NTNU
Norwegian University of
Science and Technology

Parameter Detection in real time Drilling Data

Create a Matlab agent to forecast changes in formation hardness

Clement Pierre Jean Donne

Petroleum Engineering
Submission date: June 2017
Supervisor: Pål Skalle, IGP

Norwegian University of Science and Technology
Department of Geoscience and Petroleum

i. Abstract

Nowadays, being cost efficient is a major issue of the oil and gas sector. Reducing Non Productive Time (NPT) is one of the solutions to tackle this challenge. As a result, the industry is always looking for more solutions to prevent failures and to reduce the number of incidents in order to increase the operational time.

The hardness of a formation is mainly linked to its mineral composition and its degree of cementing, but unexpected changes in formation hardness can create an inefficient rate of penetration (ROP), a too high torque on the drill string, a too high weight on bit (WOB), and a fast drilling bit wear. These phenomena can create severe failures such as drill pipe failures, stuck pipe or washout, and local doglegs (Donne, 2016).

The objective of this master thesis is to provide a data agent in Matlab able to forecast these changes in formation hardness using real time drilling data (RTDD) and to subsequently provide a hardness classification of the complete well path.

Regarding the hardness computation, the mathematical model behind the algorithm of the program is a simple version of ROP model proposed by Bourgoyne and Young in 1986. The forecast is an autoregressive moving average (ARMA) model. Several further assumptions are made to finally compute and forecast formation hardness through drillability using ROP, WOP and revolution per minute (RPM), all available in real time drilling data. According to previous studies made on the same Bourgoyne and Young model, relevant exponents are selected for WOB and RPM depending on the formation being drilled. ARMA coefficients are selected using an iterative process within the agent algorithm.

To notice the efficiency of the agent proposed, a comparison is made between hardness classification results given by the program and changes in formation hardness looking through RTDD of ten wells and their geological reports. Information on how to select relevant real RTDD cases, how to clean raw files, and how to process them is also provided.

Finally, the agent provides hardness classification reports and forecast while drilling but the precision and the accuracy is quite low. These irregularities are mainly due to the assumption made in the mathematical model used for hardness computation and in the forecasting model. Possible imprecision in test cases implementation is another factor of the final low accuracy of the agent.

ii. Acknowledgment

This master thesis in Drilling Technology is carried out at the Norwegian University of Science and Technology (NTNU) at the Department of Department of Geoscience and Petroleum. The report has been written by Clement Donne, 5th year student in a Petroleum Engineering master with specialization within drilling. Professor Pål Skalle has been the supervisor for present project.

I am sincerely thankful to my supervisor Pål Skalle for his high expertise, his guidance, his advises and feedbacks during all the year. Besides, I would like to thank Anisa Noor Corina, Isak Swahn and Tommy Toverud for their help and advices regarding this project. Moreover, I am also truly thankful to all professors at NTNU for teaching me all I know about oil and gas.

This project has been a great learning experience having the opportunity to use real RTDD from oil companies and to access Diskos database. Both improving my knowledge about the all industry and increasing my programming skills.

I would also like to thank my family, especially my mother, supporting me for years from abroad. Finally, I am grateful to fellow students, Erasmus and rugby friends for their high support during the writing of this thesis.

Trondheim, 2017

Clement Pierre Jean Donne

iii. Table of contents

iv. List of tables

v. List of figures

1 Introduction

Motivation

Nowadays the oil and gas market is tough, price of the crude Brent oil moved from 112,5 dollars per barrel in July 2014 to 37 dollars per barrel in January 2016 (Oilprice, 2017) and conflict and political issues will cause it to keep fluctuating. In such difficult times, the industry try to reduce costs as much as possible to stay competitive. Non Productive Time (NPT) is the biggest useless expense of this sector and finding a way to reduce NPT can provide companies with a high drilling efficiency.

Aadnoy (2010) explained that ten to twenty percent of the rig time is spent handling unforeseen problems. Fixing failures is costly, thus, methods to prevent incidents and failures are wanted.

State of Art

The Department of Geoscience and Petroleum of the Norwegian University of Science and Technology (NTNU) has proudly been developing data agents for years to detect parameter deviations using real time drilling data (RTDD) from eight Statoil wells. So far, a hydraulic friction agent, a calculated rate of penetration (ROP) agent and a mechanical specific energy agent were created (Skalle and Swahn, 2016). Donne in 2016 started to implement an agent able to detect changes in formation hardness using RTDD. When entering a hard formation which is an area where some sediments are more consolidated than others, incident such as inefficient ROP, too high weight on bit (WOB), fast drilling wear or too high torque on the drill string may occur (Donne, 2016). If not handle properly these incidents turn into failures such as stuck pipe, drill pipe failure, local doglegs or washouts and are a source of non-productive time for the companies.

Goals

Following the fact that preventing changes in formation hardness would probably lead to reduced NPT and thus increased drilling efficiency. The main purpose of this master thesis is to develop a new agent able to predict changes in formation hardness. This thesis is based on the previous agent done by Donne (2016) which was able to detect changes in formation hardness. The improvement will be to forecast those changes while drilling in order to send back information to the driller who can modify his drilling parameters to avoid incidents.

