimensional space by using Kernel functions. This classifies the non-linear boundaries, whereas the Kernel function will format the data linearly into a higher dimension. Choosing the correct Kernel function is very important for securing a good performance of the SVM model [45]. In fact, the chosen Kernel function will define the SVM model. A study with a data set of 6,976 well sample data, in which 199 samples were lost circulation, used for circulation loss prediction using SVM presented by Li et al. (2018) showed a prediction accuracy of over 99% for normal predictions [44]. However, this model only classified 55% of the 199 lost circulation samples correctly. Furthermore, in a study presented by Shi et al. (2019) on using SVM for early influx and loss detection using well data (surface parameters) from two wells in Tarimu oil field, the SVM model showed a superior prediction accuracy of 93.72% compared to the random forest model of 92.23% [6]. The SVM managed to detect loss accurately at an early stage, whereas it was detected 20 minutes earlier compared to the traditional pit volume measurement method. However, the SVM made some false detections at the start of loss as also seen in the study provided by Li et al. (2018). Additionally, the influx detection was falsely detected at the early stage of the

influx, whereas the SVM predicted accurately during normal conditions and throughout the event duration.

### **2.2.8 Machine learning models for kick detection - comparison**

The vast difference in data provided during drilling with different equipment, sensors and in different environments makes it hard to make a generic model for early kick detections. Surface parameters does seem to be the optimal choice for input parameters to the machine learning models, as this data is provided constantly at real-time compared to downhole measurements with delay. The downhole measuring equipment is also associated with a reasonably higher cost, making the use of surface parameters even more attractive for the operators. As there is a large difference from data set to data set on different locations, the different models may perform differently from case to case. Here, one model might outperform the others in one case, then be outperformed by the others in another case. The time made for predictions and computing capacity demanded is also needed to be considered, as some models require substantially more than others. Additionally, available input parameters and size of data set should be considered when choosing machine learning model. A comparison of the models is shown in Table 2 below, where all the models use some kind of a surface parameter as inputs for the machine learning algorithms. This shows that almost all the models are capable of a reasonably high prediction accuracy, where the KNN model was the only one with a prediction accuracy under 80.9%. The accuracy of the KNN is however above the 99<sup>th</sup> percentile in another study with a different data set. This reduced accuracy is a result of the model making several false predictions, whereas the model had a high prediction accuracy of the actual events (>90%) [42]. The reason for the false prediction and thereby the reduced accuracy is most likely due to poor data quality in the input data, whereas the data has too much noise disturbing the model. Consequently, the available data at the targeted location and the possibility of real-time pre-processing of these data is of great significance when choosing which model to apply. The model that performs best overall, considering all the factors, seems to be the ANN based on its robustness and ability to handle large data set with a reasonably large amount of noise without affecting the prediction accuracy.

<b>Model used</b>	<b>Parameters</b>	<b>Size of data set</b>	<b>Accuracy</b>
Decision Tree/Random Forests [6]	Flowrate in/out and pit gain (surface parameters)	18,720 data points (real well data)	92.23%
KNN [39]	Surface parameters	122,000 data points (real well data)	99.2%
KNN [42]	ROP, SPP, WOB, flowrate in/out (surface parameters)	Data from 108 field experiments	80.9%
ANN [4]	Downhole pressure, mud flow-out rate, density, and thermal conductivity	Simulated data (limited amount of noise)	100%
ANN [39]	Surface parameters	122,000 data points (real well data)	98.8%
ANN [5]	ROP, RPM, SPP, weight of hook, and pump flowrate.	Thousands of real well data	96.2%
SVM [44]	Surface parameters	6,976 well sample data	99%
SVM [6]	Flowrate in/out and pit gain (surface parameters)	18,720 data points (real well data)	93.72%

*Table 2: Comparison of machine learning models used for kick detection*

## **2.3 Potential economic aspects of an early kick detection**

Reduction of exploitation costs is a challenge the oil and gas industry continuously strive to solve in order to maintain an economic feasible oil production [10]. Here, influx and loss detection are of great importance due to keeping drilling safety and integrity. The systems used today often miss several events and trigger a substantial number of false alarms, especially throughout transient periods. As a result, the industry is in a dire need of huge improvements within kick detection systems to meet the demands of low production costs in an increasingly challenging market due to increased production difficulty. The economic aspects are therefore of great relevance for an operator to decide whether to implement this kind of machine learning technology or not. Here, the ability to prevent costly influxes and losses during drilling operations must outweigh the cost of implementing the software/technology. It is hard to give specific numbers on how the economic aspects around preventing a kick is, because this is strictly preventive costs. However, some companies make monthly reports based on alarms related to certain incidents during the drilling processes, where an estimation of prevented costs are summed up to present how much costs they saved by using preventive technology.

### **2.3.1 Cost reduction through digitalization and machine learning**

OG21, a project executed by a group consisting of a collaboration among universities, suppliers, oil companies, research institutes, public bodies and regulators, were initiated to define a national strategy for the petroleum technology in Norway [46]. Based on their experience and findings they found following three key elements to make the future of the Norwegian petroleum sector attractive and competitive:

- Lower costs to be more vigorous towards possible lower prices than historically.
- Match industry targets and stakeholder's expectations by having world class environmental performance and safety.
- Attract investments by having shorter lead-times.

The competitiveness of the NCS is dependent on the development of a comprehensive set of technologies, thereby digital technologies as an integral part [46]. This includes the use of artificial intelligence such as machine learning to improve existing processes and open up for new

possibilities. Ellingsen et al. (2020) presented in a DNV GL report for the OG21 project that application of machine learning in drilling, completions and intervention is one of the main opportunities for cost reduction on the NCS [47]. Here, the key opportunities were defined as parameter optimization and autonomous drilling for efficiency increase in drilling operations, as well as reduced non-productive time via anomaly and early incident detection using machine learning algorithms. This technology is estimated to have a probability of success equal to approximately 80%, with a potential cost reduction of NOK 3-4 billion per year. Now comparing this to the total operating costs of NOK 60 billion per year on the NCS, this would represent a significant reduction of around 5-7%. Considering the reported recommendations and potentials together with the increasing awareness of environmental safety, it is not unlikely that machine learning for early detection of leakage, loss circulation, and kick will become of great significance and maybe in some instances become mandatory on the NCS.

Another important factor to consider is the possibility of substituting expensive downhole measuring equipment with a machine learning system for safety reasons, such as early influx and loss detection. It is therefore important that surface parameters used as input provide sufficient results in the detection process. Another advantage of using the machine learning tools in terms of cost is the plug and play capability of the system. The companies does not need to invest anything in addition to get the models up and running, since the solution would be cloud based and in agnostic to Wellsite Information Transfer Standard Markup Language (WITSML) server, meaning the model can read WITSML data from different WITSML servers and send the output to whatever WITSML viewer the company uses [48]. Figure 8 shows how real-time data from the rig is pre-processed (with the same methods as for the training set) and fed to the machine learning models. The data is then sent for post-processing of data (rescaling) and presented in the WITSML viewer. As a result, the companies can use the infrastructure already in place on the rig-site.

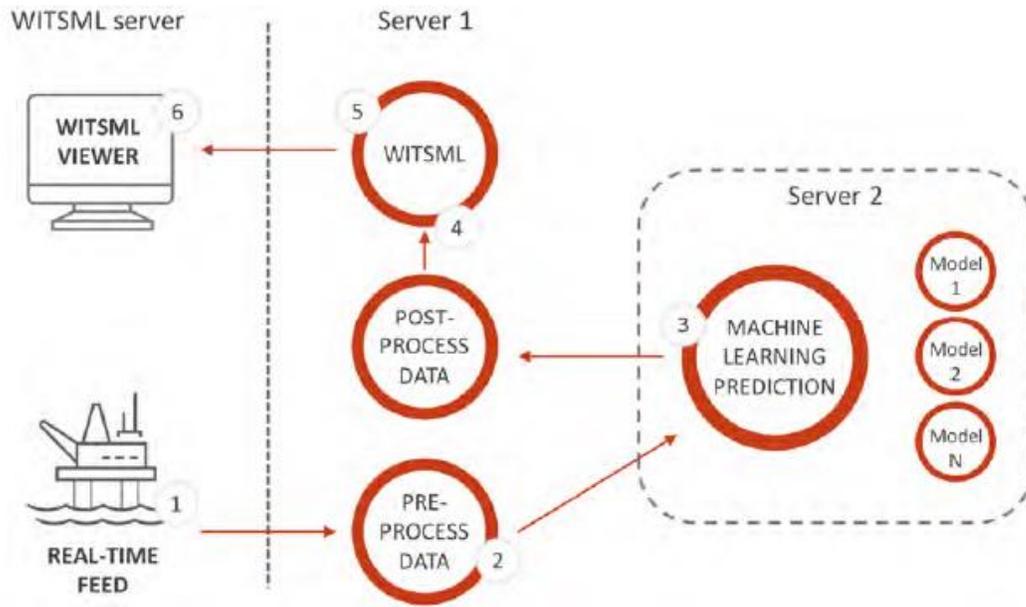


Figure 8: Structure of plug and play machine learning system [48]

### 2.3.2 Cost reduction through reduced non-productive time and early prediction

Several of the studied machine learning models researched for early kick detection are also tested for early detection of lost circulation. Lost circulation is a major expense for the petroleum industry, whereas it has been the major issue for non-productive time during well construction (12% of total non-productive time cases) over the last 60 years [49]. As a matter of fact, around 2 billion dollars are spent every year in order to treat this problem. To put this into perspective; the average cost of a drilling fluid is around 10% of the total cost of the well [50]. As a result, the mud loss might represent a significant portion of the well costs. The mud loss can be categorized into three categories; partial loss, severe loss, and complete loss, where the volumes are lost at a rate of 1-10 m<sup>3</sup>/h, >15 m<sup>3</sup>/h, and no return respectively. Applying a predictive model here, to reduce the mud loss or avoid the mud loss overall would therefore be a great method to avoid unnecessary expenses during drilling operations.

Major cost saving can be achieved through reduction of non-productive time as this makes up approximately 15-25% of the total well costs [47]. As a matter of fact, Ellingsen et al. suggests that a 50% reduction in non-productive time can be achieved using machine learning for anomaly detection such as stuck pipe, kick, and losses through monitoring real-time data and performing trend analysis. These monitoring techniques are used to some degree in today's industry, but

machine learning is not applied for its full potential here, as it only tags occurring events by sounding an alarm to the driller. The saved costs of preventing a kick will of course vary vastly from well to well and operator to operator. However, the financial benefits of early kick detection for an operator by preventing a kick can reportedly save up to \$5.7 M/day [20]. This will only further strengthen the relevance of implementing machine learning technology for well control reasons.

When applying machine learning methods for anomaly detection, the system will generate warnings through alarms to communicate to the driller that for example a kick is about to occur. These algorithms can also be trained to detect the severity of the coming issue, thus raise flags divided into a set of colors where each color represent a level of severity (ex. size of kick) for the potentially incoming issue. Each of these colors can then be associated with a cost. As a result, the prevented costs can be noted and summarized at the end of each month to provide an overview of how much the preventive tools and actions are saving the company. It is worth to mention that this would only be realistically estimated if no false kick indications are provided. Some models might tend to develop false predictions, thus raising false alarms. Consequently, the assumed prevented costs would grow unrealistically. Assuming an MPD operated well, an example of a green flag can be a signal that a small kick that is easy to circulate out of the well is incoming, a blue flag can signal that a larger kick needing a shut-in well and kill procedure is incoming, and an orange flag might indicate a late detected kick with a larger risk for developing into a blow-out. Figure 9 shows an example of how this can be summarized over a given period (for example one month) to estimate the total savings by using preventive tools and performing preventive actions to avoid the unwanted incident. This particular example shows a cost saving of 340,000 USD by not obtaining the kick in these scenarios. Note that all the numbers here are assumptions made just to visualize how the concept works. The real-life numbers will vary to a large extent for each given well based on production rate, mud costs, definition of flags and associated costs, etc. As the machine learning models can have occurrences of false predictions, the flags can be confirmed by the driller studying the data of the tagged event to verify that the warning flag is correctly assumed.

Number of flags raised to the rig in the period	Estimated cost saving (for this type of flag)	
15	10,000	150,000 (USD) Green flag
3	30,000	90,000 (USD) Blue flag
1	100,000	100,000 (USD) Orange flag
<b>19</b>		<b>340,000 (USD) Total</b>

Figure 9: Cost savings by using preventive machine learning tools for early kick detection

### 2.3.3 Costs of the Deepwater Horizon blow-out – an extreme case scenario

In extreme cases, a blowout resulting from a kick can lead to casualties and a major negative environmental impacts. A good example of such a case is the Deepwater Horizon incident in 2010 in the Gulf of Mexico. Here the driller lost control of the well, resulting in a blowout leading to the rig “Deepwater Horizon” exploding and catching on fire [51]. Consequently, 11 people lost their lives and another 17 were injured by the incident. The fire continued until the rig sank 36 hours later, and the hydrocarbons kept flowing out of the reservoir through the borehole for another 87 days. British Petroleum (BP) firstly estimated the flow rate of the oil spill to equal 1,000 barrels per day, whereas a pushback from the U.S. government later increased this estimate to be 5,000 barrels per day [52]. However, the current U.S. president at the time, Barack Obama, was not convinced by these estimates and ordered the secretary of Energy to engage a team of scientists in order to calculate a more realistic estimate of the discharge rate from the well. Through comprehensive studies, using several machine learning models to estimate the oil discharge rate, the group calculated an initial flow rate equal to 62,000 barrels per day. Ultimately, this led to an estimated oil spill of 4.9 billion barrels after the 87 days long period. Although BP of course challenged these estimates, the actual measurements of the oil discharge rate was calculated to be around 57,400 barrels per day. This resulted in becoming one of the worst environmental accidents throughout the history of the oil and gas industry.

BP were held responsible for all the costs related to cleaning of the oil spill, where BP in the annual report for 2016 documented the final costs before taxation to be \$62.59 billion [53]. However, there are hidden costs that needs to be taken into consideration for a more realistic picture of the actual loss. Hidden costs can arise from reputation damaged by the disaster, lost revenue, and non-earned profits. Lee et al. (2018) presented a study on the hidden costs carried by BP resulting from the

Deepwater Horizon blow-out, where The hidden costs were estimated using cost of global settlements, liabilities, Securities and Exchange Commission settlements, legal fees, and loss in market capitalization due to the incident [53]. Here, the loss in market capitalization (loss in stock value) was the definite largest hidden costs. One day before the blow-out, BP had an estimated stock market value of \$187.46 billion. This dropped by about one-third just a few months after the accident, equaling a loss of \$61.41 billion in capitalization after market-adjustments. By including all the hidden costs, the ultimate cost of the blow-out was estimated to be \$144.89 billion.

### 3. Random Influx Simulator

Obtaining proper kick data is a scarce resource, which is a key element when aiming to develop an early kick prediction machine learning model. When looking at drilling data, the actual kick and early indicators might be hard to determine, as well as noise might additionally increase this difficulty. Data samples from a real well will also be specific towards that particular well, meaning the machine learning model might experience difficulties adjusting to other wells. Consequently, a simulation model adjusted for acceptable noise, as well as the ability to adjust calculations towards any defined well characteristics would be of great value in generating both training and testing data for machine learning models. The focus of this chapter is to present the physics, mathematics, and behavior of the developed random influx simulator. Here, the main objective is to drill a defined number of stands, whereas there is a probability of randomly generating an influx during drilling the last stand. The size of the influx will be rather small in this simulator, as the objective is to generate data for a machine learning model that will detect kicks that is not easily seen by the driller, but it can however be set to a higher value. The size will be determined by randomly drawing a volume from a uniform distribution with a given minimum and maximum volume.

When initiating the simulator, it will loop/drill forward in time with a given timestep for each iteration. When the defined number of stands are drilled, the simulator will stop and plot graphs representing surface parameters. These parameters are chosen as they are the easiest to obtain, and would be the most efficient and economic input parameters to use in a real working environment in implementation of a machine learning model for early kick detection. Additionally, surface parameters are not suffering from the same restrictions as downhole measurements that require transmission of signals through for example sonic waves, meaning surface parameters can be collected at a reasonably high efficiency [39]. Parameters plotted by the simulator are pump pressure, bottom hole pressure, drilled length, ECD, ROP, flowrate out, connection gas, pit gain, and hook load.

The code is written in Jupyter Notebook using python as coding language. This is a great software for presenting the code in a structured way. The file can be divided into several sections (cells), where the coder can choose to write code or use markdown to present an introductory text or to provide analysis in-between the coding sections. By importing libraries such as matplotlib, the data can be presented in a graphical manner in plots, whereas Jupyter Notebook has a great layout for

presenting the plots at the end of the coding window. This build up provided by the software makes it easier for the user to understand the structure, and the output from the code. However, this software is limited from a developer's point of view, where there exists no possibility of debugging line for line downwards in the code. Potential errors in the code can also be hard to detect when coding and receiving error messages. That being said, the advantages of the easy layout and structure for the end-user outweighs the limitations from a developer's point of view in this case.

### 3.1 Basic assumption of the Simulator

Figure 10 shows a shut-in well after taken an influx into the well that is situated as a single bubble in the drilling fluid (blue field). The dynamics of the kick phenomenon and its effect on the drilling parameters will be assessed through modeling and simulation studies. For this, the physics of the kick modelling will be formulated under some simplified assumptions.

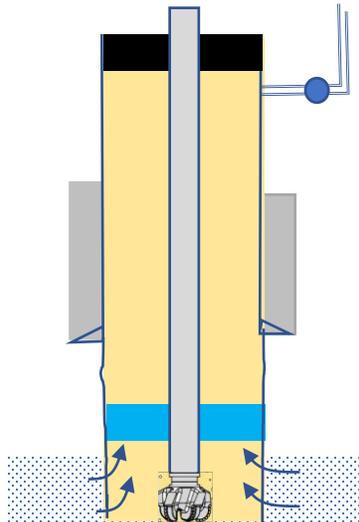


Figure 10: Visualization of single bubble approximated kick trapped in drilling fluid

The assumptions made for developing the simulator are defined as following:

- Only methane gas in influx.
- Real gas law used to calculate gas density as a function of pressure and temperature.
- Only vertical drilling.

- Kick/influx only occurs in the last drilling section. The kick will occur with an alterable probability defined by the user.
- Small kick volumes in the initial base case.
- Kick occupies the whole cross-sectional area.
- A 12 ¼ in hole size in the initial base case.
- Connection gas in previous 4 connection assumed if kick is taken in the last drilling section.
- No thermal driven volume expansion during connection.
- No expansion of kick.
- Constant influx rate of kick.
- Only water-based drilling fluid used.
- Single bubble approximation for the gas bubble, meaning the whole cross-sectional area is occupied by the gas bubble while it is completely separated from the mud [54].
- Kick will only be taken and will not be circulated out. It will only be present in the well during the event period. Hence, this code is primarily for generating data for kick detection purposes, and not circulating out the kick.
- Kicks are only situated around BHA due to their small size.
- Pump pressure is updated vs. depth since friction is proportional to well length.

### **3.2 Simulator structure**

The break-down of the code structure is shown in the appendix. Here appendix A.1) shows how the complete simulator structure is broken down. Furthermore appendices A.2) and A.3) shows the structure of the sub-systems; Drilling and Drilling with Kick, and Connection respectively. The sub-systems are broken down to show which parameters are updated in the different segments, and how the sub-system iterates forward. Each sub-system is shown as a single figure in the full simulation model, A.1). Here, these are marked as following:

- Sub-system Drilling = Drilling with circulation
- Sub-system Drilling with kick = Generate kick
- Sub-system Connection = Make connection

### 3.3 Impact of kick on drilling parameters

The True Vertical Depth (TVD) and the Measured Depth (MD) will be equal in this simulator as the well is assumed to be vertical. The drilled depth is measured by adding the ROP multiplied by the simulated time to the initial drilled depth at start of the simulation. The simulated time multiplied by the ROP is of course adjusted for the time spent during connection, when there is no drilling, such that the measured depth does not skip several time steps ahead for these periods. This is implemented in the code as shown in Figure 11 below.

```
measureddeptdata.append(drilleddepthstart+rop*(time-number_of_connections*connectiontime))
tvddata.append(drilleddepthstart+rop*(time-number_of_connections*connectiontime))
```

Figure 11: Code calculating the measured depth/TVD during drilling

This update on measured depth is important when calculating for example the bottom hole pressure and the pump pressure as these depend on the depth in the well. The calculations of parameters are shown below.

#### 3.3.1 Pump Pressure

Under certain conditions there is a possibility of mud flocculating, leading to an increase in viscosity when taking a kick [9]. As a result, the pump pressure might have a small increase, but will then drop down corresponding to the reduction in hydrostatic pressure as more and more formation fluid/methane gas enters the well. This is caused by the ‘U-tube’ effect happening as the lighter formation fluid displaces the heavier mud from the annulus. However, this simulator does not take account for the rapid and slight increase in pump pressure before the resulting reduction as we assume annulus friction is relatively low, meaning this effect most likely won’t happen or show on measurements, resulting in the plots only showing the reduction. The reduction in pump pressure during an influx incident is rather small compared to the usual pump pressure. As a result, the reduction is hard to notice on a full-scale plot for the complete simulation period, especially with noise. To obtain these results, the differential pressure during an influx incident is calculated as following:

$$\Delta P_{kick} [bar] = BHP - BHP_{kick} = (\rho_{mud} * 0.0981 * h + \Delta P_{ann}) - (\rho_{mud} * 0.0981 * (h - h_{kick}) + \rho_{kick} * 0.0981 * h_{kick} + \Delta P_{ann}) \quad (2)$$

Where:

- $BHP$  = bottom hole pressure [bar]

- $BHP_{kick}$  = bottom hole pressure with kick [bar]
- $\rho_{mud}$  = density of mud/drilling fluid [sg]
- $h$  = drilled depth [m]
- $h_{kick}$  = height of kick around BHA [m]
- $\rho_{kick}$  = density of kick [sg]
- $\Delta P_{ann}$  = annulus friction [bar]

The density of the kick is found by using the real gas law at conditions equal to pressure and temperature at the location where the kick is taken in the well. Pressure will equal bottom hole pressure, while temperature is calculated by using a geothermal gradient equal to 0.04 degrees Celsius per meter and with a temperature of 4 degrees Celsius at sea bottom. The temperature is then calculated by the following formula:

$$T_{bottom} = T_{seabottom} + (h - sea\ depth) * geogradient \quad (3)$$

Where:

- $T_{seabottom}$  = temperature at sea bottom [ $^{\circ}C$ ]
- $h$  = drilled depth [m]
- $geogradient$  = geothermal gradient [ $^{\circ}C/m$ ]

The z-factor can then be calculated at current temperature and pressure levels, along with gas gravity of methane gas as this is the assumed influx gas in the simulations. Combining all these, the density of the methane at given conditions can be found by following formula:

$$\rho_{kick} = \frac{M}{(R * T_{bottom})} * \frac{P}{z} \quad (4)$$

Where:

- $M$  = molar mass of methane [g/mol]
- $R$  = universal gas constant [J/(mol\*K)]
- $z$  = compressibility factor
- $P$  = pressure at influx location [Pa]
- $T_{bottom}$  = temperature at bottom [K]

The resulting pump pressure during influx/kick while circulating can then be calculated as:

$$\text{pump pressure with kick [bar]} = \text{pump pressure} + \Delta P_{inc} * h - \Delta P_{kick} \quad (5)$$

Where:

- $\text{pump pressure}$  = initial pump pressure at started drilled depth [bar]
- $\Delta P_{inc}$  = pump pressure increase needed per meter drilled [bar/m]
- $h$  = drilled depth [m]
- $\Delta P_{kick}$  = pressure reduction in bottom hole when kick is taken [bar]

Here, the  $\Delta P_{inc}$  will update as the simulator drills downwards, since the pump needs to compensate for more pressure loss due to friction. This is updated simply by just multiplying a factor for pressure loss with the increasing depth as the drilling process progresses. The pressure loss factor is found by dividing the initial pressure in the drill pipe by the initial drill pipe length.

The influx's effect on the pump pressure is rather small (depending on the kick size of course), meaning it is hard to detect by the naked eye. This can easily be exemplified by showing a simple calculation of how a 1 m<sup>3</sup> kick will impact the pump pressure, assuming the kick occupies the cross-sectional area around the BHA. Considering following conditions:

- Drilled depth = 4970 m, Sea depth = 100 m, and height of kick = 25.36 m
- Geothermal gradient = 0.04 °C/m and  $T_{seabot} = 4$  °C
- $\Delta P_{inc} = 0.0205$  bar/m, pump pressure = 200 bar, and  $\Delta P_{ann} = 5$  bar
- $\rho_{mud} = 1.7$  sg,  $z = 1.4677$ ,  $P = 4970m * 0.0981 * 1.7sg = 828.85$  bar, and  $M_{methane} = 16.04$  g/mol

The drop in pump pressure during influx can then be calculated by first finding the density of the kick at given conditions by using equation (3) and (4):

$$T_{bottom} = 4^{\circ}\text{C} + (4970\text{m} - 100\text{m}) * 0.04^{\circ}\text{C}/\text{m} = 198.8^{\circ}\text{C} = 471.95\text{K}$$

$$\rho_{kick} = \frac{16.04 \text{ g/mol}}{\left(8.314 \frac{\text{J}}{\text{mol} * \text{K}} * 471.95\text{K}\right)} * \frac{828.85 * 10^5 \text{ Pa}}{1.4677} = 230,853 \frac{\text{g}}{\text{m}^3} = 0.2308\text{sg}$$

then using equation (2) to calculate pressure lost by the kick:

$$\begin{aligned}\Delta P_{kick} &= (1.7sg * 0.0981 * 4970m + 5 \text{ bar}) \\ &\quad - (1.7 \text{ sg} * 0.0981 * (4970m - 25.36m) + 0.2308sg * 0.0981 * 25.36m \\ &\quad + 5 \text{ bar}) = 3.65 \text{ bar}\end{aligned}$$

The resulting pump pressure can then be calculated using equation (5):

$$\text{pump pressure with kick} = 200 \text{ bar} - 3.65 \text{ bar} = 196.35 \text{ bar}$$

As easily seen by these calculations, the drop in pump pressure is only 3.65 bar in the case of a small influx of 1 m<sup>3</sup>. Considering substantial amounts of noise in the measuring sensors, the driller will have a hard time noticing such a small change and not exchange this with noise or outliers.

### 3.3.2 Bottom Hole Pressure

The hydrostatic pressure and the frictional pressure in the annulus combined defines the bottom hole pressure in the well. Under normal drilling conditions with circulation the bottom hole pressure can be calculated as:

$$\text{Bottom hole pressure [bar]} = \rho_{mud} * 0.0981 * h + \Delta P_{ann} \quad (6)$$

Where:

- $\rho_{mud}$  = density of mud/drilling fluid [sg]
- $h$  = vertical drilled depth [m]
- $\Delta P_{ann}$  = annulus friction [bar]

Here the height ( $h$ ) will be updated in the simulator as the drilling proceeds. However, during an influx/kick incident some of the mud will be displaced by formation fluids or gas, resulting in a reduced hydrostatic pressure. The reduction in pressure comes from the difference in density between the drilling fluid and the influx fluid/gas over the length covered by the kick,  $h_{kick}$ . The height of the kick in the annulus around the BHA is calculated as:

$$h_{kick} [m] = \frac{\text{influx volume}}{\text{annular capacity}} \quad (7)$$

Where:

- *influx volume* = volume of methane flowing into well [ $m^3$ ]
- *annular capacity* = area between BHA and annulus [ $m^2$ ]

The annular capacity around the BHA is calculated as [55]:

$$\text{Annular capacity around BHA [m}^2\text{]} = \frac{\pi}{4} * ((di_{ann} * 0.0254)^2 - (do_{BHA} * 0.0254)^2) \quad (8)$$

Where:

- $di_{ann}$  = inner diameter of annulus [in]
- $do_{BHA}$  = outer diameter of BHA [in]

Note that the bottom hole pressure will increase when the height for the hydrostatic pressure increases, which is updated corresponding to the measured vertical depth as previously shown in Figure 11. The bottom hole pressure during an influx incident can then be calculated as:

$$\text{Bottom hole pressure with kick [bar]} = \rho_{mud} * 0.0981 * (h - h_{kick}) + \rho_{kick} * 0.0981 * h_{kick} + \Delta P_{ann} \quad (9)$$

Where:

- $\rho_{mud}$  = density of mud/drilling fluid [sg]
- $h$  = vertical drilled depth [m]
- $h_{kick}$  = height of kick around BHA [m]
- $\rho_{kick}$  = density of kick [sg]
- $\Delta P_{ann}$  = annulus friction [bar]

### 3.3.3 Equivalent Circulating Density

The ECD is kept constant during drilling except during connection where the ECD drops down to equal the mud weight as no circulation is ongoing in the well, and during an influx incident where the ECD drops corresponding to the reduced hydrostatic pressure in the well due to the influx. Thus, a reduction in ECD is expected when performing connections. However, a reduction in ECD during drilling with circulation will indicate an influx incident.

The ECD during drilling with circulation is calculated as [55]:

$$ECD [sg] = \rho_{mud} + \frac{\Delta P_{ann}}{h * 0.0981} \quad (10)$$

Where:

- $\rho_{mud}$  = density of drilling fluid [sg]
- $\Delta P_{ann}$  = annulus friction [bar]
- $h$  = vertical drilled depth [m]

During an influx incident, there will be a reduction in the hydrostatic pressure due to influx of formation fluids giving a reduction equal to  $\Delta P_{kick}$  as shown in section 3.3.1. Now the reduced ECD during this incident can be calculated as:

$$ECD_{kick} [sg] = \rho_{mud} + \frac{\Delta P_{ann} - \Delta P_{kick}}{h * 0.0981} \quad (11)$$

Where:

- $\rho_{mud}$  = density of mud/drilling fluid [sg]
- $\Delta P_{ann}$  = annulus friction [bar]
- $\Delta P_{kick}$  = pressure reduction in bottom hole when kick is taken [bar]
- $h$  = Drilled depth [m]

### 3.3.4 Rate of Penetration

An early indicator for kick is a sudden increase in the ROP (called drilling break), as this is altered before the kick occurs [9]. This parameter can be affected by many factors; however, the increase could be caused by an increase in permeability, formation porosity or formation pressure. As a result, a change in these factors may lead to taking an influx of formation fluids/gas into the well. However, as this factor is heavily dependent on rock compressive strength, this should be checked against surrounding lithology (the character of the rock formation) [56]. The ROP will often increase during a transition from drilling a hard formation to a soft formation [57]. As a result, this increase might lead to a false indication that a kick is occurring.

ROP is an extremely important factor in early kick detection, as this is an early indicator that a kick is occurring as stated above. Yin, et al. (2021) presented a study on the importance of different input parameters for early kick detection using machine learning [42]. Here, 9 parameters were tested using the random forests method to find their importance for contribution during early kick detection by measuring the detection accuracy while randomly varying each feature. Their findings are presented in Figure 12 below. This study shows that the ROP is in fact the most important parameter for early kick detection.

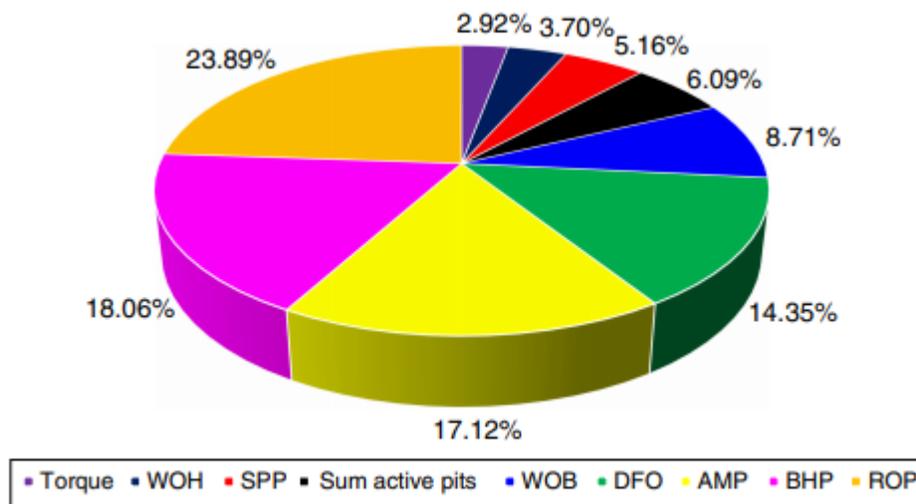


Figure 12: Importance of different input parameters for ML kick detection [42]

Note that we keep the ROP constant in this simulator, thus any increase in ROP will indicate that a kick is incoming as no lithology is implemented in the model. A drilling break is often characterized by an ROP increase of more than 20% [42]. The simulator's ROP increase can be adjusted in the initial parameters and is initially set to be a random increase of 20-25% in the base case.

The sudden ROP increase is generated by overwriting the previous 12 data-points in the ROP when a kick is generated in the simulator. As the current timestep is set to 10 seconds, the sudden increase will happen over a period of 120 seconds. This period might however be shorter if the overwriting is moving back into a connection, whereas the ROP increase will start at the end of the connection period and move up until the kick is taken with a maximum of 12 steps. The code for doing this is shown in Figure 13 below:

```

"Adding a sudden ROP increase before kick"
for i in range(12):
    sudden_increase = int(len(ropdata)-1-i) # Adding a point for sudden increase some steps before kick
    "Breaking out of loop if ROP increase is moving into the connection"
    if(np.any(np.in1d(connection_array,timedata[sudden_increase]))):
        break
    else:
        ropdata[sudden_increase] \
        = ropdata[sudden_increase]*rop_increase # Adding a sudden increase in rop before kick

```

Figure 13: Code for adding a sudden increase in ROP before kick is taken

The increased ROP will then be kept constant during the whole event period, giving a total increased ROP duration of sudden increase period + event period. Since the event duration is initially set equal to 60 seconds, we will have an increased ROP in a total of 180 seconds for the base case. Note that both the sudden increase period and the event period can be adjusted in the code.

### 3.3.5 Flowrate Out

The flowrate out is an important parameter to track in order to detect a possible influx or mud loss [21]. A difference in the flowrate in and the flowrate out gives a rapid indication on whether there is a loss of mud or influx of formation fluid occurring in the well. An increase in the flowrate out is the first indication that a kick is taken into the well [9]. However, this increase might be at such a small size that it is difficult to register, depending on equipment installed. In such an event, the increase in the active mud volume might confirm that an influx has occurred. Another challenge that can make a kick hard to detect using the flowrate out is the widely use of inaccurate measurement equipment such as mechanical flow paddles [58]. There is of course other existing equipment as an alternative measurement method of fluid level in return line, but they are all associated with containing a large amount of random noise along with varying biased values as mentioned in section 2.1.3.

Formation fluid with low density, as for example gas, reduces the hydrostatic pressure in the annulus rapidly as more and more influx are taken into the well [9]. Consequently, the influx-rate might increase very quickly. This effect has, however, been excluded in this simulator. The influx rate is kept constant during the kick event where the rate is set to be equal to the size of the kick distributed evenly over the event period. The formation fluid flowing into the well will displace the mud, forcing it out of the well, resulting in an increase in the flowrate out.

Under normal conditions the flowrate out is kept equal to the flowrate in. The flowrate out will however increase by the magnitude of the influx rate during the kick event. Here, the flowrate out during a kick event is calculated as in following equation:

$$\text{Flowrate out with kick [lpm]} = \text{flow out [lpm]} + \text{influx rate}[m^3/s] * 1000 * 60 \quad (12)$$

Where:

- *Flow out* = flow out of well during circulation [lpm]
- *influx rate* = rate of influx into well [m<sup>3</sup>/s]

### 3.3.6 Connection gas

The bottom hole pressure will be reduced during a connection as the pumps are stopped, resulting in the possibility of an influx gas occurring if the formation pressure becomes higher than the well pressure [9]. This reduction is due to the mud being static during connection, thus removing the frictional pressure from the annulus obtained during circulation. As a result, gas from the formation might start to seep into the well, generating a small influx. The presence of connection gas can be used as an indicator for underbalanced or near-balance drilling, meaning the formation pressure is greater than or near the equivalent density of the static mud column [59]. An increasing trend in connection gas can therefore be used as an early kick detection indicator.

Influx taken during connection usually has a small influx rate and volume, leading to difficulties of identifying this incident at the surface immediately. The connection gas is measured at the surface as changes in density in return mud, meaning the measurements will happen with a time lag as it needs to be circulated to the top of the well. The trend of the connection gas is also important to carefully observe, as this might be the first factor to signal that the well is near or at balance [59].

The simulator calculates the delay of the influx as a function of the volume of the annulus and the flowrate into the well. Following formula is used to calculate the volume of the annulus:

$$V_{annulus} = \text{annular capacity DP} * \text{length DP} + \text{annular capacity BHA} * \text{length BHA} \quad (13)$$

Here, the annular capacity of the BHA is calculated as shown in equation (8), and the annular capacity around the drill pipe is calculated by following formula:

$$\text{Annular capacity around DP [m}^2\text{]} = \frac{\pi}{4} * ((di_{ann} * 0.0254)^2 - (do_{DP} * 0.0254)^2) \quad (14)$$

Where:

- $di_{ann}$  = inner diameter of annulus [in]
- $do_{DP}$  = outer diameter of drill pipe [in]

After calculating the annular volume, the delay of the measurement time can be found by simply dividing the annular volume by the flowrate into the well, plus adding any potential connection periods where the well is kept static. Note that the single bubble approximation does not apply for the connection gas as this is of a very small size. The gas is therefore assumed to be mixed in the mud, and then identified by measuring density of return mud. The simulator calculates this delay as shown in Figure 14.

```

if(len(kick_det_times)>1):
    for i in range(4): # Assuming connection gas in previous 3 connections before kick occurs
        annular_volume = annular_capacity*length_BHA+annular_cap_DP*(drilleddepthstart\
            +standlength*(connections-i-1)-length_BHA)

        influx_timeindex = int((connectiontime/timestep +\
            measureddeptdata.index(drilleddepthstart+standlength*(connections-i-1)))
            "Checking if the delayed connection gas measurement is inside the simulation time range"
        if(timedata[influx_timeindex]+round(annular_volume/(flowin/1000),-1)*60+connectiontime < timedata[-1]):
            "Checking if delayed connection gas i marked in connection, moving it a connection period forth in such a case"
            if(np.any(np.in1d(connection_array,timedata[influx_timeindex]))):
                influx_delays.append(timedata[influx_timeindex]+round(annular_volume/(flowin/1000),-1)*60+connectiontime)
            else: # Adding the delayed connection gas measurement to an array storing the time they were measured
                influx_delays.append(timedata[influx_timeindex]+round(annular_volume/(flowin/1000),-1)*60)

```

Figure 14: Iterations to find time index of delayed connection gas indicator

This code will only run if there is in fact obtained a kick during the last drilling section, under the assumption that there is no connection gas if no kick is obtained. If there is in fact a kick during the last drilling section, the simulator will enter a loop to calculate the annular volume and gas measurement delays of the previous 4 connections before kick is taken. However, only 3 of these will show as the last delay will occur after the simulated time period ends. This is because one has to drill more than one section for the gas to be circulated to the surface with assumed wellbore geometry and circulation rate. The code will identify if the delayed measurement is inside a connection and move it one connection period out of the connection. Lastly, the specific time the delayed influx fluid measurement is observed is added to an array.