Approach

Reaching the goal is based on seven steps. First, an investigation of the literature about formation hardness to understand mechanical processes involved and how it does cause failures. Then an improvement of the mathematical model created used by Donne (2016) which should detect formation hardness while drilling. Later, a selection of relevant RTDD cases from the industry and a study of these cases. In associated drilling and geological reports, changes in formation hardness will be detected manually. It provides a test base for the agent. Then an updated and improved implementation of Donne's agent in Matlab. Finally, after testing the agent to evaluate its performance, a self-assessment of the work conclude the master thesis.

2 Relevant published knowledge

This chapter presents background information needed to reach the aim of this thesis. This include the definition of formation hardness, and what incidents this physical aspect of the rock involve during the drilling process. Furthermore, it presents knowledge about forecasting hardness using time data and how to test the efficiency of a numerical agent.
This chapter is an extended version of the project done by Donne (2016).

2.1 Hardness

Using work done by Hoseinie, Ataei and Al. (2012) and their comparison of different rock hardness scales, hardness can be define as a mineral or rock's resistance to penetration by the drilling tool. This present definition is very close to the description of the drillability. This term is used and defined by the mathematical model in chapter 3.1. Hardness is a physical property of the rock and many empirical methods exist to define it as Mohs hardness, Indentation Hardness Index or L-type Schmidt hammer (NDT, 2017). The importance of this rock property during the drilling process is the relation between hardness and rate of penetration (ROP). Undeniably, when the rock strength increases, ROP decreases while the drill bit wear increases which again increases drilling time and thus drilling cost.

2.2 Hard, Soft and laminated formation

To define difference between different type of formation hardness and describing drilling incidents, the Oilfield glossary of Schlumberger available online is helpful and used frequently in following subchapters.

Schlumberger [1] (2017) presents hard rock as a specific category of sedimentary rock which is particularly difficult to disaggregate. A factor of rock hardness is its mineral composition. For instance, according to the Mohs hardness scale, a formation full of quartz will be much harder than one containing calcite or gypsum. Although, the most important cause of hard formation is a high degree of cementing (Solberg, 2012). On the contrary, soft formation is less cemented and easier to disaggregate.

Regarding laminated formation, this phenomenon occurs when sedimentary rocks are deposited in layers (Figure 2-1).

Figure 2-1 Example of a shale formations laminate structure according Norton (2016)

The problem with laminated formation is that layers have different permeabilities leading to an anisotropic formation. Drilling through anisotropic rock is thus challenging and incidents may occur as described in following sub chapters.

2.3 Incidents linked with changes in formation hardness while drilling

This sub chapter is a modified version of the project wrote by Donne (2016).

2.3.1 Incidents linked with hard formation

When sedimentary rocks strength increases suddenly, it may create issues during drilling operations.

2.3.1.1 Low rate of penetration

First, as the formation encountered is suddenly harder to disaggregate, the rate of penetration (ROP) decreases. If the hard formation was not expected, then the bit chosen may be incorrect and the ROP will stay low all along this section. In that case, the cuttings bed downhole is quite low and can cause reactive formation failures (Skalle, Aamodt and Gundersen, 2013). Moreover, with low ROP, the drilling time becomes significantly higher, which has a direct influence on drilling expenses (DRILLEX) for companies as they have to rent drilling facilities for a longer period of time.

2.3.1.2 Drill string failure

We described before that ROP decreases in hard formation. In addition, operators will have to increase the weight on the bit (WOB) to compensate the high rock strength and keep going down. Yet, a higher WOB induces higher torque on the drill string. The main failure resulting from this situation is a hole in the string (Solberg, 2012).

2.3.1.3 Fast drill bit wear

When encountering an unexpected hard formation area, the drilling process occurs with equipment not designed for such rock strength properties. In that case, this drilling tool will wear out faster than planned. It leads to a non-productive time (NPT) due to a faster change of the bit, which significantly increases the cost of the well.

2.3.2 Incidents linked with a sudden change in formation hardness

2.3.2.1 Unplanned Dogleg

Dogleg are deviated or curved sections of a borehole, as shown in Figure 2-2. They can be desired and chosen (Schlumberger [2], 2017), or unexpected in case of changes between hard and soft formation. In this situation, the increase WOB on the drill string to enter the hard formation will tend to buckle the drill string inside the soft formation where wellbore is weaker. It changes the penetration angle of the drilling bit and starts an unexpected directional drilling in this well. According to Schlumberger [2], (2017), consequences of a dogleg are significant. First, the wellbore is not located in the planned path, which can prevent a planned casing string from fitting. In addition, the repeated abrasion of the drill string in the curved region of the borehole may create a keyseat, also presented in Figure 2-2, in which the bottom hole assembly components can become stuck as they are pulled through the section (Donne, 2016). A stuck pipe failure is expensive as it stops drilling operations in favor of jar or fishing operations. In the worst scenario, it leads to a sidetrack and thus to even more DRILLEX.