As there is no specific volume measurement of the connection gas, the plot will be binary where 1 indicates there has been a seepage of gas into the well during connection and 0 indicates no connection gas. To show this clearly in a plot, the delayed timers are tagged as 1 in an array consisting only of zeros. This is tagged in an array by inserting a 1 at the delayed time index in an array filled with zeros as shown in Figure 15. The tag is therefore placed at the same position as the time the connection gas is measured at the surface, whereas the last delayed connection gas measurement is not taken into the array due to the maximum simulated time.

```
"Adding the influx event to an array storing 0 for no influx during connection and 1 for influx during connection"
connectiongas = np.zeros(len(timedata)) # To store values 0 for no gas and 1 for influx during connection
for i in range(len(influx_delays)):
    if(len(timedata)<(influx_delays[i])/timestep): # Breaking out of loop if the influx delay is outside simulation data
        break
    else:
        connectiongas[timedata.index(influx_delays[i])] = 1 # Marking the connectiongas at a delayed time
```

Figure 15: Creating an array to store delayed connection gas at correct time index

After the arrays are updated with correct data to identify when the delayed connection gas is measured in time, the connection gas is plotted as shown in Figure 16 below:

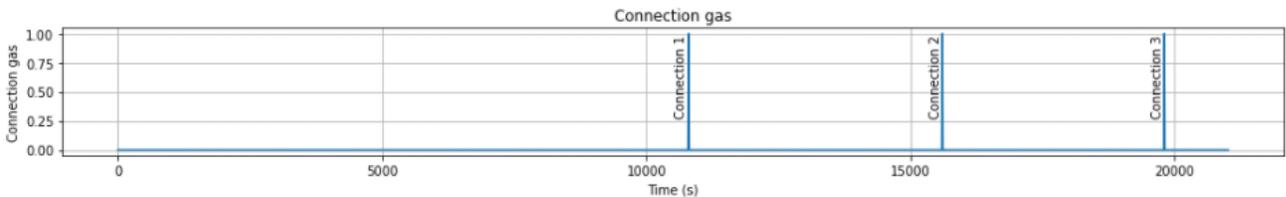


Figure 16: Example of how connection gas is presented in a 5 connection simulation

This clearly shows at what time the connection gas is identified by surface sensors. By comparing this to Figure 17, it is easily seen that there is in fact a delay on the measured connection gas, where Figure 17 shows that the first connection happens some time just before 5,000 s and the measured connection gas is not identified before just after 10,000 s.

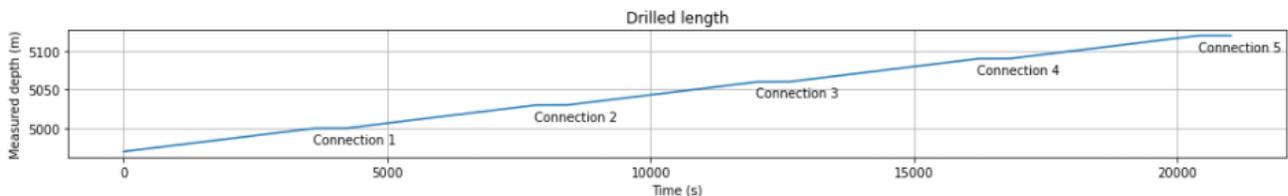


Figure 17: Example of drilled depth during a 5 connection simulation

### 3.3.7 Pit gain

Irregular changes in pit volume is one of the prime indicators of either mud loss or influx of formation fluid [60]. An obtained increase in active mud volume, that is not due to internal transfer of mud between mud pits on the rig, is almost always caused by influx of formation fluid into the well [9]. This can be used as a so-called ‘100% indication’ of a kick by correlating this to an increase in return flow from the well. Pit gain increase can also be caused by thermal expansion of mud (which is not considered in this simulator), but this takes longer time to notice in the pit gain [61]. While taking a kick on the other hand, the pit volume will increase rather rapidly. It is however worth to mention that the raw data in pit gain often are dominated by discontinuities as a consequence of removing or adding tanks from the active pit volume [29]. These data can on the other hand be removed without affecting important data as these are not related to the events occurring in the well, thus providing a set of modified pit volume data that can be used for more precise kick detection. This difference in raw data and filtered data is shown in Figure 18 below. This shows that the trend is much easier to detect using the filtered data compared to the raw data with a substantial number of outliers. There is however still some noise in the filtered data, where most of the spikes seems to be most likely generated by pump stops during drilling operations.

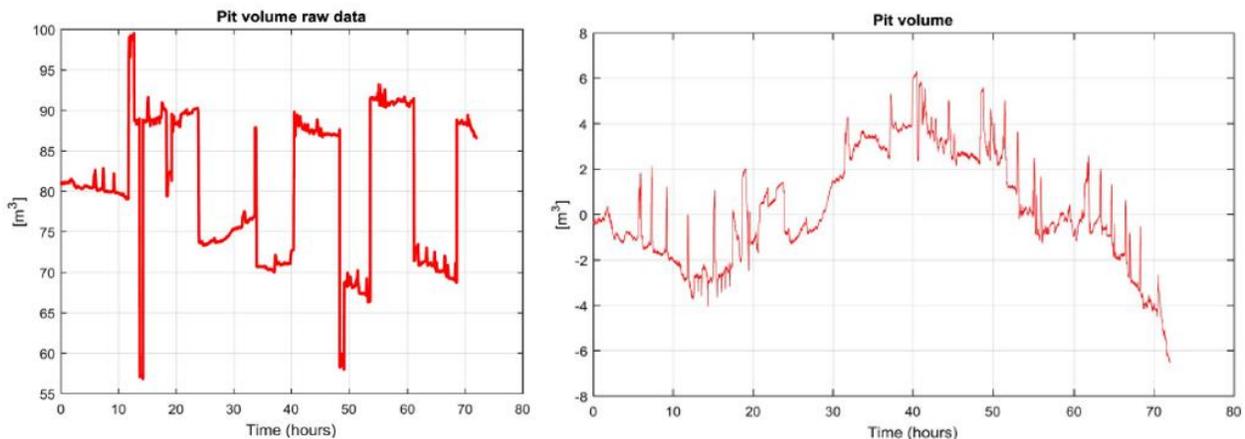


Figure 18: Comparison of raw pit volume data (left) vs. filtered pit volume data (right) [29]

The simulator in this thesis assumes an incompressible drilling fluid, meaning it will completely displace a mud volume equal to the influx volume. The pit gain increase will therefore indicate what volume of kick is taken. In a real case scenario, a correctly sized and instrumented active system can detect influxes of around  $1 \text{ m}^3$  in the pit gain measurements, even on floating drilling

units [9]. This simulator does however introduce kicks in the pit gain for any given volume, as the purpose of this code is made for training of a machine learning algorithm to detect relatively small kicks and interpret warning signals for early kick detection. Kicks are therefore additionally tagged in an array by the number 1 and non-kicks are tagged by 0. This means that an array consisting of solely zeroes and ones will be generated to identify at exactly what point in the data set kicks are taken. Consequently, one can use this array to verify that the machine learning algorithm has made the correct identifications.

### **3.3.8 Hook load**

The hook load is an important parameter for controlling drilling operations by for example controlling the weight on bit. This shows the total forces dragging the drill string and components down, thus representing how smooth the drilling operation is (i.e. that there are no irregularities such as stuck pipe, influx, etc.) [62]. Here, the driller will study potential increases and decreases in hook load to interpret different events happening downhole, often to identify potential cuttings accumulation to prevent stuck pipe incidents from occurring.

A slight increase in hook load can be observed when a kick enters the well [9]. This is a result of the lighter weight of the formation fluid compared to the drilling fluid, whereas the increase in hook load will be more significant the lighter the weight of the formation fluid. The displacement of mud by formation fluid will reduce the buoyancy effect, resulting in an increase in the effective weight of the drill string. This is noticed as an increase in hook load at the surface. It is important to state that this is not a major increase, whereas the kick would have to be very large before the increase in hook load is easily noticeable at the surface. Another contributing factor is that the data quality of hook load measurements often is of poor quality [63]. This is practically measured indirectly as either a tension in the deadline or in the travelling equipment. Consequently, the hook load is exposed to load-generating forces between the top of the string and the location of measurement. Any load contribution from these forces can contribute to a substantial amount of noise in the measuring data. As a result, this is not one of the primary methods of detecting influx of formation fluids.

When calculating the hook load, several factors needs to be considered, such as the most important factor: buoyancy effects. The fact that the buoyancy of a body equals the weight of the displaced

fluid in which it floats was discovered by Archimedes (approx. 300 B.C.) in ancient Greece [64]. By multiplying the weight of the drill pipe (including BHA) in air by a calculated buoyancy factor the submerged weight of the drill pipe is found. This calculation is valid for both vertical and deviated wells. However, this simulator assumes fully vertical well.

The buoyancy factor can be calculated as following:

$$\text{Buoyancy factor} = 1 - \frac{\rho_{mud}}{\rho_{steel}} \quad (15)$$

Where:

- $\rho_{mud}$  = density of drilling fluid [sg]
- $\rho_{steel}$  = density of steel [sg]

Note that this is only valid when the inside and the outside of the drill pipe is filled with the same fluid (i.e. the same density). Now the hook load under normal condition can be calculated as following:

$$\text{Hookload [kN]} = ((m_{topdrive} + \text{bouyancy factor} * (m_{DP} * \text{length DP} + m_{BHA} * \text{length BHA}) - \text{wob}) * 9.81) / 1000 \quad (16)$$

Where:

- $m_{topdrive}$  = mass of topdrive [kg]
- $m_{DP}$  = mass of DP [kg]
- $\text{length DP}$  = length of DP [m]
- $m_{BHA}$  = mass of BHA [kg]
- $\text{length BHA}$  = length of BHA [m]
- $\text{wob}$  = weight on bit [kg]

During an influx of formation fluid, a section of the annulus will be filled with another fluid than the inside of the drill string, which is still filled with mud. This will affect the buoyancy effect for that specific section. Now the buoyancy for the zone in the annulus around the BHA occupied by the kick can be calculated as following [64]:

$$\text{Buoyancy factor kick} = 1 - \frac{\rho_{kick} * do_{BHA}^2 - \rho_{mud} * di_{BHA}^2}{\rho_{steel} * (do_{BHA}^2 - di_{BHA}^2)} \quad (17)$$

Where:

- $\rho_{kick}$  = density of kick/methane gas [sg]
- $d_{o_{BHA}}$  = outer diameter of BHA [m]
- $\rho_{mud}$  = density of drilling fluid [sg]
- $d_{i_{BHA}}$  = inner diameter of BHA [m]
- $\rho_{steel}$  = density of steel [sg]

It is worth mentioning that this formula can vary depending on assumptions made, whereas the kick density can be a mix of drilling fluid and formation fluid and that the kick does not necessarily have to be situated around the BHA. This simulator does however assume only methane gas, single bubble approximation, and that the kick is only situated around the BHA, meaning that the density of the kick will equal the calculated density of the methane gas at given depth and pressure.

Now, using the updated buoyancy factor for the area covered by the kick, the increased hook load can be calculated by following formula:

$$\text{Hookload with kick [kN]} = ((m_{topdrive} + \text{buoyancy factor} * (m_{DP} * \text{length DP} + m_{BHA} * (\text{length BHA} - h_{kick}))) + \text{buoyancy factor kick} * (m_{BHA} * h_{kick}) - \text{wob}) * 9.81 / 1000 \quad (18)$$

Where:

- $m_{topdrive}$  = mass of topdrive [kg]
- $m_{DP}$  = mass of DP [kg]
- $\text{length DP}$  = length of DP [m]
- $m_{BHA}$  = mass of BHA [kg]
- $\text{length BHA}$  = length of BHA [m]
- $\text{wob}$  = weight on bit [kg]
- $h_{kick}$  = height of kick around BHA [m]

It is worth to mention that, as previously stated as well, this is not easy to detect at the top of the well. Considering a 1 m<sup>3</sup> kick taken around the BHA, there will be a hook load increase of around 1.36 tons, which is extremely hard to notice at a hook load of around 167 tons consisting of substantial amounts of noise.

### **3.4 Application of model**

Due to difficulties in generating real-like data sets in laboratory environments, this thesis focused on developing a simulator for generating large data sets with parameters typically used as input parameters for machine learning models made for early kick detection. For the model to make as realistic predictions as possible, it is important to simulate an environment close to realistic conditions. Consequently, the simulator is built to take 30 initial parameters in order for the user to create an environment described to be as similar to expected well conditions as possible in the simulations.

The sub-sections below will present the initial base case for the simulator, as well as two varied simulations to visualize the flexibility of the simulator. Here, case 1 will be a simulation of the simulator base case as presented in section 3.4.1 below, and case 2 will be a simulation of equal conditions with larger influx size, number of connections, and event duration. This is simply to show how the simulator can be extended to simulate larger kicks over larger event periods, and to visualize how the results will differ in similar plots based on duration and size.

#### **3.4.1 Simulator base case**

We assume a vertical well starting of at 4970 meters drilling one stand, then making a 10-minute connection, then drilling the next stand, repeating for a decided number of connections (initially set to 5). Each drilling section will equal the length of one stand, initially set to 30 meters, meaning a 5 connection simulation will drill down from 4970 meters to 5120 meters. We assume a 12 1/4-inch section with a 5-inch drill pipe and a 100-meter-long BHA with outer diameter of 8.5 inches. Mud weight is 1.7 sg and only water-based drilling fluid is considered. Top drive weighs 25 ton and we drill with 5-ton WOB. The drill pipe weighs 33.57 kg/m in air and the BHA weighs 251.59 kg/m in air. We assume the BHA consist mainly of drill collars with inner diameter of 3 inches, with steel density set to 7.85 sg. The annular friction while circulating is assumed to be 5 bar based on relatively low friction for this section. Pump pressure is assumed to be 200 bar at 3000 lpm. We will assume that we take a kick while drilling the last section, but also assume that it is concentrated and located around the BHA where it will have maximum height. The simulation is set into a scenario where we are drilling with a bottom hole pressure close to the pore pressure, such that connection gas can be seen during static periods. Meaning that small amounts of influx are taken

and circulated towards the surface. When drilling the last section, the pore pressure increases rapidly, and a larger influx is likely to be taken. However, this is based on the probability of taking a kick during this section, initially set to 0.2% for each timestep. This seems like a pretty low chance of obtaining a kick, but as the simulator uses many timesteps to drill ahead it is likely to happen at least once during the last drilled section.

Only methane gas influx is assumed. Here, the volume assumed to initially be set to be randomly drawn value from a uniform distribution considering kick volumes between  $0.05 - 1 \text{ m}^3$ . This can, however, be changed by the user to obtain either a bigger or smaller kick volume. The bounds of the randomly drawn kick from a uniform distribution is set by two alterable parameters defining the minimum and maximum size of the kick in  $\text{m}^3$ . The mentioned probability of obtaining a kick can of course be altered by the user to either lower the possibility of a kick or to heighten the possibility of obtaining a kick during this drilling section. We will use the single bubble approximation, assuming that the gas occupies the whole cross-sectional area and is separated from the mud [54]. Assuming that it is only around the BHA will make the drop in hook load more evident than if it was around the drill string since BHA is much heavier than drill pipe and a longer section of it is exposed to gas. Note that the gas influx will only be on the outside of the BHA and a modified equation has to be used to calculate the buoyancy [64]. The position of the bit is initially at 4970 meters. The water depth is 100 meters with 4 Celsius at sea bottom and the geothermal gradient is 0.04 C/m. The gas influx will have a temperature equal to the formation temperature and the pressure can be assumed to be the same as the circulating pressure (hydrostatic pressure + friction)

We will here investigate the effect of gas influx on:

- Pressure drop in the well and how this affects the pump pressure and the bottom hole pressure.
- The reduction in ECD when kick is taken .
- How ROP suddenly increase before a kick as an early warning signal .
- How the influx rate of a kick affects the flowrate out of the well.
- Influx volume of kick and how this affects the pit gain under the assumption that the drilling fluid is incompressible, meaning increased flowrate out at top will equal influx rate of kick into annulus.

- The buoyancy and consequently how much the hook load will increase if a kick is taken.

Detection of connection gas is also implemented, functioning as early warning signals of incoming kick. The simulator is built using the physics and formulas presented earlier in sections 3.3.1-3.3.8.

### 3.4.2 Case 1 - base case simulation

The first presented case is a run of the base case as described in 3.4.1. Here the following initial conditions are set:

Timestep = 10 s	Outer diameter BHA = 8.5 in	Block lift time = 120 s
Stand length = 30 m	Inner diameter BHA = 3 in	Number of connections = 5
ROP = 0.00833 m/s	Inner diameter annulus = 12.25 in	Probability of kick = 0.2% per timestep
Connection time = 600 s	Outer diameter drill pipe = 5 in	Event duration = 60 s
Drilled depth at start = 4970 m	Weight drill pipe = 33.57 kg/m	Kick size minimum = 0.05 m <sup>3</sup>
Pump pressure at start = 200 bar	Weight BHA = 251.59 kg/m	Kick size maximum = 1 m <sup>3</sup>
Annulus friction = 5 bar	Length BHA = 100 m	ROP increase max = 1.25
Flowrate in = 3000 lpm	Weight top drive = 25,000 kg	ROP increase min = 1.20
Mud weight = 1.7 sg	WOB = 5,000 kg	Pressure drill pipe = 100 bar
Steel weight = 7.85 sg	Block height = 30 m	Starting time = 0

Table 3: Initial input parameters base case simulator run

The simulation will generate a kick at random times during the last drilled section each time it is run. However, kicks might not be taken in all simulations as this is based on a given probability. As noise in the data is a challenging factor in detecting kicks during drilling, the simulation is run once with fictional noise and once without to visualize the kick detection difficulty. Here, the noise is set with mean value of 0 as noise is added to the calculated and simulated data in the simulator, meaning the real mean value would be the calculated value. The standard deviation of each parameter is presented in Table 4 below. It is worth to mention that the noise is only set as fictional values that does not necessarily represent the realistic noise in the sensor equipment used in drilling operations. This will be represented as a proposed further work in the conclusion chapter. The purpose of adding noise is first of all to demonstrate that it may be difficult to track changes in

drilling data that are warning about a kick for certain parameters. But, as pointed out, the noise assumed here may not represent the real accuracy of the sensors.

Parameter	Standard deviation	Unit
Flowrate out	101.11	lpm
Hook load	15.42	kN
Pump pressure	2.5	bar
Bottom hole pressure	2	bar
ROP	0.0005	m/s
Pit gain	0.05	m <sup>3</sup>

Table 4: Overview of standard deviation in noise data

Now, running the simulator twice; once with noise and once without noise several drilling data will be generated, plotted, and available for comparison. Here, a 0.79 m<sup>3</sup> influx was generated in the simulation without noise and a 0.85 m<sup>3</sup> influx was generated for the simulation with noise. Firstly, comparing the pump pressure:

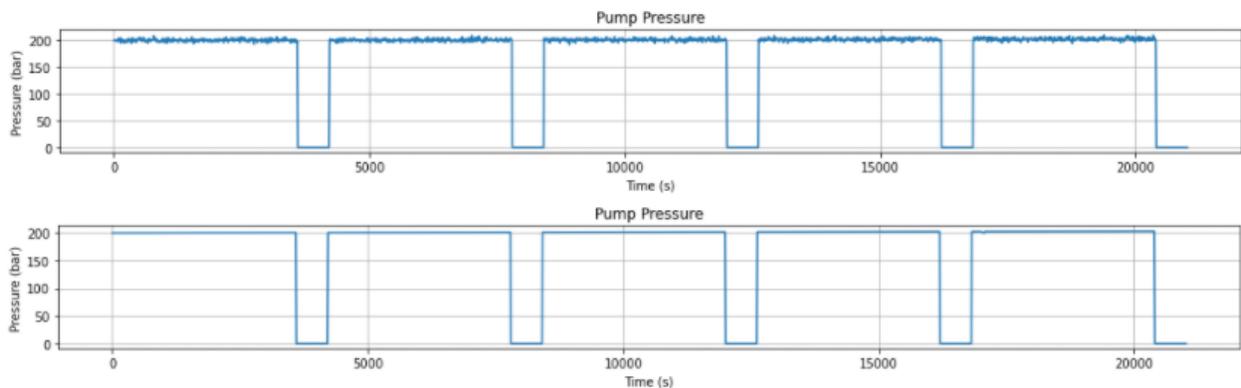


Figure 19: Comparison of pump pressure data with and without noise case 1

The reduction in pump pressure in Figure 19 is not easy to detect, as also previously shown in calculations of a 1 m<sup>3</sup> influx in section 3.3.1. In this case, the reduction will be below 3.65 bar, meaning it is very hard to detect using the naked eye, especially with noise in the data. However, this can be easier to detect when zooming in at the graph as in the cut-out of the last drilled section shown in Figure 20.

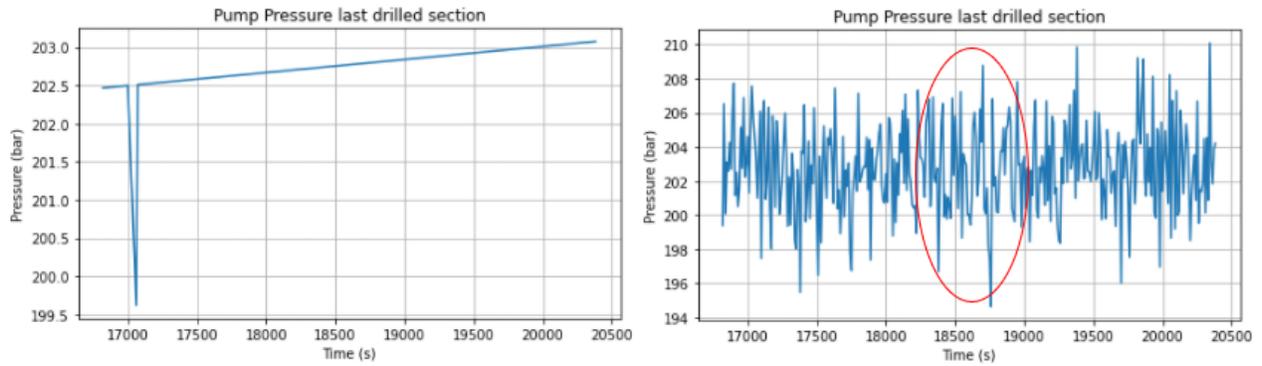


Figure 20: Pump pressure last drilled section with and without noise

Here, the graph to the left shows a reduction of pump pressure of just below 3 bar, whereas the graph with the noise seems to have an undetectable influx due to a substantial amount of noise. However, the influx should be somewhere inside the marked red circle. It is worth to mention that this is a relatively small influx, making it even harder to detect.

Now, looking at the bottom hole pressure. As stated in 3.3.2 this should be reduced when obtaining an influx due to lighter formation fluid displacing the drilling fluid. This is somewhat easy to detect in a data set without noise, however it is extremely hard to detect in a data set with noise at such a small influx. This is visualized in Figure 21 below. The large sequenced drops are the drop in bottom hole pressure during connection as the pumps are turned off.

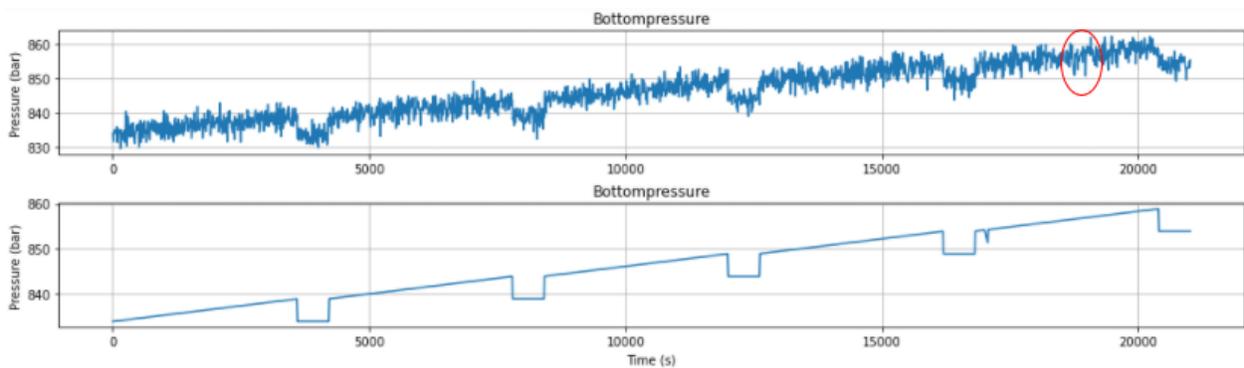


Figure 21: Comparison of bottom hole pressure data with and without noise case 1 (influx marked with red circle)

The ECD data are made without noise, thus only the plot from the simulation without noise will be represented as shown in Figure 22. This is not one of the primary indicators used for kick, but it does however get affected by the influx. As explained in section 3.3.3, the ECD will drop during

connections due to the pumps being turned off, and when obtaining an influx as lighter density formation fluids are entering the well, reducing the overall density. The reduction resulting from the influx can easily be seen in the plot below, even though it is not of large significance on a bigger scale.

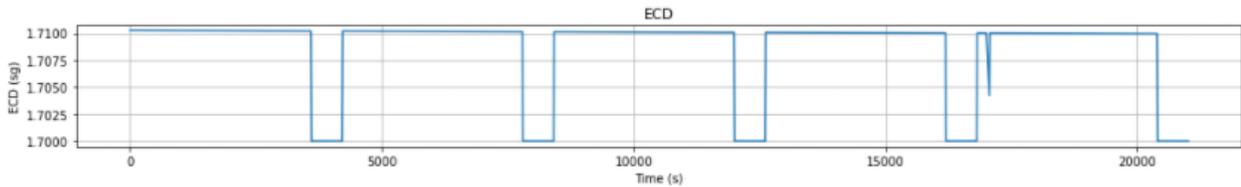


Figure 22: ECD data from simulation without noise case 1

Now, looking at the ROP data, it is seen that an increase is observed both in the noisy data and in the data without noise. But this could of course be harder to detect with a larger amount of noise. This parameter is known to be a preliminary warning signal, as the ROP will have a sudden increase just before an influx is taken. This is depicted in Figure 23 below. The ROP is also one of the most important parameters in early kick detection, as stated in section 3.3.4. It is worth to mention, again, that this can be affected by the rock hardness, resulting in sudden increases not related to an incoming influx.

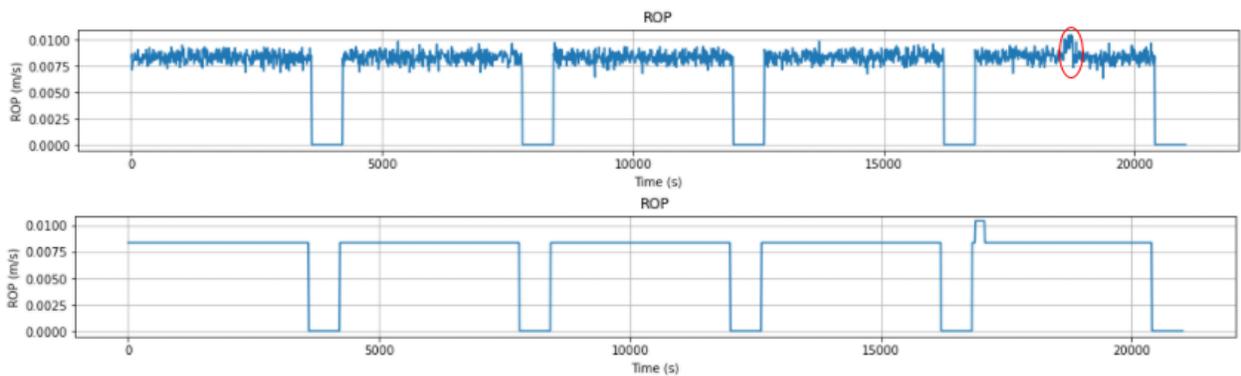


Figure 23: Comparison of ROP data with and without noise case 1 (influx marked with red circle)

The flowrate out of the well is an important parameter to track as it is one of the primary indicators of a kick. Observing an increase in flowrate out, without changing the flowrate in will be a solid indicator that a kick is occurring. This is due to the influx pushing out volumes of drilling fluid out of the well at a rate equal to the influx rate, where the impact will look somewhat like represented

in Figure 24. This is usually measured in liters per minute, whereas the influx rate is in  $m^3$ , meaning the impact will be significant as  $1 m^3$  equals 1000 liter. The influx can therefore be easily observed by the increase in the flowrate out both with and without noise. To make this a more realistic scenario, calculations for thermal expansion of drilling fluid during connection should be implemented, as well as the possibility of the influx to get mixed in the mud or dissolve into for example an oil based drilling fluid.

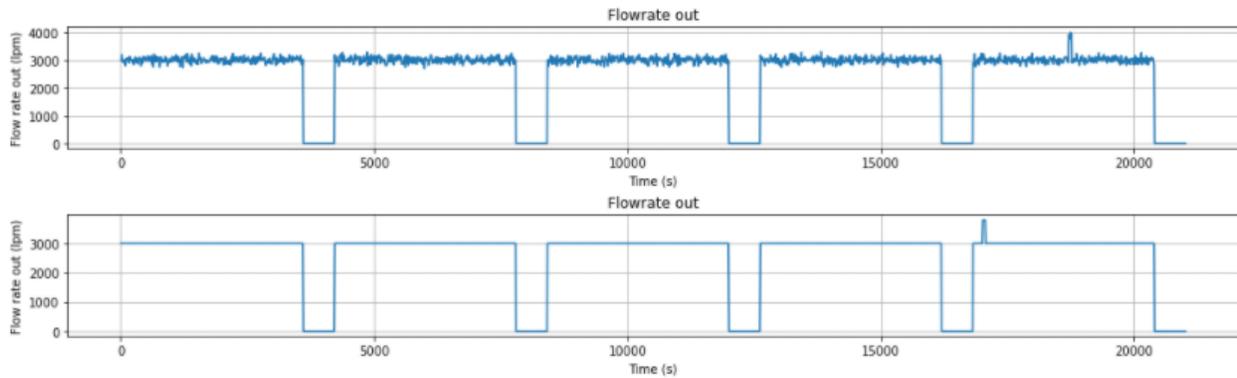


Figure 24: Comparison of flowrate out data with and without noise case 1

Now looking at the pit gain. This is assumed to stay at 0 for the whole simulation period due to no thermal expansion or other affecting instances not being implemented. It is therefore easy to detect the occurring influx, even for the noisy data, as shown in Figure 25. This is also one of the primary indicators of an incoming kick. The pit gain will equal the inflowed volume at the given time, as the simulator is under the assumption of single bubble approximation and incompressible drilling fluid. The spike in pit gain will therefore be quite noticeable.

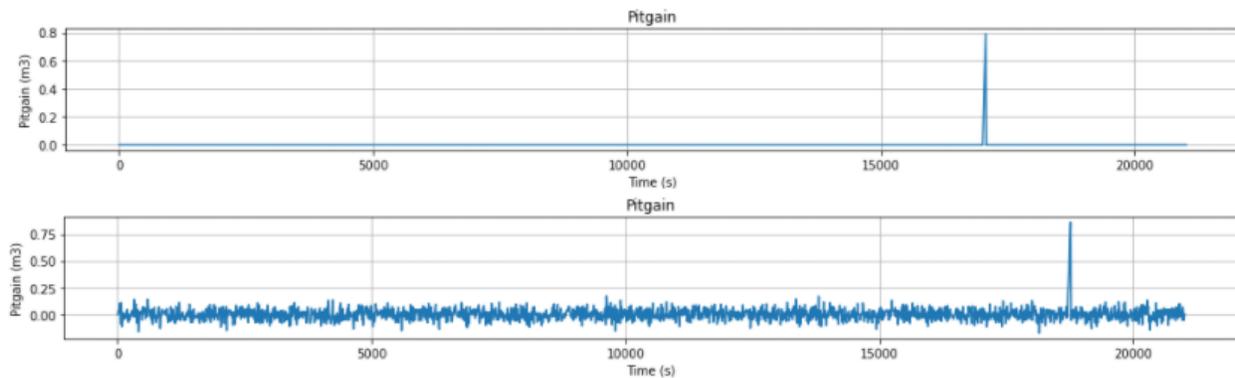


Figure 25: Comparison of pit gain data with and without noise case 1

Now, moving on to the last surface parameter, the hook load. As described in section 3.3.8, this will not be affected noticeably by small influxes. Here, a 1 m<sup>3</sup> influx would only reduce the hook load by 1.36 tons, which is very hard to detect at hook loads in a magnitude of around 167 tons. In Figure 26 it is clearly seen that the increase in hook load cannot even be seen in the plot without noise. The 5 sudden drops in hook load is a representation of the drill string being locked in slips during connections, whereas the block (the black line) moves to the top position to connect to a new stand.

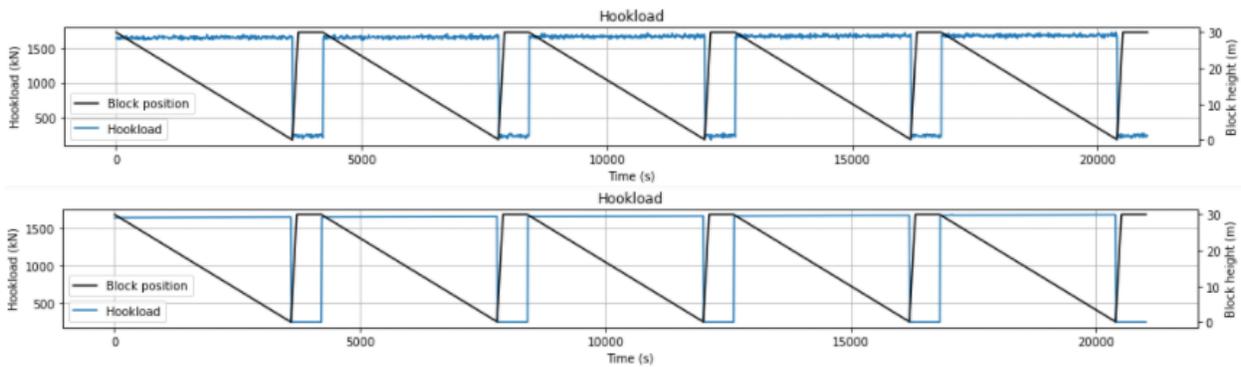


Figure 26: Comparison of hook load data with and without noise case 1

However, studying this closer by looking solely at the last drilled section as done with the pump pressure the increase in hook load is much more detectable as seen in Figure 27. However, looking at the plot to the right, the increase in hook load will be undetectable for the naked eye in this noisy data. As a result, this would need a large amount of filtering/pre-processing in order for the machine learning algorithms to make precise predictions.

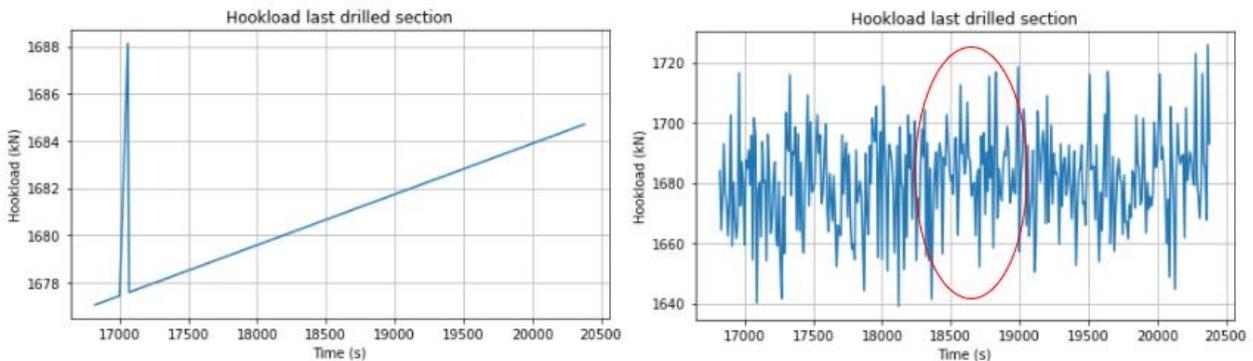


Figure 27: Hook load last drilled section with and without noise case 1 (influx section marked with red circle)

It is also worth to mention that both these simulations will generate plots exactly like the one shown in Figure 16 and explained in section 3.3.6. This shows that connection gas has been occurring during the previous 3 connections prior to the kick. The connection gas can therefore serve as a preliminary warning signal that a kick is occurring, as with the sudden ROP increase. Furthermore, section 3.4.3 will show how the plots are affected when obtaining a larger influx and with more connections made. This is to show how a kick the driller typically can detect without hesitation looks.

### 3.4.3 Case 2 – more connections, longer event period, and larger influx

To show the flexibility of the simulator, the influx size, number of connections, and event duration is changed while other factors are kept equal to case 1 in order to keep a rather comparable result as shown in Table 5 below. This is also to demonstrate how the well parameters will behave for a kick size that should definitely be detected by the driller.

Timestep = 10 s	Outer diameter BHA = 8.5 in	Block lift time = 120 s
Stand length = 30 m	Inner diameter BHA = 3 in	Number of connections = 8
ROP = 0.00833 m/s	Inner diameter annulus = 12.25 in	Probability of kick = 0.2% per timestep
Connection time = 10 min	Outer diameter drill pipe = 5 in	Event duration = 180 s
Drilled depth at start = 4970 m	Weight drill pipe = 33.57 kg/m	Kick size minimum = 3 m <sup>3</sup>
Pump pressure at start = 200 bar	Weight BHA = 251.59 kg/m	Kick size maximum = 4 m <sup>3</sup>
Annulus friction = 5 bar	Length BHA = 100 m	ROP increase max = 1.25
Flowrate in = 3000 lpm	Weight top drive = 25,000 kg	ROP increase min = 1.20
Mud weight = 1.7 sg	WOB = 5,000 kg	Pressure drill pipe = 100 bar
Steel weight = 7.85 sg	Block height = 30 m	Starting time = 0

Table 5: Initial input parameters case 2 simulator run

After adjusting the initial parameters, the simulation was again run twice, one time with noise and one time without noise. This generated a 3.57 m<sup>3</sup> influx for the simulation with noise, and a 3.42 m<sup>3</sup> for the simulation without noise, both obtained in the last drilling section as the simulator is programmed to do. Now, looking back at the first two simulations in case 1, the change in pump pressure was only noticeable in the plot of the last section without noise. This is because an influx

of small size has a relatively low impact on the pump pressure. However, studying the substantially larger influx generated in this case, the change in pump pressure is quite detectable both for the data with and without noise as seen in Figure 28.

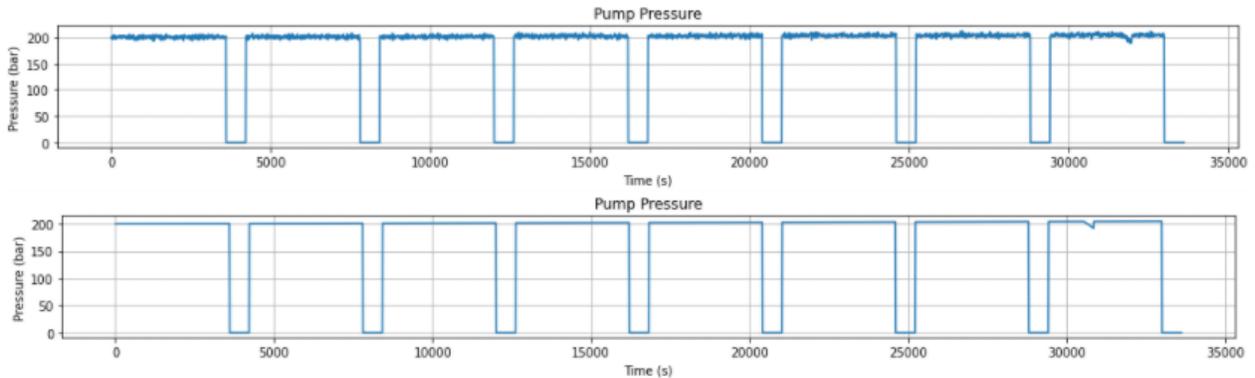


Figure 28: Comparison of pump pressure data with and without noise case 2

Looking solely at the last drilled section in Figure 29, it is easily seen that there is as a matter of fact a noticeable reduction in pump pressure for both simulations as expected by the calculations. The reduction by these influx sizes are around 12 bar for both simulations. However, adding more noise to the pump pressure data would make it even harder to detect, whereas for example increasing the standard deviation to around 10-15 bar would make the reduction in pump pressure with noise almost undetectable by the naked eye.