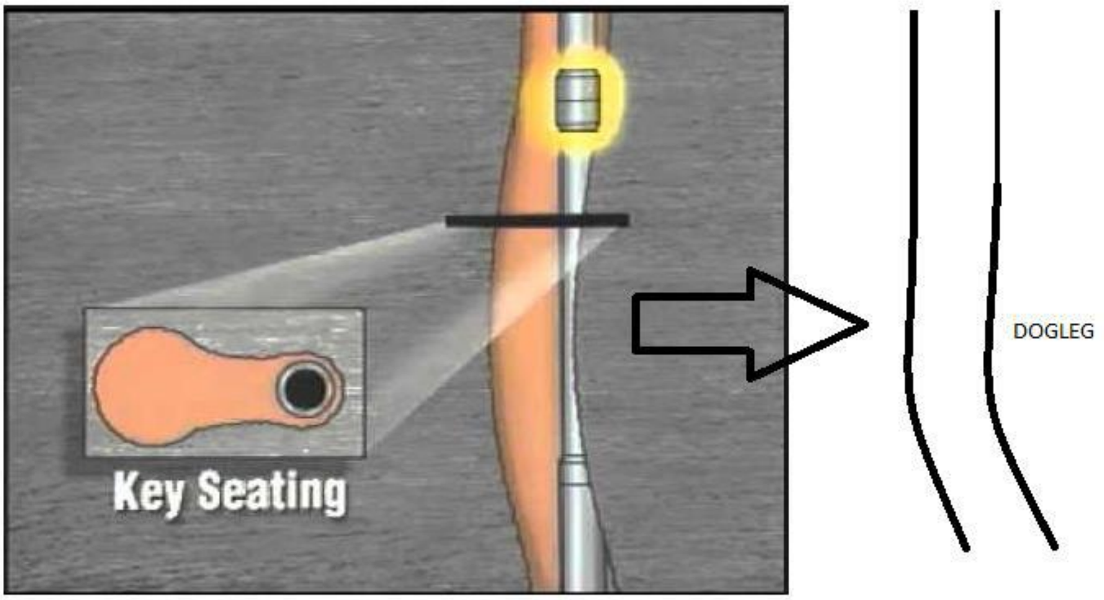


Figure 2-2 Illustration of a keyseat and a dogleg (Donne, 2016)

2.3.2.2 Washout

The previous section introduced the concept of dogleg. However, the first step is often a washout. The Schlumberger oilfield glossary define it as an enlarged section of the wellbore (Schlumberger [3], 2017). Like the dogleg, it is due to the buckled pipe in the soft area. The main problem with washout is that the pipe tends to stick at the ledges and the shoulders of the wellbore when pulling out or tripping in (Donne, 2016).

2.3.3 Incidents linked with laminated formation

The main problem resulting from laminated formation is the deviation of the expected well trajectory (a dogleg), we already discuss problems linked by doglegs in sub chapter 2.3.2.2 yet, the drilling process into these specific formations need to be more detailed. Brown, Green and Sinha (1981) explained that vertical well would tend to deviate following the perpendicular to the laminations if the laminations are not steeply dipping. It is called up-dip. On the opposite in steeply dipping strata, hole deviation tends to be down-dip meaning it is parallel to the laminations. The figure 2-3 presents this phenomenon:

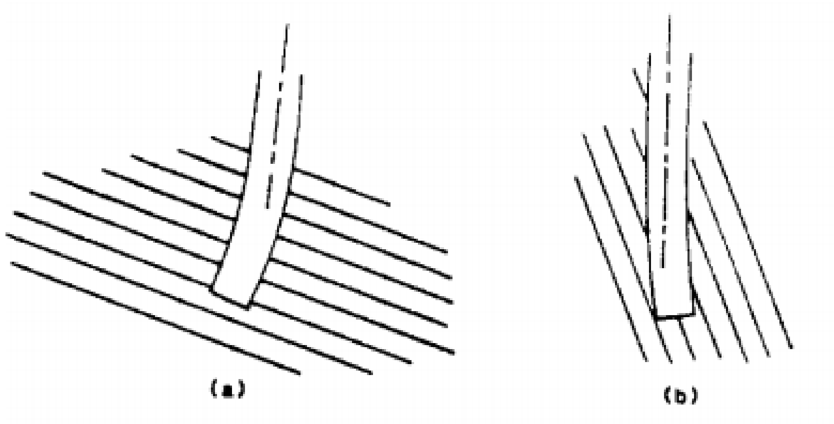


Figure 2-3 Borehole deviation in shallow dipping (a) and steeply dipping (b) in laminated area, (Brown, Green and Sinha, 1981)

2.4 Forecasting using time series

Understand formation hardness is important is important as we want to prevent related issues. However, we also want to be able to predict these issues using real time drilling data (RTDD), thus, it is important to understand forecasting methods.

2.4.1 Real time drilling data and Time series

Time series is a sequence of random variables taken sequentially in time (Box, George and al. 2015). It appears that many sets of data are considered as a time series, for instance a monthly sequence of goods sold by a shop, or hourly observations made on the yield of a chemical process.

RTDD are surface recorded data (Schlumberger [4], 2017) providing values of many parameters (such as ROP or WOB) at a precise time. In practice, RTDD are recorded every three to five seconds.

This sequential record of random variables over an extended period of time allow us to consider RTDD as a time series and thus allow us to use mathematical methods dedicated to these series to reach the goal of this master thesis.