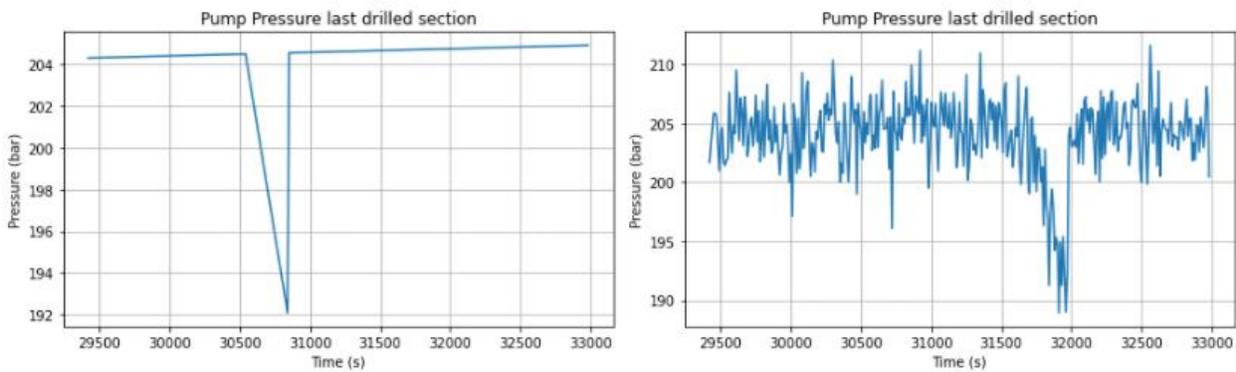


Figure 29: Pump pressure last drilled section with and without noise

Now, looking at the bottom hole pressure, this is expected to decrease when an influx is taken as previously stated. Studying the larger influx volume in this case as depicted in Figure 30 it is clearly

seen that the reduction in bottom hole pressure is noticeable in both cases, even with noise. This supports the fact that the drillers should be able to detect kicks at such a large size.

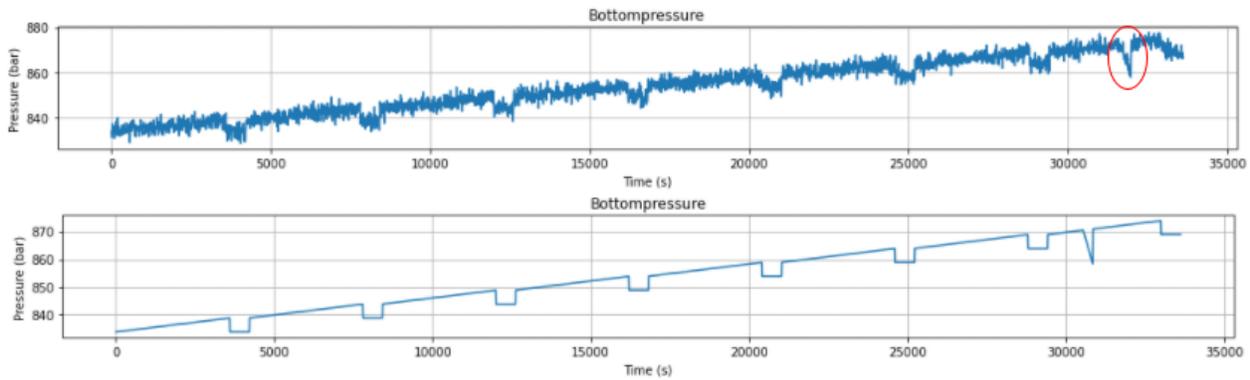


Figure 30: Comparison of bottom hole pressure data with and without noise case 2 (influx section marked with red circle)

The ECD data shows an impact no matter the size of the kick, as there is no noise in this data in the simulator as well as the plot shows changes down to a low change. This is clearly shown in Figure 31 below where a change from 1.71 to 1.69 sg is noticeable in the plot. As a matter of fact, the ECD will fall below the original mud density, whereas this reduction also will lead to a reduction in the bottom hole pressure increasing the possibility of generating a larger influx volume as the pressure falls even more below the pore pressure.

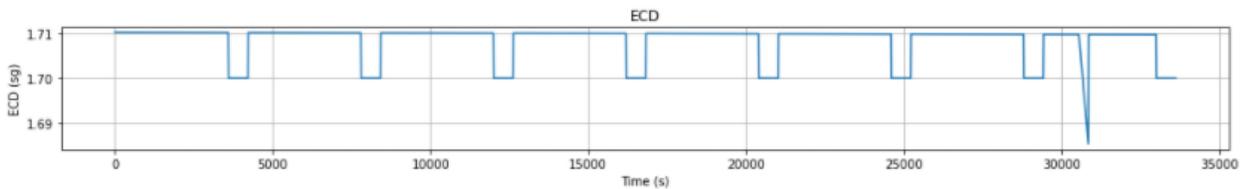


Figure 31: ECD data from simulation without noise case 2

Moreover, looking at the ROP data. This would not be affected by the size of the kick, as the increase is a preliminary warning signal as well as showing an increased value throughout the whole event duration. However, the ROP increase is likely to lie at somewhere above 20% no matter kick size, meaning the only change from case 1 to case 2 in this data would be that more connections are made and that the influx event duration is longer in case 2. The longer event

duration might in some cases with a lot of noise make it easier to detect a trend of increased values and distinguish this from regular noise. The results can be seen in Figure 32.

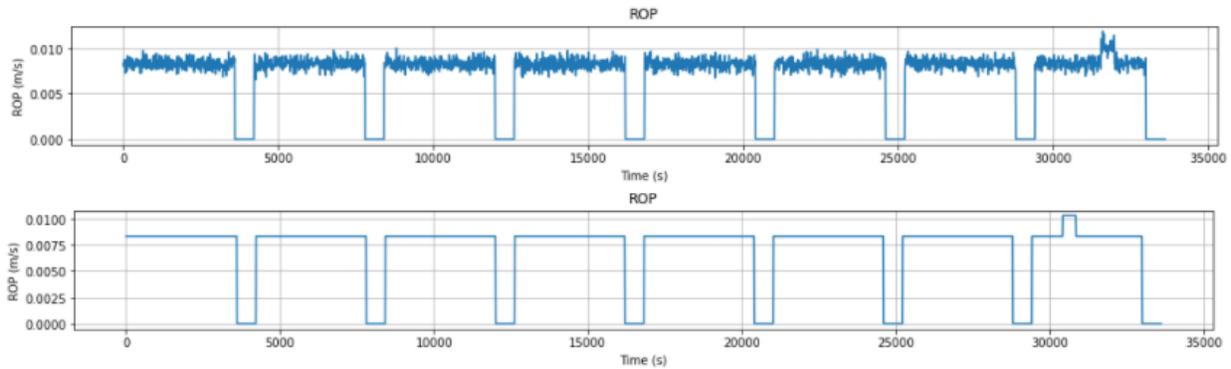


Figure 32: Comparison of ROP data with and without noise case 2

Now, studying the flowrate out with an extended event period, the flowrate impact will become smaller as the total volume is divided into an influx rate for each timestep throughout the whole event duration. However, a larger influx volume would give contradictive results, meaning increased flowrate out. Since both the event period is extended and the influx volume is larger, the flowrate out would remain more or less similar to case 1 even though the influx volume is larger. On the other hand, since the event duration is extended in case 2, the increase flowrate out can be easier seen as it is plotted over a significant amount of increased time period. This is shown in Figure 33 below.

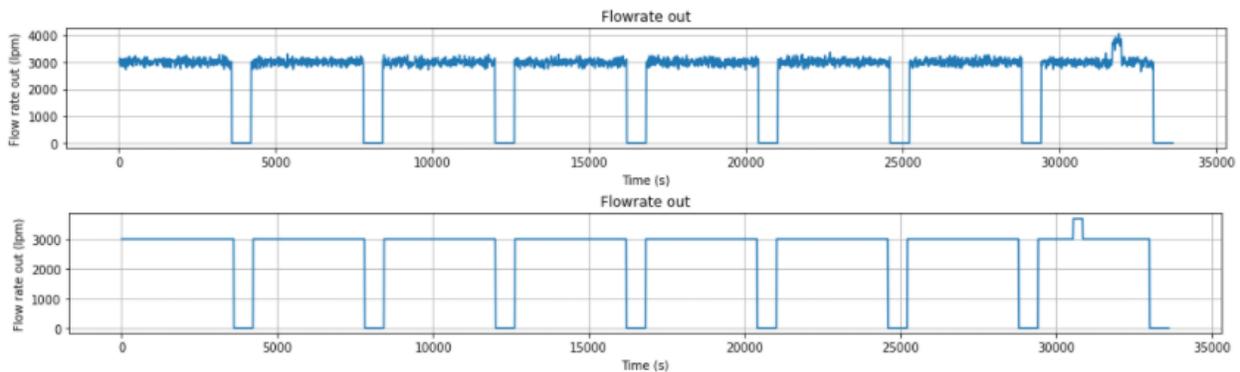


Figure 33: Comparison of flowrate out data with and without noise case 2

By increasing the number of stands drilled, the connections obtaining small influxes of gas should be shown in later connections. Figure 34 shows that connection gas is observed in connection 4, 5, and 6 for case 2, whereas it is shown for 1, 2, and 3 in case 1. This is because the influx is obtained in drilling section number 8 instead of 5 as in case 1.

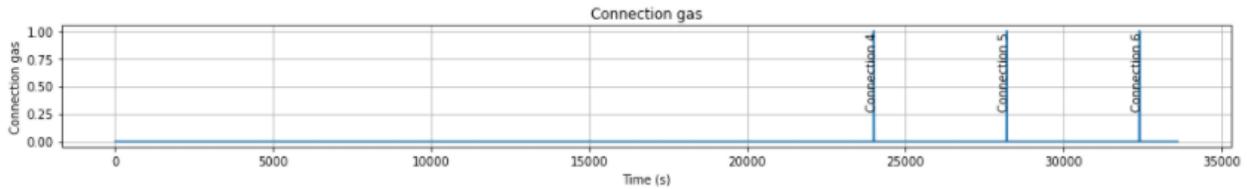


Figure 34: Connection gas in an extended simulated number of connections for case 2

The pit gain is still relatively easy to detect as a result of the influx. However, the pit gain is increased over three times, as the taken influx is over three times larger in case 2 than in case 1 for all simulations. In the case of increased noise in pit gain data, the increased value in case 2 would have a greater impact on the influx detection as pit gains lower than 1 m<sup>3</sup> easier could be misinterpreted as noise. Although, as both cases are presented here, the influx can be easily seen either way.

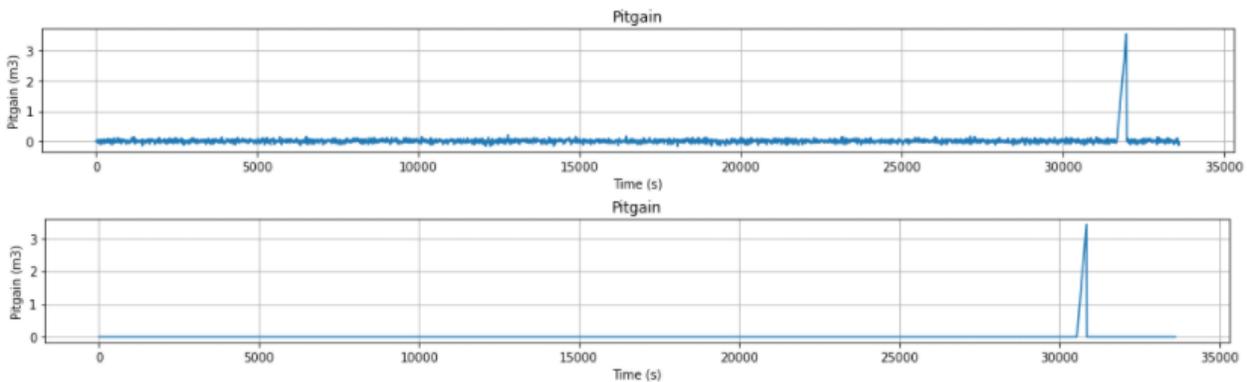


Figure 35: Comparison of pit gain data with and without noise case 2

Lastly, looking at the hook load. This was fairly undetectable in case 1 for both with and without noise. Now, considering a larger influx and a longer event period, it should be easier to detect hook load changes in case 2. In Figure 36 it is shown that the increased hook load (marked by a red circle) during influx is in fact easier to observe when obtaining a larger kick over a longer event

period. It is however not that easy to detect the influx in the hook load data with noise, as this easy can be misinterpreted as noise. The extended event period will on the other hand support an increased trend over a longer time, making it somewhat easier to differentiate from regular noise.

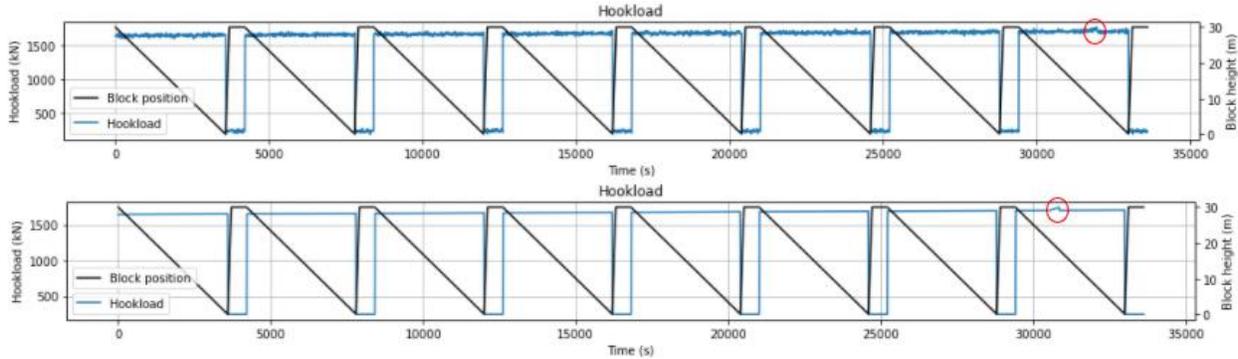


Figure 36: Comparison of hook load data with and without noise case 2

Now, since the influx event is in the last drilled section, a zoomed cut-out of this section will better present the increased hook load. The representation of the hook load data shown in Figure 37 makes it easier to detect that the hook load is in fact increased during the influx event. This is fairly easy to interpret on the hook load data without noise (to the left), but somewhat harder to interpret on the hook load data with noise (to the right). However, here it is seen that the extended event period supports an easier interpretation of the data to identify the influx. Otherwise it could be easier misinterpreted as noise, such as in the plot with noise in Figure 27.

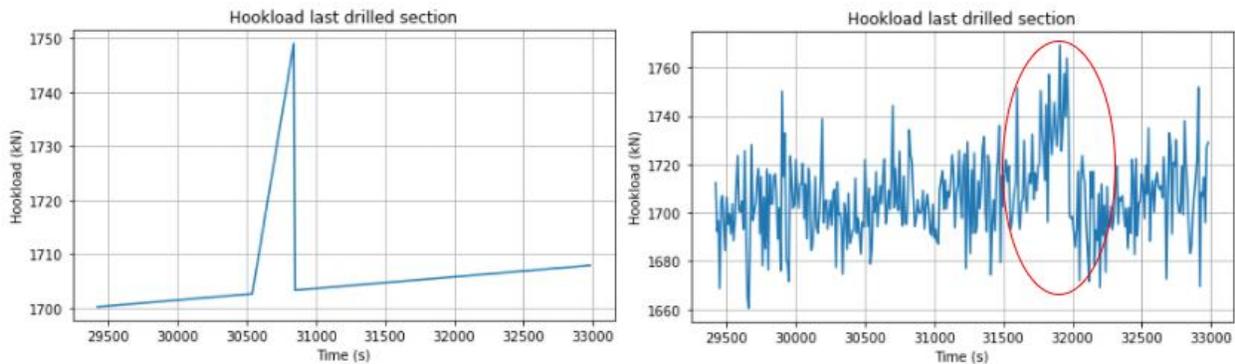


Figure 37: Hook load last drilled section with and without noise case 2 (influx section marked with red circle)

The presentation of data in case 2 versus case 1 shows that a larger influx and longer event period in fact does make it easier to detect an influx, as expected. However, not all the parameters are that

dependent on a larger influx over a longer period to be easier to detect. Here, the connection gas is something that is expected to occur no matter the incoming volume, meaning this is as easily seen either way. The drop in ECD as well as the gain in ROP for the influx is also easy to detect in both cases, where the ROP usually increase by 20% no matter the size of the incoming influx. On the other hand, it was easier to detect the impact of an influx on pump pressure, bottom hole pressure and hook load when considering a larger influx volume (case 2) compared to a smaller volume (case 1), due to the significant increase in influx volume and event period. It is again worth to state that the noise in these data does not necessarily represent real noise in data samples gathered from a well. The need for filtering and pre-processing is however deemed necessary in order for the machine learning models to make the correct predictions without being too affected by noise and outliers. As a result, real data can be filtered and pre-processed to reach an acceptable amount of noise as well as the noise in this simulator can be adjusted towards minimum amount of noise possible to reach in these processes. This will help generate large data sets of random influx data, in order to train, test and verify different machine learning models for application in real operating environments.

## 4. Discussion

As the world moves towards deeper waters in search for new resources, the drilling operations become more challenging. This increases the chance of obtaining a kick, thus the need for better kick detection tools are seen as an important factor to keep these operations safe and economically possible. The rapidly emerging interests in machine learning has initiated a series of experiments and studies for understanding how machine learning can be used for early kick prediction, as detection time is a crucial parameter in kick detection in order to mitigate the risk and impact. However, these machine learning models are dependent on sufficient training data in order to perform efficiently. Obtaining good training data can be hard and requires comprehensive pre-processing in order to be able to train the machine learning model, as real well data contains a substantial amount of noise. On the other hand, most simulated data today are lacking realistic noise and are made up of too many assumptions to generate training data that would suffice as realistic for training machine learning models without generating an unrealistic prediction accuracy. Consequently, the machine learning model would not perform to a high accuracy when applied to real working environments as the model would not be trained to deal with noise, resulting in making false and wrong predictions.

As previously stated, surface sensor equipment used in drilling operations are associated with a lot of noise during data gathering. Naturally, some sensors will provide better data gathering than others, as for example Coriolis flowmeters versus flow-paddle as described in section 2.1.3. The use of proper sensor equipment is therefore critical as the machine learning models are heavily dependent on the input data used in order to make accurate predictions. These measurements are also seen as the eye of the driller, as the driller is heavily dependent on observing changes in these parameters to interpret what is happening downhole. Another important factor to consider is what kind of drilling fluid that is used. As stated in section 2.1.4 it is important to consider how differently an influx of gas would behave in oil based drilling fluid and water based drilling fluid. Using an oil based drilling fluid would open up for the possibility of the influx to completely be dissolved into the drilling fluid, whereas the gas would liberate from the drilling fluid when reaching a certain pressure when traveling up the well. The dissolution of the gas in the drilling fluid would mean that the surface parameters would not get affected by the influx, resulting in the kick being hidden until reaching a point where it might be too late to perform early mitigating

actions in order to regain control of the well. Most machine learning models trained for kick detection does not take account for this difference, meaning the abnormalities in the data resulting from a kick in an oil based mud could be of such a small size that the machine learning model possibly would neglect it. This thesis is also focused on kick detection using water based drilling fluid, as the use of oil based drilling fluid would complicate the development of the simulator to a large extent. The implementation of using oil based drilling fluid would of course be of great interest in further work of the subject.

One of the main motivations for developing the simulator is to reduce the comprehensive pre-processing time when generating training and testing data for the machine learning models. The aim is to produce realistic-like data with an acceptable amount of noise in order to easily generate training and testing data for machine learning algorithms. It is however worth to mention that the simulator has some limitations, but these are discussed later in this section. Furthermore, pre-processing and cleaning of data accounts for over half the development time in developing machine learning models as stated in section 2.2.2. An effective development of machine learning models would therefore depend on the ability to generate data sets with an acceptable amount of noise and outliers. The importance of training data used is reflected by the comparison of the different machine learning models used for early kick detection in Table 2. Here, the same kind of machine learning algorithms show different results based on the input data used in the training and testing samples. A good example is the drastic accuracy loss in the KNN model, whereas change in training data made the prediction accuracy drop from 99.2% to 80.9%. This is a result of the amount of noise used in the data sample, whereas more noise led to a drastic loss of accuracy due to several false predictions due to outliers being interpreted as abnormalities. The machine learning model that seems to be most effective, as well as robust towards changes in data set is the ANN model. However, several results here are based on simulated data, meaning that there might be limitations towards amount of noise, resulting in an unrealistic increase in prediction accuracy.

Safety is of course a major factor when choosing to implement certain tools in drilling operations, but the economic aspects will also contribute as a vital factor. It is however hard to build a case of the economic aspects of early kick prevention as this is strictly preventive costs, meaning no costs of the consequences will be reported as lost costs. An interesting study to perform would be to compare for example 100 wells using preventive tools like early kick detection using machine

learning with 100 wells not using preventive tools, then see the difference in costs for these two categories. However, costs connected to minor or larger kicks are not isolated and reported publicly by operators, meaning it is hard to come up with a reasonable estimate of saved costs. A possibility would be to gather costs of non-productive time and gather information about how much non-productive time that is associated with influxes. It is however worth to mention that the non-productive costs would vary from well to well as different procedures performed at the wells would have varied costs, but more importantly the costs related to the lost production during that period which would vary enormously from reservoir to reservoir. This is on the other hand interesting data to research for a further work of the subject, whereas the OG21 project presents the importance of cost reduction using digitalization for the NCS as presented in section 2.3.1. Here, anomaly detection for well control are reported to have a potential cost reduction of NOK 3-4 billion per year for the NCS with a success probability of around 80%. Another cost reducing factor to consider is to completely remove other sensors such as logging while drilling tools and use solely the surface parameters to make calculations and interpret the drilling conditions downhole by using machine learning as an efficiency and safety tool. This implies that the research of machine learning used for kick detection is vital for the future of both the NCS and the oil and gas industry globally.

As stated, this thesis proposes a preliminary developed simulator that aims to calculate and produce data that are typically found in surface measurements when obtaining an influx, where the ultimate goal is to use this for generating realistic training sets for machine learning algorithms. It is however worth to mention that since this is in the preliminary stages of development, the simulator will be based on a series of assumptions, as presented in section 3.1. Some of these assumptions would be natural to include, such as real gas law applied to calculate gas density at given conditions, pump pressure will be adjusted for drilled depth to maintain pressure, and connection gas in some defined number of connection prior to the influx. However, other assumptions are made to simplify the simulator, which will result in some limitations. Here, assumptions such as no thermal driven volume expansion during connection, no expansion of kick, and single bubble approximation for the gas bubble should be avoided in further development of the simulator. The influx is assumed to be solely methane gas, which is the most likely case, but other fluids might occur as well. However, this is not deemed as a limitation in the simulator as the other gases would produce somewhat similar results, that will not affect the machine learning models ability to make accurate predictions. The fact that the well is assumed to be vertical, fairly small influx volumes are

considered, and that the influx will only occur in the last drilled section is just assumptions and will not limit the machine learning models ability to make accurate predictions. Furthermore, the assumptions that there is no thermal expansion during connection is something that should be researched for further development of the simulator. The thermal expansion will cause the drilling fluid to expand, thus giving increased pit gain volumes that may result in the machine learning model to assume this is a kick. However, machine learning models should be made to compare patterns between previous connection and the coming one to mitigate this problem as stated in section 2.1.2. The simulator does also assume that the influx will not expand, as the event period are designed only for the total flow in time of the influx. In real scenarios the influx will expand as it travels up the well due to reduced pressure as explained in 2.1.4. This will further affect the surface parameters, whereas an expanding kick would make larger and larger impacts on the already affected parameters. Since the simulator is made for machine learning models used for early kick detection, this should not be a case as the system should warn the driller before the kick is travelling up the well and expanding. Consequently, this could be implemented in the simulator for further development, but it is not deemed necessary due to the machine learning model's ability to identify kicks before this happens. In this case, the influx will only affect the parameters during the influx event period, it will not be circulated up the well. This means that the influx will only be situated around the BHA (assuming only small influxes). Furthermore, under the assumption of single bubble approximation and that only water based drilling fluid, the kick will completely displace the drilling fluid around the BHA. Here, the possibility of methane dissolving into the drilling fluid should be further researched (especially in oil based drilling fluid), as this would have a significant impact on the parameters by camouflaging the kick. Here, several simulators today are limited by assuming that the methane gas dissolves instantly into the drilling fluid, whereas this takes some time in real environments. It is worth to again state that these assumptions are only made to be able to develop this simulator as a preliminary simulator. By further developing this simulator, all these assumptions should be taken under consideration, aiming to minimize the amount of assumptions to make a more realistic simulation.

The functionality and flexibility of the simulator is presented in section 3.4.2 and 3.4.3. This shows that the simulator easily can take account for changes in initial parameters and adjust simulations accordingly. This is important in order to generate several different training sets for the machine learning models to strengthen the robustness of the kick detection system. Additionally, both cases

were simulated by some kind of fictional noise to show the effect noise has on the predictions. Gathering typical noise levels is fairly time consuming, as well as different equipment will be associated with different noise levels. However, this is deemed necessary as the machine learning models would obtain unrealistically high prediction accuracies without noise in the data sets. Hence, the models would make several false predictions when subjected to noise in real drilling data due to the noise and outliers being observed as abnormalities. Small influxes are fairly hard to detect when subjected to a substantial amount of noise, whereas they are easier seen at larger influx volumes. Here, the machine learning model's job is to detect these small influxes that are not easily seen by the driller in order to perform rapid mitigating actions.

Lastly, the proposed machine learning algorithms to test in further work of the subject is recommended to be a Long Short-Term Memory (LSTM) or an ANN model. The LSTM algorithms are usually applied when recurrent neural networks (a subset of ANN) suffer from short-term memory, i.e. when the data set is so long that the algorithms can't carry information from early time steps to a later one [65]. Here, the algorithm is structured with memory blocks, a set of recurring connected subnets. Each cell consists of three gates: input, output, and forget, where these perform write, read, and reset actions respectively. The LSTM work similar to a recurring neural network, however, where the memory blocks are substitutes for the hidden layers in a recurring neural network. The LSTM and the ANN are suggested due to their ability to take several input parameters and make correlations between them. In the case of early kick detection, it would therefore be of great value to include ROP and connection gas data as these provide preliminary signals for incoming kick. However, these should not be used to detect a kick alone, as they might give indications of kicks based on other influencing factors, making a false prediction. Combining these factors with for example an increased flowrate out, pit gain, and hook load, this could be used to flag kick events when detecting abnormalities in the data.

The simulator was designed for generating data to what is most likely going to be an ANN model developed by Exebenus. This model is intended for using surface parameters to enable early kick predictions. Consequently, they would be able to provide a tool that can reduce the need for using downhole sensors as well as mitigating risks of losing well control. As a result, the system would be very attractive for the industry looking at the economic aspects of the case.

## 5. Conclusion

In this thesis, a method for generating data sets for training and testing machine learning models was explored. Firstly, a literature study was conducted to create a foundation and understanding of how kicks are measured, how they affect the well, and which operations are associated with kicks. Additionally, economic consequences of kicks, and the current status on machine learning used for kick detection today was studied. Lastly, a simulator for random influx generation was developed.

Several machine learning models have been tested for early kick detection applications, but the prediction accuracy will vary vastly depending on the input data used. Many of the studies today include the use of simulated data that are subjected to a minimum amount of noise, meaning the machine learning models will provide unrealistically good prediction accuracies. Furthermore, OG21 suggest that the use of machine learning and digitalization is essential for the oil and gas industry on the NCS in order to survive more challenging times in the future. Here, they presented that the potential cost reductions using such technology for anomaly detection in order to keep well control was around NOK 3-4 billion per year, which again represents around 5-7% of the total operating costs per year of the NCS.

The simulator developed in the thesis showed great potential by generating data sets based on several physical and mathematical calculations for influx in a well. Additionally, the random generating element is important in order to generate enough different data to prevent overfitting of the machine learning models. Each new simulation will generate a completely new data set that is unique due to the random nature of the simulator. However, the simulator is currently limited by several assumptions, as well as lack of realistic noise levels in the drilling data. These are crucial elements for generating realistic data in order to train the machine learning models for accurate predictions in real operating environments (which is the ultimate goal). Noise will make the anomalies significantly harder to detect, which can clearly be seen in the two simulated cases presented in section 3.4.2 and 3.4.3. Here, it is clearly shown that noise in especially pump pressure, bottom hole pressure, and hook load data would make influx detection challenging as the impact blends in with the noise. However, a further development of the simulator can be of possible significance for training machine learning models in early kick detection, especially for ANN or LSTM models, based on the large amount of data it is able to generate in several surface parameters.

The implementations that are seen as necessary before achieving this goal is presented in section 5.1 below.

## **5.1 Further Work**

Although the simulator shows promising results, it is subjected to some limitations as previously stated. The further work proposed to meet these limitations are therefore presented in this section. Firstly, one of the most important implementations would be to add noise at a realistic level based on sensor equipment used at the rig. Here, a comprehensive data gathering of noise levels in different sensor equipment must be conducted. This should then be adjusted towards pre-processing needs before implementation in machine learning algorithms, in order to reduce the pre-processing needs substantially.

Other factors that needs to be considered for implementation in the simulator is thermal expansion of drilling fluid during connections, possibility of different lithologies while drilling and the corresponding impact on ROP, and gas dissolving into the drilling fluid, i.e. the use of oil based drilling fluid. These factors would all impact the generated data and then consequently the provided plots. The use of oil based drilling fluid and random influx generation will be more complex to implement since the gas dissolution process has to be modelled where probably kinematic process has to be included. This means that the dissolution process becomes time dependent. In these cases, kick detection might be too late to keep well control by mitigating actions, and a shut-in procedure needs to be performed quickly.

Lastly, the simulator data should be tested in LSTM or ANN algorithms, whereas several input parameters can be used. This should enable early kick detection by studying the relationships between flowrate out /pit gain increases, sudden ROP increases, and connection gases obtained. Here, the model should be able to make predictions based on parameters occurring at different times such as sudden ROP increase, connection gas, and flowrate out in order to foresee the incoming kick event. However, the mentioned further work necessary for the simulator should be implemented before applying these data in machine learning algorithms.

# References

- [1] M. E. Paté-Cornell, "Learning from the Piper Alpha Accident: A Postmortem Analysis of Technical and Organizational Factors," *Risk analysis*, vol. 13, no. 2, 1993.
- [2] T. Azwell and e. al., "Final Report on the Investigation of the Macando Well Blowout," Deepwater Horizon Study Group, 2011, March 1.
- [3] S. Unrau, P. Torriane, M. Hibbard, R. Smith, L. Olesen-Pason and J. Watson, "Machine Learning Algorithms Applied to Detection of Well Control Events," in *Paper SPE-188104-MS presented at the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, 2017, 24-27 April.
- [4] S. Muojeke, R. Venkatesan and F. Khan, "Supervised data-driven approach to early kick detection during drilling operation," *Journal of Petroleum Science and Engineering*, vol. 192, no. 107324, 2020.
- [5] J. Yang, T. Sun, Y. Zhao, A. T. Borujeni, H. Shi and H. Yang, "Advanced Real-time Gas Kick Detection Using Machine Learning Technology," in *Paper ISBN 978-1 880653 85-2 proceeding of the Twenty-ninth International Ocean and Polar Engineering Conference*, Honolulu, Hawaii, USA, 2019, 16-21 June.
- [6] X. Shi, T. Zhou, Q. Zhao, H. Jiand, L. Zhao, Y. Liu and G. Yang, "A New Method to Detect Influx and Loss During Drilling Based on Machine Learning," in *Paper IPTC-19489-MS presented at the International Petroleum Technology Conference*, Beijing, China, 2019, 26-28 March.
- [7] D. Fraser, R. Lindley, D. Moore and M. V. Staak, "Early Kick Detection Methods and Technologies," in *Paper SPE-170756-MS presented at the SPE Annual Technical conference and Exhibition*, Amsterdam, The Netherlands, 2014, 27-29 October.
- [8] E. D. Toskey, "Kick Detection at the Subsea Mudline," in *Paper OTC-25847-MS presented at the Offshore Technology Conference*, Houston, Texas, USA, 2015, 4-7 May.
- [9] R. B. Vik, *Well Control Manual*, Vols. ISBN 82-412-0530-9, Norway: Vett& Viten, 2001.
- [10] A. Lafond, F. Leblay, G. Roguin and M. Ringer, "Automated Influx and Loss Detection System Based on Advance Mud Flow Modeling," in *Paper SPE-195835-MS presented at the SPE Annual Technical Conference and Exhibition*, Calgary, Alberta, Canada, 2019, 30 September - 2 October.
- [11] L. A. Carlsen, G. Nygaard, J. E. Gravdal, M. Nikolaou and J. Schubert, "Performing the Dynamic Shut-In Procedure Because of a Kick Incident When Using Automatic Coordinated Control of Pump Rates and Choke-Valve Opening," in *Paper SPE/IADC 113693 presented at the 2008 SPE/IADC Managed Pressure Drilling and Underbalanced Operations and Exhibition*, Abu Dhabi, UAE, 2008, 28-29 January.

- [12] R. A. G. Luis, J. Bedoya, S. Cenberlitas, W. Bacon, O. Gabaldon and P. R. Brand, "Managed Pressure Drilling Influx Management: A Comprehensive Analysis of IME, MPD Operations Matrix, Dynamic Kick Tolerance Concepts, and Parameter Sensitivity," in *Paper SPE/IADC-200514-MS presented at the SPE/IADC Managed Pressure Drilling and Underbalanced Operations Conference and Exhibition*, Virtual event, 2020, 29-30 October.
- [13] J. E. Chirinos, J. R. Smith and D. A. Bourgoyne, "Alternative Shut-In and Pump Start-Up Procedures for Kicks Taken During MPD Operations," in *Paper IADC/SPE 143094 presented at the IADC/SPE Managed Pressure Drilling and Underbalanced Operations Conference*, Denver, Colorado, USA, 2011, 5-6 April.
- [14] K. Kinik, F. Gumus and N. Osayande, "A Case Study: First Field Application of Fully Automated Kick Detection and Control by MPD System in Western Canada," in *Paper SPE/IADC-168948-MS presented at the SPE/IADC Managed Pressure Drilling and Underbalanced Operations Conference and Exhibition*, Madrid, Spain, 2014, 8-9 April.
- [15] P. Vieira, M. Arnone, F. Torres and F. Barragan, "Roles of Managed Pressure Drilling Technique in Kick Detection and Wellcontrol - The Beginning of the New Conventional Drilling Way," in *Paper SPE/IADC 124664 presented at the SPE/IADC Middle East Drilling Technology Conference & Exhibition*, Manama, Bahrain, 2009, 26-28 October.
- [16] J. D. Brakel, B. A. Tarr, W. Cox, F. Jørgensen and H. Straume, "SMART Kick Detection: First Step on the Well-Control Automation Journey," in *Paper SPE 173052 presented at the SPE/IADC Drilling Conference and Exhibition*, London, UK, 2015, 17-19 March.
- [17] R. L. Rudolf and P. V. R. Suryanarayana, "Kicks Caused by Tripping-In the Hole on Deep, High Temperature Wells," in *Paper SPE 38055 presented at the 1997 SPE Asia Pacific Oil and Gas Conference*, Kuala Lumpur, Malaysia, 1997, 14-16 April.
- [18] J.-M. Godhavn, B. Olorunju, D. Gorski, M. Kvernland, M. Sant'Ana, O. M. Aamo and S. Sangesland, "Significant Surge and Swab Offshore Brazil Induced by Rig Heave During Drillpipe Connections," in *Paper SPE 200518 presented at the SPE/IADC Managed Pressure Drilling and Underbalanced Operations Conference and Exhibition*, Virtual event, 2020, 12 November.
- [19] B. A. Tarr, D. W. Ladendorf, D. Sanchez and G. M. Milner, "Next-Generation Kick Detection During Connections: Influx Detection at Pumps Stop (IDAPS) Software," in *Paper SPE 178821 presented at the IADC/SPE Drilling Conference and Exhibition*, Fort Worth, Texas, USA, 2016, 1-3 March.
- [20] S. C. H. Geekiyange, A. Ambrus and D. Sui, "FEATURE SELECTION FOR KICK DETECTION WITH MACHINE LEARNING USING LABORATORY DATA," in *Paper OMAE2019-95496 proceedings of the ASME 2019 38th International Conference on Ocean, Offshore and Arctic Engineering*, Glasgow, Scotland, UK, 2019, 09-14 June.
- [21] E. Cayeux, "Measurement of the Flowrate Out of a Well for Conventional Drilling Operations," in *Paper IADC/SPE-199661-MS presented at the IADC/SPE International Drilling Conference and Exhibition*, Galveston, Texas, USA, 2020, 3-5 March.

- [22] H. Linga, A. Torsvik and A. Saasen, "Kick Detection Capability of Oil-Based Muds in Well Control Situations," in *Paper SPE-180039-MS presented at the SPE Bergen One Day Seminar*, Bergen, Norway, 2016, 20 April.
- [23] P. L. O'Bryan and A. T. B. Jr., "Swelling of Oil-Base Drilling Fluids Due to Dissolved Gas," in *Paper SPE 16676 presented at the 62nd Annual Technical conference and Exhibition of the Society of Petroleum Engineers*, Dallas, Texas, USA, 1987, 27-30 September.
- [24] N. D. Bradley, E. Low, B. Aas, R. Rommetveit and H. F. Larsen, "Gas Diffusion- Its Impact on a Horizontal HPHT Well," in *Paper SPE 77474 presented at the SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, 2002, 29 September - 2 October.
- [25] D. Gomes, M. S. Nilsen, J. Frøyen, K. S. Bjørkevoll, A. C. V. M. Lage, K. K. Fjelde and D. Sui, "A Transient Flow Model for Investigating Parameters Affecting Kick Behavior in OBM for HPHT Wells and Backpressure MPD systems," in *Paper OMAE2018-77547 proceedings of the ASME 2017 37th International Conference on Ocean, Offshore and Arctic Engineering*, Madrid, Spain, 2018, 17-22 June.
- [26] P. L. O'Bryan, A. T. B. Jr., T. G. Monger and D. P. Kocso, "An Experimental Study of Gas Solubility in Oil-Based Drilling Fluids," in *Paper SPE 15414 presented at the 1986 SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 1988, 5 October.
- [27] H. Linga, K. S. Bjørkevoll, J. O. Skogestad and A. Saasen, "Gas Influx into Drilling Fluids During Flow Check Operations as Affected by Gas Absorption Characteristics of the Drilling Fluid," in *Paper SPE/IADC-184686-MS presented at the SPE/IADC Drilling Conference and Exhibition*, Hague, The Netherlands, 2017, 14-16 March.
- [28] R. Rommetveit, K. K. Fjelde, B. Aas, N. F. Day and E. Low, "HPHT Well Control; An Integrated Approach," in *Paper OTC 15322 presented at the 2003 Offshore Technology Conference*, Houston, Texas, USA, 2003, 5-8 May.
- [29] J. O. Skogestad, K. S. Bjørkevoll, J. Frøyen, H. Linga, E. Lenning and S. T. Håvardstein, "Well Control Incident in the North Sea as Interpreted with Advanced Gas Influx Modeling," in *Paper SPE/IADC-194145-MS presented at the SPE/IADC Drilling International Conference and Exhibition*, The Hague, The Netherlands, 2019, 5-7 March.
- [30] A. McDonald, "DATA QUALITY CONSIDERATIONS FOR PETROPHYSICAL MACHINE LEARNING MODELS," in *Paper SPWLA-2021-0036 presented at the SPWLA 62nd Annual Logging Symposium*, Virtual Event, 2021, 17-20 May.
- [31] A. U. Osarogiagbon, F. Khan and R. Venkatesan, "Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations," *Process Safety and Environmental Protection* 147, p. 367–384, 2020, 18 September.
- [32] T. A. Olukoga and Y. Feng, "Practical Machine-Learning Applications in Well-Drilling Operations," *SPE Drilling & Completion*, no. SPE 205480, 2020, 27 December.