2.4.2 Forecast methods using time series

The analysis of time series and their utilization in long and short-term prevision is well documented and frequently used in several sectors such as finance, meteorology, or statistics. In his introduction to time series analysis, Mélard (2006) explains that many different forecasting methods of varying complexity and accuracy exist. They are divided into various categories such as moving average, growth curves, seasonal decomposition, exponential smoothing, and multiple regression. The selection of an appropriate method is based on the following information:

- Extrapolative methods such as exponential smoothing, moving average, and auto regression use present and past recorded data to predict coming data.
- Explanatory methods such as simple or multiple linear regression use present and past recorded data but also present and previous recorded time events to predict coming data and time.
- Systemic methods use the relationship between data and time to predict future.

RTDD measurements are independent but recorded using a constant interval, so the study of extrapolative methods is relevant and sufficient for this master thesis.

We choose to focus on two methods and will select the more relevant after tests on RTDD: Exponential smoothing and autoregressive integrated moving average (ARIMA).
Chapter 3 describes further and implements these two methods.

2.5 Computer agent testing

Software testing, defined as a process of executing a program to find errors, is an important component of software quality assurance and many software organizations are spending up to 40 % of their resources on testing (Jovanović, 2006). We do not have such resources but we still need to understand how it works to assess the quality of the data agent.

Botella, Burgués and Al. (2004), wrote that the International Organization for Standardization (ISO) has defined a set of quality standards widely used to describe the quality of a product. In terms of software quality, the most widespread one is the ISO/IEC 9126, used as a framework for software evaluation that we need for the agent.
The model is defined by means of general characteristics of software, which are further refined into sub characteristics as shown in Table 2-1. At the bottom of the hierarchy are the measurable software attributes computed using some metric. When a software fulfills all these requirements, it is then regarded as a high quality program.
Considering time available for this master thesis, we will not test reliability, usability, maintainability and portability, as the agent we build is not a commercial product ready to be set up in operational drilling facilities in front of technical operators. However, we can find a test to assess functionality and efficiency of the software by challenging sub characteristics like accuracy and time behavior of the agent. Chapter 7 describes further in detail and implements the test methodology chosen for the thesis.

TABLE 2-1 THE ISO/IEC 9126-1 INTERNAL/EXTERNAL QUALITY MODEL (ISO/IEC, 2001)

Characteristics	Sub characteristics
Functionality	Suitability Accuracy interoperability security functionality compliance
Reliability	maturity fault tolerance recoverability reliability compliance
Usability	understandability learnability operability attractiveness usability compliance
Efficiency	time behavior resource utilization efficiency compliance
Maintainability	analyzability changeability stability testability maintainability compliance
Portability	adaptability installability co-existence replaceability portability compliance

3 Mathematical model of the agent

The purpose of the project is to propose a data agent using RTDD to predict changes in formation hardness. In order to create useful algorithm for this agent, we must be able to assess a value of hardness. However, hardness is not quantifiable, that is why we need a quantifiable indicator of hardness. We choose the drillability.

3.1 Drillability

The complete chapter is an extended version of the project by Donne (2016). Manutchehr-Danai and Mohsen (2013) wrote that the drillability is the specific value or relative speed at which a material may be penetrated by a drill bit. For this reason, we understand that drillability can be a good indicator of the formation hardness as it indicates if the penetration of the formation is relatively easy or difficult:

$$Hardness = \frac{1}{Drillability} \tag{3.1}$$

Since there is a lack of existing tools for measuring the "drillability" at any point in a well (Rakhleev, 1965), we will have to estimate it according to several other drilling parameters. Following the work of Rashidi, Hareland, and Wu (2010), we will consider a strong link between the drillability and the rate of penetration.

Indeed, on one hand, drillability indicates if penetration is easy or difficult while on the other hand the rate of penetration indicates if the penetration is slow or fast. The idea behind using ROP is that it is a data easily recordable during drilling and so easily useable by the agent.

3.2 Bourgoyne and Young Rate of Penetration Model

This subchapter was taken from the previous specialization project by Donne (2016).

The manner in which the important drilling variables (weight on bit, rate of penetration, rotation per minute…) affect penetration rate is quite complex and only partially understood (Bourgoyne, Millhelm and Chenevert and al., 1986). That is why; we will have to use a model that we know will not be perfect but as complete as possible. Previous specialization project proved that mathematical model known as the Bourgoyne and Young Rate of Penetration model could be relevant for this agent.

It was normally designed to be used for rolling cutter bits but according Miska, Rajabov et Al. (2012); it also can be used for PDC bits used today.

The Bourgoyne-Young drilling model (1974) can be defined as follows:

$$Rate\ of\ Penetration = ROP = (f_1) * (f_2) * (f_3) * (f_4) * (f_5) * (f_6) * (f_7) * (f_8) \quad (3.2)$$

Each of these eight functions model a specific phenomenon:

f_1 represents effects of formation strength and bit:

$$f_1 = e^{2,303*a_1} = K \quad (3.3)$$

f_2 represents the increase in rock strength due to normal compaction with depth:

$$f_2 = e^{2,303*a_2*(10000-D)} \quad (3.4)$$

f_3 is the effect of under compaction experienced in abnormally pressured formations:

$$f_3 = e^{2,303*a_3*D^{0,69}*(g_p-9)} \quad (3.5)$$

f_4 models the effect of overbalanced drilling: $f_4 = e^{2,303*a_4*D*(g_p-\rho_c)}$ (3.6)

f_5 represents the effect of bit weight : $f_5 = [\frac{\frac{W}{d_b} - (\frac{W}{d_b})_t}{4 - (\frac{W}{d_b})_t}]^{a_5}$ (3.7)

f_6 represents the effect of rotary speed : $f_6 = (\frac{N}{60})^{a_6}$ (3.8)

f_7 models the effect of tooth wear : $f_7 = e^{-a_7*h}$ (3.9)

f_8 represents the effect of bit hydraulics : $f_8 = (\frac{F_j}{1000})^{a_8}$ (3.10)

Here a_2 to a_8 are constants that must be chosen based on local drilling conditions through drill-off tests

All terms in these equations are defined in the nomenclature (chapter 11).

This model is useful if we know all of these parameters. However, a_1 to a_8 exponents have to be determined using prior drilling data obtained in the area (Bourgoyne, Millhelm and Chenevert and al., 1986). The problem is we do not have all of these data yet as the agent is here to predict formation changes while drilling, meaning before getting all of these data. The other problem is that frequent changes in formation parameters with depth will make the selection of relevant constants for all formation types difficult. That is why we have to both use assumptions and simplify this model (Donne, 2016).

3.3 Assumptions on the model

This sub chapter is also part of the project made by Donne in 2016.

We compute our full hardness formula by using equation (3.1) in f1 (3. 3) and replacing it in the ROP formula (3.2) with K representing drillability:

$$hardness = \frac{1}{drillability} = \frac{1}{K} = \frac{1}{f1} = \frac{(f_2) * (f_3) * (f_4) * (f_5) * (f_6) * (f_7) * (f_8)}{ROP} \tag{3.11}$$

Now we proceed to several assumptions.

3.3.1 Neglect factors due to their slow changing with depth or low level of influence

Table 3-1 presents two types of factors existing in the formula 3.11:

TABLE 3-1 CLASSIFICATION OF BOURGOYNE AND YOUNG MODEL FACTORS ACCORDING THEIR VARIATION TIME

Factors changing over long drilling interval	Factors changing instantly
-linked to formation f2 - linked to bit wear: f7	- linked to formation: f1, f3 and f4 - linked to our drilling parameters ROP and WOB : f5,f6, f8

The purpose of the agent is to predict changes in formation hardness over a short time scale, as we know it is impossible to predict directly the formation hardness of the entire path. Thus, we will consider just a small drilling window and not a long drilling interval.

Consequently, we consider factors changing over long drilling interval as a new constant C, and Eq 3.11 become:

$$hardness = \frac{C * (f_3) * (f_4) * (f_5) * (f_6) * (f_8)}{ROP} \tag{4.12}$$

With C = $(f_2) * (f_7)$ and is assumed to be constant

According Solberg (2012) this first assumption looks relevant as she did it in her similar research in detecting formation hardness she includes f_2 factor but demonstrated there is almost no change in hardness due to applying the depth correction exponent, which can be removed in further analyses (Donne, 2016).

3.3.2 Assumptions on factor linked to formation

Let's have a look back to factors linked to formation:

$$f_3 = e^{2,303*a_3*D^{0,69}*(g_p-9)} \tag{3.5}$$

$$f_4 = e^{2,303*a_4*D*(g_p-\rho_c)} \tag{3.6}$$

The first assumption done by Donne (2016) is to ignore the 2,303 coefficients. Indeed, Bourgoyne, Millhelm, Chenevert and al. (1986) wrote that this number allows the constant a_1 to be defined easily in terms of the common logarithm of a penetration rate. As our coefficient a_1 is included in our hardness, we do not need to have the exact value of the drillability but its variation. Therefore, we can neglect the 2,303 coefficient. g_p represents the pore pressure gradient. It can change quickly in abnormal pressurized formations. Thus, we need an efficient tool to compute those variations. According, Jorden and Shirley (1966), the utilization of the normalized rate of penetration (d-exponent) will be necessary to correlate it to the differential pressure. Coupled to the dc-exponent (Rehm and McClendon, 1971) we should be able to plot two curves describing differential pressure and obtain formation pressure. The d-exponent is defined as:

$$d = \frac{\log(\frac{R}{60N})}{\log(\frac{12W}{10^6 db})} \tag{3.13}$$

And the dc-exponent defined as:

$$d_c = d * \frac{MW1}{MW2} \tag{3.14}$$

All terms in these equations are also defined in the nomenclature (chapter 11). However, an issue remains with this method. Indeed, 3.13 and 3.14 show that the d component computation need weight on bit (W) and rotation per minute (N) despite the fact that we already use them to compute drillability. A technique to include this formation pressure parameter could be to ask the agent to do some iterative loops in order to compute the right value needed for the drillability (Donne, 2016). This implementation is a challenging for the time available in this master thesis, thus we assume a value of 1 for the function f_4, which is the case for zero overbalance situation (Bourgoyne, Millhelm and Chenevert and al., 1986). We will assume f_3 equal to 1 as if we drill on normal pressurized formation. In order to make this last assumption relevant, and as done in 2016 project, we will test the agent in intervals where no formation with abnormal pressure were detected according to the end of well reports if available. Chapter 5 presents this interval selection.