- [33] V. Elichev, A. Bilogan, K. Litvinenko, R. Khabibullin, A. Alferov and A. Vodopyan, "Understanding Well Events with Machine Learning," in *Paper SPE-196861-MS presented at the SPE Russian Petroleum Technology Conference*, Moscow, Russia, 2019, 22-24 October.
- [34] N. Askham, D. Cook, M. Doyle, H. Fereday, M. Gibson, U. Landbeck, R. Lee, C. Maynard, G. Palmer and J. Schwarzenbach, "THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT; Defining Data Quality Dimensions," DAMA UK Working Group, UK, 2013.
- [35] M. C. Storey, "Demystifying Log Quality Control," in *Paper SPE-182313-MS presented at the SPE Asia Pacific Oil & Gas Conference and Exhibition*, Perth, Australia, 2016, 25-27 October.
- [36] C. Xu, S. Misra, P. Srinivasan and S. Ma, "When Petrophysics Meets Big Data: What can Machine Do?," in *Paper SPE-195068-MS presented at the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2019, 18-21 March.
- [37] S. A. Gharbi, M. Ahmed and S. Elkatatny, "Use Metaheuristics to Improve the Quality of Drilling Real-Time Data for Advance Artificial Intelligent and Machine Learning Modeling. Case Study: Cleanse Hook-Load Real-Time Data," in *Paper SPE-192810-MS presented at the Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, 2018, 12-15 November.
- [38] C. I. Noshi and J. J. Schubert, "The Role of Machine Learning in Drilling Operations; A Review," in *Paper SPE-191823-18ERM-MS presented at the SPE Eastern Regional Meeting*, Pittsburgh, Pennsylvania, USA, 2018, 7-11 October.
- [39] R. Alouhali, M. Aljubran, S. Gharbi and A. Al-yami, "Drilling Through Data: Automated Kick Detection Using Data Mining," in *Paper SPE-193687-MS presented at the SPE International Heavy Oil Conference and Exhibition*, Kuwait City, Kuwait, 2018, 10-12 December.
- [40] X. Hou, J. Yang, Q. Yin, L. Chen, B. Cao, J. Xu, L. Meng, Y. Zhang, D. Xu and X. Zhao, "Automatic Gas Influxes Detection in Offshore Drilling Based on Machine Learning Technology," in *Paper SPE-198534-MS presented at the SPE Gas & Oil Technology Showcase and Conference*, Dubai, UAE, 2019, 21-23 October.
- [41] M. Aghito and K. S. Bjørkevoll, "Hybrid Approach for Drilling Automation," in *Paper SPE-200734-MS presented at the SPE Norway Subsurface conference*, Virtual event (originally Bergen, Norway), 2020, 2-3 November.
- [42] Q. Yin, J. Yang, M. Tyagi, X. Zhou, X. Hou, N. Wang, G. Tong and B. Cao, "Machine Learning for Deepwater Drilling: Gas-Kick-Alarm Classification Using Pilot-Scale Rig Data with Combined Surface-Riser-Downhole Monitoring," *SPE Journal*, no. SPE 205365, 2021, 19 January.
- [43] A. Ahmed, S. Elkatatny, A. Ali and A. Abdurraheem, "Prediction of Lost Circulation Zones Using Artificial Neural Network and Functional Network," in *Paper SPE-203268-MS presented at the Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, 2020, 9-12 November.
- [44] Z. Li, M. chen, Y. Jin, Y. Lu, H. Wang, Z. Geng and S. Wei, "Study on Intelligent Prediction for Risk Level of Lost Circulation While Drilling Based on Machine Learning," in *Paper ARMA 18-105 presented at*

*the 52nd US Rock Mechanics/Geomechanics Symposium*, Seattle, Washington, USA, 2018, 17-20 June.

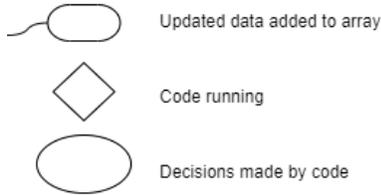
- [45] A. A. S., A. A. Mahmoud, S. Elkatatny, M. Mahmoud and A. Abdurraheem, "Prediction of Pore and Fracture Pressures Using Support Vector Machine," in *Paper IPTC-19523-MS presented at the International Petroleum Technology Conference*, Beijing, China, 2019, 26-28 March.
- [46] e. a. Gunnar H. Lille, "Machine Learning in the Petroleum Industry," OG21, 2020, 3 November.
- [47] H. P. Ellingsen, S. Angelsen, O. K. Sollie, A. Suyuthi, R. Mistry, P. Myrseth and O. G. Tveiten, "OG21 - Study on Machine Learning in the Norwegian petroleum industry," DNV GL, Høvik, Norway, 2020.
- [48] M. M. H. M. Hashim, M. H. Yusoff, M. F. Arriffin, A. Mohamad, D. Gomes, M. Jose and T. E. T. Bidin, "Utilizing Artificial Neural Network for Real-Time Prediction of Differential Sticking Symptoms," in *Paper IPTC-21221-MS presented at the International Petroleum Technology Conference*, Virtual event, 2021, 23 March - 1 April.
- [49] U. Arshad, B. Jain, M. Ramzan, W. Alward, L. Diaz, I. Hasan, A. Aliyev and C. Riji., "Engineered Solution to Reduce the Impact of Lost Circulation During Drilling and Cementing in Rumaila Field, Iraq," in *Paper IPTC-18245-MS presented at the International Petroleum Technology Conference*, Doha, Qatar, 2015, 6-9 December.
- [50] H. H. Alkinani, A. T. T. Al-Hameedi and S. Dunn-Norman, "Miniizing Lost Circulation Non-Productive Time Using Expected Monetary Value and Decision Tree Analysis," in *Paper SPE-200844-MS presented at the SPE Western Regional Meeting*, Virtual event (Originally scheduled to be held in Bakersfield, California, USA), 2020, 20-22 April.
- [51] C. A. Willemse and P. H. A. J. M. v. Gelder, "Analysis of the Deepwater Horizon Accident in Relation to Arctic Waters," in *Proceedings of the Twent-first (2011) International Offshore and Polar Engineering Conference*, Maui, Hawaii, USA, 2011, 19-24 June.
- [52] S. D. Mohagegh, "Contribution of Artificial Intelligence and Machine Learning in U.S. DOE's Efforts During the Aftermath of Deepwater Horizon," in *Paper SPE-191613-MS presented at the 2018 SPE Annual Technical Conference and Exhibitoin*, Dallas, Texas, USA, 2018, 24-26 September.
- [53] Y. G. Lee, X. Garza-Gomez and R. M. Lee, "Ultimate Costs of the Disaster: Seven Years After the Deepwater Horizon Oil Spill," *The Journal of corporate accounting & finance*, vol. 29, pp. 69-79, 2018.
- [54] D. Gomes, K. K. Fjelde, K. S. Bjørkevoll and J. Frøyen, "Gas suspension effects in riser unloading and appropriate modeling approaches," in *Paper OMAE2020-18049 presented at the ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering*, Fort Lauderdale, Florida, USA, 2020, June 28 - July 3.
- [55] M. Zamora, "Operation And Development Of A Well-Control Calculator Module," in *Paper SPE-10385-MS presented at the SPE Offshore Europe*, Houston, Texas, USA, 1981, 15-18 September.

- [56] A. D. Purkayastha, R. Rana, R. Talreja and N. Rajaiah, "Drilling Event Chart: A Kick Prevention Tool," in *Paper SPE-198660-MS presented at the SPE Gas & Oil Technology Showcase and Conference*, Dubai, UAE, 2019, 21 - 23 October.
- [57] Y. Mao and P. Zhang, "An Automated Kick Alarm System Based on Statistical Analysis of Real-Time Drilling Data," in *Paper SPE-197275-MS presented at the Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, 2019, 11-14 November.
- [58] K. Gjerstad, R. Bergerud and S. T. Thorsen, "Exploiting the Full Potential in Automated Drilling Control by Increased Data Exchange and Multi Disciplinary Collaboration," in *Paper SPE-201763-MS presented at the SPE Annual Technical Conference & Exhibition*, Denver, Colorado, USA, 2020, 26-29 October.
- [59] M. A. Ahmed, O. A. Hegab and A. Sabry, "Early detection enhancement of the kick and near-balance drilling using mud logging warning sign," *Elsevier B. V.*, 2015.
- [60] E. Cayeux and B. Daireaux, "Precise Gain and Loss Detection Using a Transient Hydraulic Model of the Return Flow to the Pit," in *Paper SPE/IADC 166801 presented at the SPE/IADC Middle East Drilling Technology Conference and Exhibition*, Dubai, UAE, 2013, 7-9 October.
- [61] Z. Yuan, D. Morrell, A. G. Mayans, Y. H. Adariani and M. Bogan, "Differentiate Drilling Fluid Thermal Expansion, Wellbore Ballooning and Real Kick during Flow Check with an Innovative Combination of Transient Simulation and Pumps off Annular Pressure While Drilling," in *Paper IADC/SPE-178835-MS presented at the IADC/SPE Drilling Conference and Exhibition*, Fort Worth, Texas, USA, 2016, 1-3 March.
- [62] S. H. A. Gharbi, F. S. A. Sanie and M. R. A. Zayer, "Automated Real time Data Cleansing and Summarization; Case Study on Drilling Hook load Real time Data," in *Paper SPE-176755-MS presented at the SPE Middle East Intelligent Oil & Gas Conference & Exhibition*, Abu Dhabi, UAE, 2015, 15-16 September.
- [63] E. Cayeux, H. J. Skadsem and R. Kluge, "Accuracy and Correction of Hook Load Measurements During Drilling Operations," in *Paper SPE/IADC-173035-MS presented at the SPE/IADC Drilling Conference and Exhibition*, London, United Kingdom, 2015, 17-19 March.
- [64] B. S. Aadnoy and E. Kaarstad, "Theory and Application of Buoyancy in Wells," in *Paper IADC/SPE 101795 presented at the Asia Pacific Drilling Technology Conference and Exhibition*, Bangkok, Thailand, 2006, 13-15 November.
- [65] L. F. Santos, M. Gattass, A. C. Silva, F. Miranda, C. Siedschlag and R. Ribeiro, "Natural gas detection in onshore data using transfer learning from a LSTM pre-trained with offshore data," in *Paper SEG-2020-3426303 presented at the SEG International Exposition and Annual Meeting*, Virtual event, 2020, 11-16 October.

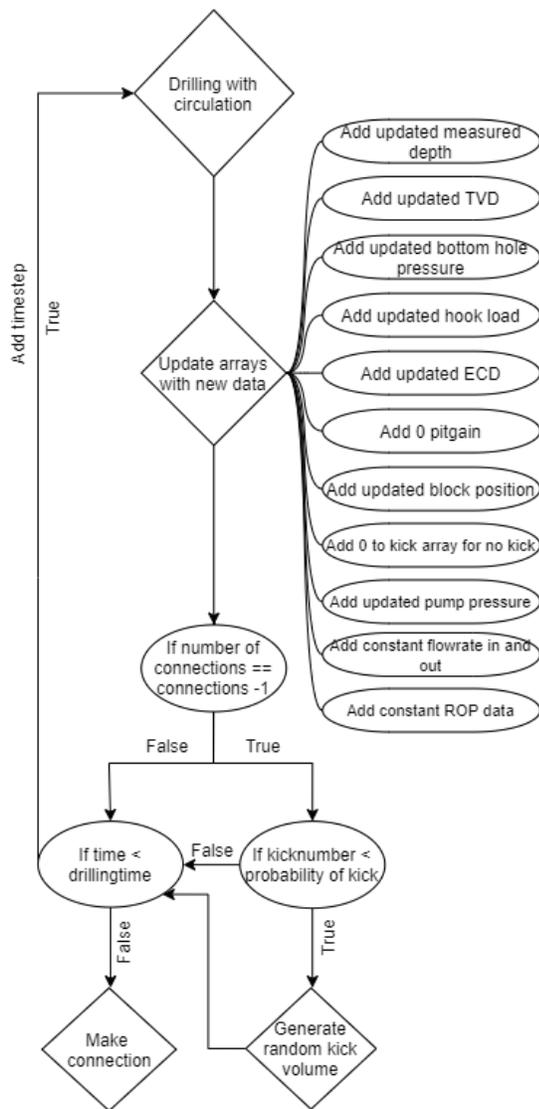


## A.2) Structure of sub-systems: Drilling and Drilling with Kick

### Definition of figures:



### Sub-system: Drilling

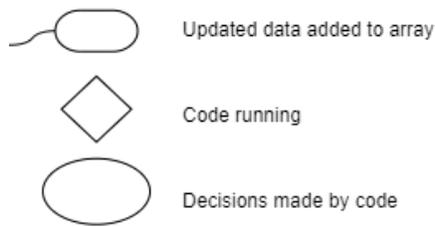


### Sub-system: Drilling with Kick

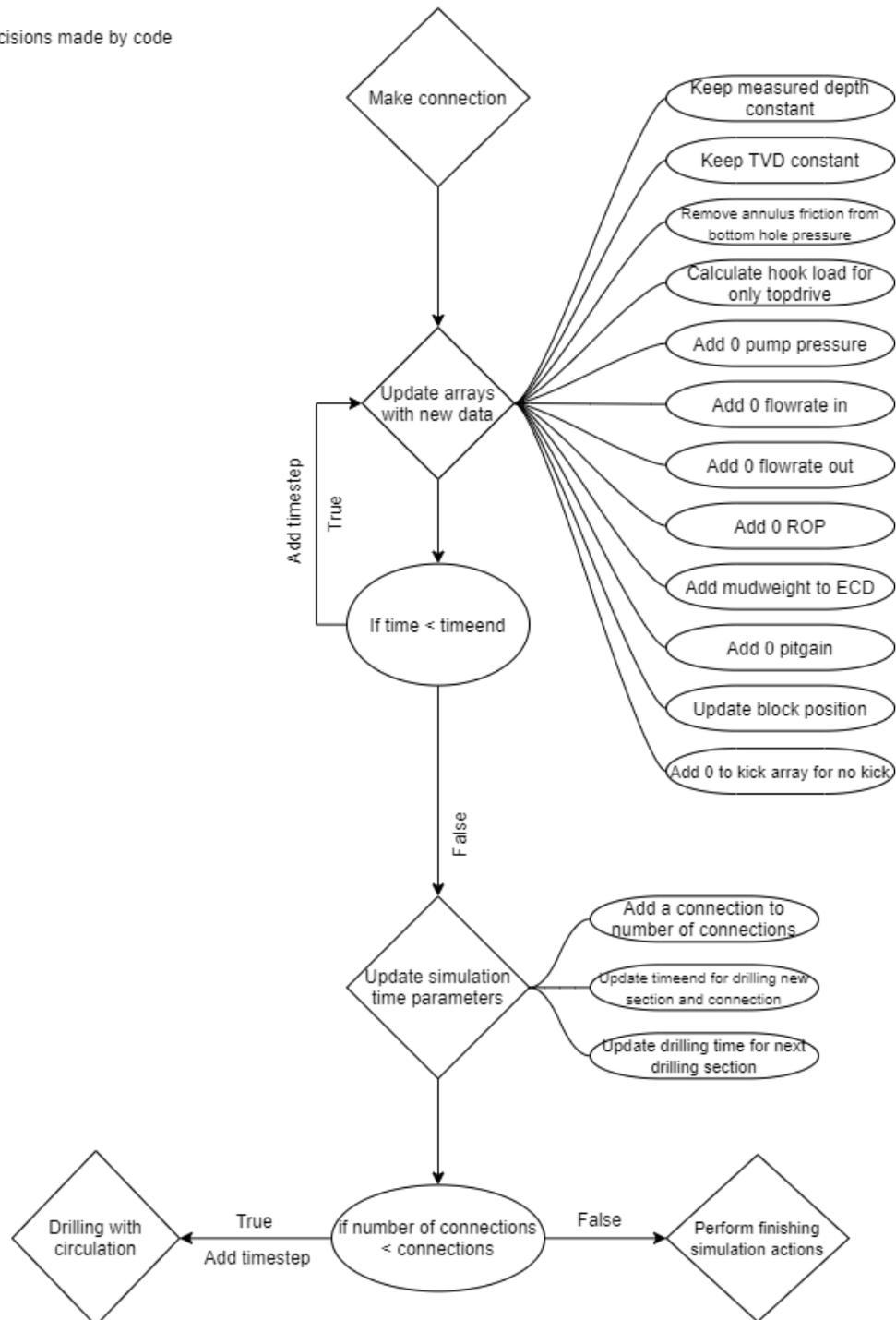


### A.3) Structure of sub-system: Connection

#### Definition of figures:



### Sub-system: Connection



## B.1) Simulator base-code:

```
"Importing necessary Libraries"  
import random  
from numpy import random  
import matplotlib.pyplot as plt  
import numpy as np
```

```
"Assuming only methane gas in kick and defining functions to calculate z-factor, density of gas  
and temperature"
```

```
def Zgas(gg,pp,tt):  
    """  
    Parameters:  
    -----  
    gg - gas gravity  
    pp - pressure Pascal  
    tt - temperature Kelvin  
    -----  
    """  
    p = pp/100000*14.5 # Converting pressure to psi  
    t = (tt*9/5)-459.67 # Converting temperature to Fahrenheit  
  
    A1 = 0.31506237  
    A2 = -1.0467099  
    A3 = -0.57832729  
    A4 = 0.53530771  
    A5 = -0.61232032  
    A6 = -0.10488813  
    A7 = 0.68157001  
    A8 = 0.68446549  
  
    ppc = 702.5-50*gg # Critical pressure [psi]  
    tpc = 167+316.67*gg # Critical temperature [Rankine]  
  
    ppr = p/ppc # Reduced pressure  
    tpr = (t+459.67)/tpc # Reduced temperature  
  
    Z = 1  
    error = 999  
    iteration = 0  
  
    while(error>0.001):  
        iteration += 1  
        if(iteration>100):  
            break  
        ropr = 0.27*ppr/Z/tpr  
        Z1 = 1+(A1+A2/tpr+A3/tpr/tpr/tpr)*ropr  
        Z1 = Z1+(A4+A5/tpr)*ropr*ropr  
        Z1 = Z1+(A5*A6*ropr*ropr*ropr*ropr*ropr)/tpr  
        Z1 = Z1+(A7*ropr*ropr/tpr/tpr/tpr)*(1+A8*ropr*ropr)*np.exp(-A8*ropr*ropr)  
        error = 2*abs((Z-Z1)/(Z+Z1))  
        Z = (Z1+Z)/2  
        zfactor = Z  
    return zfactor
```

```

def rhog(pressure,temp):
    """
    Parameters:
    -----
    pressure - pressure [Pa]
    temp     - temperature [K]
    -----

    """

    M      = 16.04 # Molar mass of methane g/mol
    Mair   = 28.96 # Molar mass air g/mol
    R      = 8.314 # Universal gas constant J/(mol*K)

    gamma = M/Mair

    z      = Zgas(gamma,pressure,temp)
    rhog   = M/(R*temp)*pressure/z      # g/m3
    rhog   = rhog/1000000                # sg
    return rhog

def temp_calc(drilleddepth):
    """
    Parameters:
    -----
    drilleddepth - depth at which the temperature should be calculated [m]
    -----

    """

    geo_gradient = 0.04 # Celsius/meter
    T_seabot    = 4    # Celsius
    T_bottom    = T_seabot + (drilleddepth-100)*geo_gradient
    return T_bottom

"Following parameters are initial parameters that can be changed for a wanted case"
timestart      = 0.0      # seconds (s)
timestep       = 10      # s
standlength    = 30      # meter (m)
rop            = 30/(60*60) # m/s Assumed 30 meter per hour
connectiontime = 10*60   # s
drilleddepthstart = 4970 # m, Assumed vertical well
pumppressurestart = 200  # bar, assumed at flowrate of 3000 lpm and at depth 4970 m
annulusfriction = 5      # Assume 5 bar pressure friction in annulus for 3000 lpm in 12
1/4 inch section
flowin         = 3000    # lpm
mudweight      = 1.7     # sg
steelweight    = 7.85    # sg
do_BHA         = 8.5     # inches (in) - outer diameter of BHA
di_BHA         = 3       # in - inner diameter of BHA
di_ann         = 12.25   # in - inner diameter of annulus
do_DP          = 5       # in - outer diameter of drill pipe
weight_DP      = 33.57   # kg/m - weight per meter drill pipe
weight_BHA     = 251.59  # kg/m - weight per meter BHA
length_BHA     = 100     # m
topdrive       = 25000   # kg
wob            = 5000    # kg
block_height   = 30      # m
block_liftime  = 2*60    # s (time taken to lift block up at connection)
connections    = 5       # Number of connections for simulation

```

```

probkick          = 0.002      # Probability of kick at fraction 0-1
eventduration    = 60         # s
kicksize_min     = 0.05       # m3 - minimum size of kick generated by uniform random
distribution
kicksize_max     = 1         # m3 - maximum size of kick generated by uniform random
distribution
rop_inc_min      = 1.20       # minimum percentage increase in rop (1.20 give 20% increase)
generated in random distribution
rop_inc_max      = 1.25       # maximum percentage increase in rop (1.25 give 25% increase)
generated in random distribution
pressure_DP      = 100        # bar -pressure in drillpipe, used to calculate additional pump
pressure needed when drilling
noise            = True       # Set to True if noise is to be included and False if not

"Calculating parameters based on initial parameter input"
drillpipelength = drilleddepthstart - length_BHA                # m
bouyancy_factor = 1 - mudweight/steelweight                    # Bouyancy factor for hook
Load calculations
annular_capacity = np.pi/4*((di_ann*0.0254)**2-(do_BHA*0.0254)**2) # m2 - annular capacity
around BHA
annular_cap_DP   = np.pi/4*((di_ann*0.0254)**2-(do_DP*0.0254)**2) # m2 - annular capacity
around drillpipe
block_pos        = block_height                                # m - initial position of
block
flowout          = flowin
pumppressure_gain = pressure_DP/drillpipelength                # bar/m - pressure gained
per meter drilled

"Defining parameters to be used for kick event calculations"
timeevent        = 0      # Start time of kick event
triggevent       = False # Setting kick event to false when no kick is obtained
kicksize         = 0      # m3
inflowvolume     = 0      # m3
number_of_connections = 0 # setting a variable for counting connections during drilling

"Parameters defined to control the length of segments in simulation"
section_drillingtime = standlength/rop # Drilling time per drilled section
drillingtime        = section_drillingtime # Setting a starting point for the drilling time,
stacking for each connection
timeend             = drillingtime+connectiontime # Setting one drilling iteration to equal drilling
of section and connection
time                = timestart # Setting initial time to defined starting time

"List to be used for storage of data for plotting parameters vs time."
timedata           = [] # timedata during the simulation
sppdata           = [] # pumppressuredata
flowrateoutdata   = []
flowrateindata    = []
measureddeptdata  = [] # Measured depth
tvddata           = [] # True vertical depth
bottompressure    = [] # Bottom hole pressure
ecddata           = []
ropdata           = [] # Rate of penetration
hookloaddata      = [] # Hookload calculated in kN
kick_array        = [] # storing 0 if no kick and 1 if kick
kick_det_times    = [] # To store at what times kick occur and stops
connection_array  = [] # connection times
pitgain           = [] # pit gain during kick event
influx_delays     = [] # possible gas influxes during connection
block_pos_array   = [] # position of block/topdrive (m)

```

```

"Filling in the first elements in lists at time = 0."
timedata.append(timestart)
sppdata.append(pumppressurestart)
flowrateoutdata.append(flowout )
flowrateindata.append(flowin)
measureddeptdata.append(drilleddepthstart)
tvddata.append(drilleddepthstart)
bottompressure.append(mudweight*0.0981*drilleddepthstart+annulusfriction)
hookloaddata.append(((topdrive+bouyancy_factor*(weight_DP*drillpipelength+\
+weight_BHA*length_BHA)-wob)*9.81)/1000)

ropdata.append(rop)
ecddata.append((mudweight+annulusfriction/(drilleddepthstart*0.0981)))
pitgain.append(0)
kick_array.append(0)
block_pos_array.append(block_height)

"Running simulation for decided number of connections"
while number_of_connections < connections:

    "Running a session of drilling and connection, before adding on new drilling section and connection"
    while time<=timeend:

        time += timestep      # Moving one timestep forward in time
        timedata.append(time) # Storing time values into an array of times
        if noise == True:
            "Creating simulated real-like noise in the measuring equipment"
            flowrateout_noise = random.normal(0,101.11) # Lmp
            hookload_noise    = random.normal(0,15.42)  # kN
            pump_noise        = random.normal(0,2.5)   # bar
            botpres_noise     = random.normal(0,2)     # bar
            ROP_noise         = random.normal(0,0.0005) # m/s
            pitgain_noise     = random.normal(0,0.05)  # m3

        else:
            "Setting all noise to 0 if noise is not to be used"
            flowrateout_noise = 0 # Lmp
            hookload_noise    = 0 # kN
            pump_noise        = 0 # bar
            botpres_noise     = 0 # bar
            ROP_noise         = 0 # m/s
            pitgain_noise     = 0 # m3

        "Drilling with circulation and without influx/kick"
        if (time<drillingtime and triggevent == False):
            """
            Appending data to the measured depth, true vertical depth (TVD), bottom hole pressure (BHP),
            hookload, ECD, pitgain and block position arrays.
            Calculations:
            -----
            * drilleddepthstart+rop*(time-number_of_connections*connectiontime) = drilleddepth
            based on time spent drilling
            * block_height/(section_drillingtime/timestep) = block movement based on equal
            steps for drilling duration
            -----
            """

```

```

"Adding data to arrays as drilling progresses"
measureddeptdata.append(drilleddepthstart+rop*(time-number_of_connections*
connectiontime))
tvddata.append(drilleddepthstart+rop*(time-number_of_connections*connectiontime))
bottompressure.append(mudweight*0.0981*(drilleddepthstart+\
+rop*(time-number_of_connections*connectiontime))+\
+annulusfriction+botpres_noise)
hookloaddata.append(((topdrive+\
+bouyancy_factor*(weight_DP*(drillpipelength+\
+rop*(time-number_of_connections*connectiontime))+\
+weight_BHA*length_BHA)-wob)*9.81)/1000+hookload_noise)
ecddata.append(mudweight+annulusfriction/(0.0981*(drilleddepthstart+\
+rop*(time-number_of_connections*connectiontime))))
pitgain.append(0+pitgain_noise) # Adding 0 pitgain when no influx/kick is taken

block_pos -= block_height/(section_drillingtime/timestep) # Moving block position
equally downwards per step
block_pos_array.append(block_pos)

"Setting the possibility of kick before last connection"
if(number_of_connections == connections-1):

    kicknumber = random.uniform(0,1) # Random possibility of kick from 0-1

    "Generating kick if kicknumber is lower than the probability of kick"
    if kicknumber<probkick:
        triggevent = True # Activating kick event by setting
triggevent to True
        kicksize = random.uniform(kicksize_min,kicksize_max) # m3
        rop_increase = random.uniform(rop_inc_min,rop_inc_max) # Random percentage
increase in ROP
        kick_det_times.append(time) # Storing time when kick is obtained
        kick_start = drilleddepthstart+rop*(time-number_of_connections*
connectiontime) # at what depth kick is taken
        "Getting density of kick at relevant conditions through above defined
functions"
        temp_kick = temp_calc(kick_start)+273.15 # Calc-
ulate temp at kick pos. [K]
        botpres_kick= (mudweight*0.0981*kick_start+annulusfriction)*(10**5) # Calc-
ulate pressure at kick pos[Pa]
        kickweight = rhog(botpres_kick,temp_kick) # Calc-
ulate density of kick

        "Adding a sudden ROP increase before kick"
        for i in range(12):
            sudden_increase = int(len(ropdata)-1-i) # Adding a point for sudden
increase some steps before kick
            "Breaking out of Loop if ROP increase is moving into the connection"
            if(np.any(np.in1d(connection_array,timedata[sudden_increase]))):
                break
            else:
                ropdata[sudden_increase] \
                = ropdata[sudden_increase]*rop_increase+ROP_noise # Adding a sudden
increase in rop before kick

        "Adding the gained pump pressure while drilling downward, and constant flow in,
flow out and ROP"
        sppdata.append(pumppressurestart+pumppressure_gain*(rop*(time-number_of_connections
*connectiontime))+pump_noise)
        flowrateindata.append(flowin)

```

```

flowrateoutdata.append(flowout+flowrateout_noise)
if(triggevent == True): # Adding a single step of increased ROP at time kick is
detected
    ropdata.append(rop*rop_increase+ROP_noise)
else:
    ropdata.append(rop+ROP_noise)
kick_array.append(0) # Adding 0 for no kick

elif (time<drillingtime and triggevent == True):
    ""Drilling with circulation and kick/influx.

Calculations:
-----
    * drilleddepthstart+rop*(time-number_of_connections*connectiontime) = drilled depth
based on time spent drilling
    * mudweight*0.0981*(drilleddepthstart-height_kick+rop*(time-number_of_connections*
connectiontime)) + kickweight*height_kick*0.0981+annulusfriction = finding bottomhole pressure
when kick is obtained in well * topdrive+(bouyancy_factor*(weight_DP*(drillpipelength+rop*(time
- number_of_connections*connectiontime)) + weight_BHA*(length_BHA-height_kick))+
bouyancy_factor_kick*(weight_BHA*height_kick)-wob))*9.81)/1000 = hookload calculated when kick
is in well, subtracting the kick height from BHA and using the kick buoyancy factor to
calculate for area covered by the kick height.
-----
    ""

    "Creating an even influx over the eventduration:"
influxrate = kicksize/(eventduration)
inflowvolume += influxrate*timestep

    pitgain.append(inflowvolume+pitgain_noise) # Adding up inflow volume to the pit
gain, increasing over event period
    measureddeptdata.append(drilleddepthstart+rop*(time-number_of_connections*
connectiontime))
    tvddata.append(drilleddepthstart+rop*(time-number_of_connections*connectiontime))

    height_kick = inflowvolume/annular_capacity # Height of kick in annulus
based on inflowvolume

    bouyancy_factor_kick = 1 - (kickweight*do_BHA**2-mudweight*di_BHA**2)\
/(steelweight*(do_BHA**2-di_BHA**2)) # Calculating buoyancy factor for hook load
calculations at BHA with kick
    hookloaddata.append(((topdrive+\
    +(bouyancy_factor*(weight_DP*(drillpipelength+\
    +rop*(time-number_of_connections*connectiontime))+\
    +weight_BHA*(length_BHA-height_kick))+\
    +bouyancy_factor_kick*(weight_BHA*height_kick)-wob))*9.81)/1000
+hookload_noise)

    bottompressure.append(mudweight*0.0981*(drilleddepthstart-height_kick+\
    +rop*(time-number_of_connections*connectiontime))+\
    +kickweight*height_kick*0.0981+annulusfriction+botpres_noise)

    timeevent = timeevent + timestep

    "Bottom hole pressure with kick in well"
    P_kick = mudweight*((drilleddepthstart+rop*(time-number_of_connections*
connectiontime))-height_kick)*0.0981+kickweight*height_kick*0.0981
    "Difference in bottomhole pressure when kick is in well vs ordinary drilling
without influx"
    dP = mudweight*(drilleddepthstart+\

```

```

        +rop*(time-number_of_connections*connectiontime))*0.0981-P_kick

    ecddata.append(mudweight+annulusfriction/(0.0981*(drilleddepthstart+\
        +rop*(time-number_of_connections*connectiontime)))-dP/(0.0981*
(drilleddepthstart+rop*(time-number_of_connections*connectiontime)))
    "Reducing pump pressure by pressure difference due to kick/influx"
    sppdata.append((pumppressurestart+pumppressure_gain*(rop*(time-
number_of_connections*connectiontime)))-dP+pump_noise)
    flowrateindata.append(flowin)
    flowrateoutdata.append(flowout+(influxrate*1000*60)+flowrateout_noise) # Adding
influx in lpm to the flow out
    ropdata.append(rop*rop_increase+ROP_noise)
    kick_array.append(1) # Adding 1 for kick occurrence
    block_pos -= block_height/(section_drillingtime/timestep) # Moving block position
equally downwards per step
    block_pos_array.append(block_pos)

    # Check on if duration of event is over. Go back to normal conditions.
    if timeevent>=eventduration:
        triggevent = False
        timeevent = 0
        inflowvolume = 0
        kick_det_times.append(time)

    else:
        """
        Here is the connection.
        Calculations:
        -----
        * drilleddepthstart+(number_of_connections+1)*30 = adding the same depth of drilled
depthstart + standpipelength
per connection.
        * topdrive*9.81/1000 = hookload at connection is just topdrive force in kN
        * block_height-(block_height/(block_lifttime/timestep)) = finds the largest height
before stacking over the
block's maximum height
        -----
        """

    measureddeptdata.append(drilleddepthstart+(number_of_connections+1)*standlength)
    tvddata.append(drilleddepthstart+(number_of_connections+1)*standlength)
    bottompressure.append(mudweight*0.0981*(drilleddepthstart+(number_of_connections+1)
*standlength)+botpres_noise)
    hookloaddata.append(((topdrive)*9.81)/1000+hookload_noise)
    sppdata.append(0)
    flowrateindata.append(0)
    flowrateoutdata.append(0)
    ropdata.append(0)
    ecddata.append(mudweight)
    connection_array.append(time)
    pitgain.append(0+pitgain_noise)
    kick_array.append(0)

    "Moving block from bottom to top to position in new drillpipe for drilling new
section"
    if(block_pos < block_height-(block_height/(block_lifttime/timestep))):
        block_pos += block_height/(block_lifttime/timestep)
        block_pos_array.append(block_pos)
    else:
        block_pos = block_height

```

```

        block_pos_array.append(block_pos)

    number_of_connections += 1 # Adding up performed
    connection
    timeend += section_drillingtime + connectiontime # Extending timeperiod for
drilling for a new section
    drillingtime += section_drillingtime + connectiontime # Extending drilling period
for new section

"""
Adding connection gas for previous 3 connections if kick has occurred
Calculations:
-----
* annular_capacity*(drilleddepthstart+standlength*(connections-i-1)) = calculating the annular
volume by multiplying the
annular capacity by the drilled depth at the respective connection.
* int(connectiontime/timestep + measureddeptdata.index(drilleddepthstart+standlength*
(connections-i-1))) = finding the
time index of the respective connection and moving to the end of the connection
* round(annular_volume/(flowin/1000),-1)*60 = finding time to circulate annulus and rounding to
nearest 10, and converting
to seconds.
-----
"""
if(len(kick_det_times)>1):
    for i in range(4): # Assuming connection gas in previous 3 connections before kick occurs
        annular_volume = annular_capacity*length_BHA+annular_cap_DP*(drilleddepthstart\
+standlength*(connections-i-1)-length_BHA)
        influx_timeindex = int(connectiontime/timestep +
measureddeptdata.index(drilleddepthstart+standlength*(connections-i-1)))
        "Checking if the delayed connection gas measurement is inside the simulation time range"
        if(timedata[influx_timeindex]+round(annular_volume/(flowin/1000),-1)*60+connectiontime
< timedata[-1]):
            "Checking if delayed connection gas i marked in connection, moving it a connection
period forth in such a case"
            if(np.any(np.in1d(connection_array,timedata[influx_timeindex]))):
                influx_delays.append(timedata[influx_timeindex]+round(annular_volume/
(flowin/1000),-1)*60+connectiontime)
            else: # Adding the delayed connection gas measurement to an array storing the time
they were measured
                influx_delays.append(timedata[influx_timeindex]+round(annular_volume/
(flowin/1000),-1)*60)

"Adding the influx event to an array storing 0 for no influx during connection and 1 for influx
during connection"
connectiongas = np.zeros(len(timedata)) # To store values 0 for no gas and 1 for influx during
connection
for i in range(len(influx_delays)):
    if(len(timedata)<(influx_delays[i])/timestep): # Breaking out of loop if the influx delay
is outside simulation data
        break
    else:
        connectiongas[timedata.index(influx_delays[i])] = 1 # Marking the connectiongas at a
delayed time

"Creating a series of subplots"
fig, axs = plt.subplots(9,1, figsize=(15,20))

```

```

"Plotting pumppressure data vs timedata"
axs[0].plot(timedata,sppdata)
axs[0].set_title('Pump Pressure')
axs[0].set(xlabel='Time (s)',ylabel='Pressure (bar)')
axs[0].grid()

"Plotting bottomhole pressure vs timedata"
axs[1].plot(timedata,bottompressure)
axs[1].set_title('Bottompressure')
axs[1].set(xlabel='Time (s)',ylabel='Pressure (bar)')
axs[1].grid()

"Plotting measured depth data vs timedata"
axs[2].plot(timedata,measureddeptdata)
for i in range(number_of_connections): # Identifying connections with their respective number
    x = connection_array[int(i*(connectiontime/timestep))+2*i]
    y = measureddeptdata[timedata.index(connection_array[int(i*(connectiontime/timestep))+2*i])]
    axs[2].annotate(f'Connection {i+1}',(x,y),xytext=(x,y-20))
axs[2].set_title('Drilled length')
axs[2].set(xlabel='Time (s)',ylabel='Measured depth (m)')
axs[2].grid()

"Plotting ECD-data vs timedata"
axs[3].plot(timedata,ecddata)
axs[3].set_title('ECD')
axs[3].set(xlabel='Time (s)',ylabel='ECD (sg)')
axs[3].grid()

"Plotting ROP-data vs timedata"
axs[4].plot(timedata,ropdata)
axs[4].set_title('ROP')
axs[4].set(xlabel='Time (s)',ylabel='ROP (m/s)')
axs[4].grid()

"Plotting flowrate out data vs timedata"
axs[5].plot(timedata,flowrateoutdata)
axs[5].set_title('Flowrate out')
axs[5].set(xlabel='Time (s)',ylabel='Flow rate out (lpm)')
axs[5].grid()

"Plotting possible connection gas where 1 implies connection gas at given time and 0 no connection gas"
axs[6].plot(timedata,connectiongas)
for i in range(len(influx_delays)): # Identifying connection gas tag with connection number
    x = timedata[timedata.index(influx_delays[i])]
    y = 1
    axs[6].annotate(f'Connection {(connections-len(influx_delays)+1)-i}',(x,y),
xytext= (x-250,0.3),rotation=90)
axs[6].set_title('Connection gas')
axs[6].set(xlabel='Time (s)',ylabel='Connection gas')
axs[6].grid()

"Plotting pitgain vs timedata"
axs[7].plot(timedata,pitgain)
axs[7].set_title('Pitgain')
axs[7].set(xlabel='Time (s)',ylabel='Pitgain (m3)')
axs[7].grid()

"Plotting hookload and block position on twinned axes vs timedata"
axs9 = axs[8].twinx() # Creating a twin axis for block positioning
axs[8].plot(timedata,hookloaddata,label='Hookload')

```