3.3.3 Assumption on the weight on bit factor

$$f_5 = [\frac{\frac{W}{d_b} - (\frac{W}{d_b})_t}{4 - (\frac{W}{d_b})_t}]^{a_5} \tag{3.7}$$

The threshold bit weight coefficient $(\frac{W}{d_b})_t$ is often quite small and can be neglected in relatively soft formations area (Bourgoyne, Millhelm, Chenevert and al., 1986).

Moreover, the drill bit diameter is a constant such as the number 4, which is here a conversion factor. We saw with f_2 that we do not need to have the exact value of the drillability but rather its variation so we can also neglect these last two constants (Donne, 2016) and use:

$$f_5 = [W]^{a_5} \tag{3.15}$$

3.3.4 Assumption on the rotary speed factor

$$f_6 = (\frac{N}{60})^{a_6} \tag{3.8}$$

Here again the 60 value is a conversion factor we can neglect such as all constant (and so the C constant of Eq. 12) still using the fact that we do not need to have the exact value of the drillability (Donne, 2016).

3.3.5 Assumption on the hydraulic factor

$$f_8 = (\frac{F_j}{1000})^{a_8} \tag{3.10}$$

This factor is linked to jet impact of the nozzle on the formation. It changes quickly following flow rate (Donne, 2016). According the microbit studies conducted by Eckel (1967) about the influence of hydraulics on the rate of penetration, there is a relationship similar to the Reynolds number controlling the combined effect of fluid properties and hydraulics on rate of penetration. However, the influence on the rate of penetration is relatively small compare to weight on bit or revolution per minute. Thus, we can normalize this f_8 term to be equal to 1.0 for a jet impact of 1,000 lbf (Nascimento, Kutas, Elmgerbi and al., 2015).

3.3.6 Final simplified mathematical model

Including all these assumptions, we obtain our final mathematical model linking hardness of the formation to the rate of penetration:

$$\boldsymbol{hardness} = \frac{[\boldsymbol{W}]^{\boldsymbol{a_5}} * [\boldsymbol{N}]^{\boldsymbol{a_6}}}{\boldsymbol{ROP}} \tag{3.16}$$

This model was used during Donne's project in 2016 and considered precise enough. The added value of this master thesis is the forecasting part in sub chapter 3.6 and the improvement of the agent internal code.

3.4 Choosing exponents values

Equation 3.16 represents the model to compute hardness along the wellbore. Five elements compose it:

- Weight on bit (W), rotation per minute (N) and rate of penetration (ROP) provided directly by real time drilling data
- a_5 and a_6 still must to be defined.

Bourgoyne, Young et al. (1986) wrote that these constant has to be determined for each formation using the drilling data taken from the previous drilling records of the field. Table 3-2 presents the recommended bounds for these coefficients, provided directly by Bourgoyne et al (1986) and based on the reported ranges various formations in different areas. This table is relevant, as it has been used by many researchers such as Rahimzadeh, Mostofi et al. (2011) to work on methods to determine these constants; or Nascimento, Kutas et al. (2015) who wanted to adapt this table for the specific case of the Presalt formation layers in Angola.

TABLE 3-2 BOURGOYNE AND YOUNG COEFFICIENTS UPPER AND LOWER BOUNDARIES (DONNE,2016)

Coefficient	Lower bound (soft)	Upper bound (hard)
a1	0,5	1,9
a2	0,000001	0,0005
a3	0,000001	0,0009
a4	0,000001	0,0001
a5	0,5	2
a6	0,4	1
a7	0,3	1,5
a8	0,3	0,6

We notice that the bit weight exponent a5 obtained from field data ranges from 0.5 to 2.0 and the rotary speed exponent a6 ranges from 0.4 to 1 depending on formation studied. We must choose a combination of these coefficients for extreme soft formation, medium soft formation, average formation, medium hard formation and extreme hard formation as this hardness scale is the one we want to get for our agent.

Solberg (2012), Stunes (2012) and Helgestad (2010) also used the Bourgoyne and Young model in their works assuming different a5 and a6 values to compute hardness. Table 3-3 presents their assumption and allow us to choose a5 and a6 for the average formation of this project.

TABLE 3-3 BOURGOYNE AND YOUNG A5 AND A6 COEFFICIENTS USED IN SIMILAR PROJECTS (DONNE, 2016)		
a5	a6	Reference
1,2	0,6	Solberg, 2012
1	0,7	Helgestad, 2010
0,9	0,4	Stunes, 2012
1,1	0,65	average representative exponent for this project

As we know a5 and a6 are average values, and we still must determine these coefficients for complete hardness scale. In their work using the Bourgoyne and Young model to compute the rate of penetration, Bahari and al. (2008) provided coefficients according to five formations drilled but also provided corresponding lithology. Bahari and al. (2008) computed coefficients for soft formations (containing mainly clay, limestone or sands) moderate hard (containing hard limestone, sandstone or dolomite) and hard ones (granites, basalts or quartzite). Dr. Skalle's oral comments (2017) about typical coefficients values worldwide emphasizes the researchers need to use average numbers as presented in table 3-4 as the selection of coefficients for a5 and a6:

TABLE 3-4 A5 AND A6 COEFFICIENTS ACCORDING TO THE FORMATION DRILLED		
Formation	a5	a6
Extreme hard	1,9	0,9
Moderate hard	1,5	0,75
Average (from table 3-3)	1,1	0,65
Moderate soft	0,8	0,6
Extreme soft	0,5	0,4

Coefficients a5 and a6 for hardness scale exist now but we still need to find the corresponding hardness value scale. In other words, when do we have to change a5 and a6 value when computing hardness?

3.5 Determination of the hardness ladder to apply exponential coefficients

3.5.1 Determination of the hardness ladder baser on the median hardness value

Originally, the project by Donne (2016) raised the idea of dividing the regular hardness by the maximal hardness along the well path. Such method provides a normalized hardness between zero and one, which is easy to classify. Close to 0 means soft and close to 1 means hard. However, to compute the maximum hardness data from the complete well path needed to be known which is irrelevant, as we want to predict it. To tackle this challenge we use a new methodology. We look for wells drilled in the same area and drilled during the same years. The reason is having similar geology and formations along all wells and then having similar technologies used by the drilling rig for all these wells as equipment have a direct influence on the Bourgoyne and Young model. The idea is to compute hardness along all these wells paths using only the average a5 and a6 coefficients values and have a look at the hardness values resulting.

Data are selected from ten wells in the same area in the North Sea. Chapter 4 will give a detail presentation of these data. However, Table 3-5 represents boundaries of the hardness found in these ten wells. Appendix I presents Matlab code of the test and results.

TABLE 3-5 BOUNDARIES OF HARDNESS VALUES COMPUTED IN 10 WELLS

Well	1	2	3	4	5	6	7	8	9	10
Min Hardness	0.002	0	1.083	0.27	0.03	0	0	0.20	4.20	0.002
Median Hardness	0.6457	1.906	172.6	30.69	8.336	1274	4.077	51.05	9.476	7.697
Max Hardness	5e+35	1e+07	8e+13	1e+31	4e+29	4e+35	2e+40	3e+12	278.31	2e+31
Log min Hardness	-6.10	-9999	0.08	-1.3	-3.41	-9999	-9999	-1.58	1.43	-6.10
Log median Hardness	-0.437	0.645	5.15	3.42	2.12	7.14	1.4	3.93	2.24	2.04
Log max Hardness	82.30	16.70	32.09	71.71	68.24	82.15	92.90	28.74	5.62	72.16

Regarding hardness unit, following equation 3.11, and making a dimensional analysis, as hardness is presented as one divided by a constant K unitless, we consider hardness as unitless. We notice that hardness ranges over a very large scale, from extreme soft hardness value at around 0 to extreme hard where hardness value is around 1E40 at the highest. Such a value is quite unrealistic but we detail it in the chapter 8. Thus, logarithmic scale is a good solution to describe evolution of hardness versus depth, as presented in end of Appendix I.

We also observe that there is no common median hardness to all wells. However, once the mean of the hardness is known for a well, we can manage to find a hardness ladder for exponential coefficients based on the computation of the logarithm of the hardness. Indeed figures in Appendix I presents strong hardness variations in softer and harder area compare to median value.

As an assumption for this master thesis, Table 3-6 propose a hardness ladder for exponential coefficients:

TABLE 3-6 HARDNESS LADER FOR EXPONENTIAL A5 AND A6 COEFFICIENTS

Coefficient to apply	Condition using H= hardness med= median of the hardness
Extreme soft	H under med/100
Moderate soft	H belongs to [med/10, med/100[
Average	H belongs to]med/10,med*10[
Moderate hard	H belongs to [10*med;100*med[
Extreme hard	H above 100*med

So far, the chapter explains how to compute hardness, what coefficients use and in which case. However, the hardness ladder uses the median of the hardness computed using the full length of the well. As we have to predict hardness during drilling hardness of full path is an unknown, and we need to find complementary method to approximate the median hardness value.

3.5.2 Determination of the median hardness value

The idea is to estimate as quickly as possible the median value of the hardness for the well to be able to use our hardness ladder. The assumption we made for this master thesis is that we are going to compute it by first using fifty meters drilled. It means that we cannot use the agent in the first 50 m, as we need it to initialize the process. However, after 50 m we should be able to predict the incoming hardness. To test if this assumption is relevant we compute the median hardness value of the ten wells of chapter 3.5.1 using only data from the first fifty meters. Then we compare it to the median hardness computed using all data available.

Table 3-7 presents these computations made using Matlab code of the appendix II.

TABLE 3-7 COMPARISON BETWEEN MEDIAN HARDNESS COMPUTED BY USING ONLY FIRST 50 M DRILLED AND MEDIAN HARDNESS COMPUTED USING COMPLETE WELL PATH

Well	1	2	3	4	5	6	7	8	9	10
Median hardness on the complete well length	0.645	1.906	172.6	30.69	8.336	1274	4.077	51.05	9.476	7.697
Median hardness of first 50 m drilled	0.705	1.981	253.05	15.37	7.610	7226	2.424	64.148	14.95	6.408

Median hardness computed using only the first 50 m drilled and median hardness computed using complete well path have the same magnitude on the ten wells. Thus, determine median hardness over the first fifty meters drilled to initialize our agent is the method chosen for this master thesis. The assessment of this choice is made in chapter 8.

3.6 Determination of the forecasting model

Chapter 2.4.2 introduces the concept of extrapolative methods for time series and their utility in forecasting. For the purpose of this thesis, we focus on only two extrapolative methods: exponential smoothing and autoregressive integrated moving average (ARIMA). This chapter presents how we determine the more suitable method to use for our model.

3.6.1 An exponential smoothing technique: The Holt method

Exponential smoothing (ES) methods (Gardner and Everette, 2006) have become very popular because of their simplicity and good overall performance, but many versions exist. To find the most useful, we follow the paper of Mélard (2006) about short terms forecasting methods. He deals with an approach using short-term past trend of a time series to predict the short-term coming trend: The Holt method. This method (Holt, 2004) is useful when there is no seasonal component, which is the case with RTDD; the formation hardness is not cyclic. Indeed, by using a local short term past trend, this method can quickly adapt with a small delay to trend changes, such as a change of hardness while drilling.
We are going to adapt this method and its formulas to our hardness using the initial equations coming from the work of Holt (2004).

We start by defining notation of our time series:

- H_t represents the time series of hardness
- t is time. Time ranges from t=0 at the start of drilling to t=n when drilling operation stop
- $H(t)$ is the value of hardness at the time t

In case of simple ES applied to series with no seasonal trend, forecast at t is an invariant constant of the horizon h (Idrissi, 2014). We define the nomenclature of the forecast:

- h is the horizon of the forecast
- $\widehat{H}_T(h)$ is hardness at horizon h, it means hardness value at time t+h

$$\widehat{H}_T(h) = \widehat{H}_T \quad (3.17)$$

Equation 3.17 is here to simplify the notation.

According Mélard (2006), forecast $\widehat{H}_T$ depends of previous measurements, so:

$$\widehat{H}_T = \alpha.\widehat{H}_T + (1-\alpha).\widehat{H}_{T-1} \quad (3.18)$$

Where α is a parameter belong to [0 , 1].

Moreover, we can also write $\widehat{H}_{T-1}$ as a function of $\widehat{H}_{T-2}$ following equation 3.18. We iterate the process with all time from t=1 to t=T to get equation 3.19:

$$\widehat{H}_T = \alpha . \sum_{i=0}^{T-1} (1-\alpha)^i . \widehat{H}_{T-1} \qquad (3.19)$$

Holt (2004), reformulate this general formula to decompose $\widehat{H}_T$ into two compound, the level $\hat{a}_1$ and the slope $\hat{a}_2$. Equation 3.20 presents this decomposition:

$$\widehat{H}_T = \hat{a}_1(T) + \hat{a}_2(T) . h \qquad (3.20)$$

With:

Level:

$$\begin{aligned} \hat{a}_1(T) &= \lambda . H_T + (1-\lambda) . [\hat{a}_1(T-1) + \hat{a}_2(T-1)] \\ &= \lambda . H_T + (1-\lambda)\widehat{H}_{T-1}(1) \end{aligned} \qquad (3.21)$$

Slope: $$\hat{a}_2(T) = \mu . [\hat{a}_1(T) + \hat{a}_1(T-1)] + (1-\mu) . \hat{a}_2(T-1) \qquad (3.22)$$

Where μ and λ are two constants with a value between 0 and 1. Their determination is explained in sub chapter 3.6.3.

In order to implement it Matlab, we need the initialization process of the method, which is:

Level initialization: $$\hat{a}_1(1) = H_1 \qquad (3.23)$$

Slope initialization: $$\hat{a}_2(1) = H_2 - H_1 \qquad (3.24)$$

To resume, using equation 3.20, we can compute hardness at the date T on the horizon h, meaning we can compute hardness at time = T+h.
The rate of penetration is the factor linking the time to the depth in the forecast.

This method suggests a local linear trend for the forecast. For each new point, the trend is defined by its ordinate called level and the slope characterizing trend of the forecast line (Idrissi, 2014).

3.6.2 The autoregressive integrated moving average (ARIMA) method

The autoregressive integrated moving average method or ARIMA is another extrapolative method. Contrary to the Holt approach only based on previous observations, ARIMA considers the forecast error e_T as presented in equation 3.25 from Mélard (2006) and is adapted to our notation:

$$e_{T+1} = H_{T+1} - \widehat{H}_T(1) \tag{3.25}$$

ARIMA requires definition of two new operators:

The delay operator B: $$BH_t = H_{t-1} \tag{3.26}$$

The ordinary differentiation operator ∇ : $$\nabla H_t = H_t - H_{t-1} \tag{3.27}$$

Using the initiation to time series analysis of Mélard (2006), we describe the ARIMA model as a junction of three parts: the autoregressive method requiring an order p, an integrative operator requiring a differentiation number d and a moving average method requiring an order q. Together these parts form a random process verifying an equation in which e_T are random variables forming a white noise. In other words, model ARIMA (p, d, q) verifies the stochastic differences equation 3.28:

$$(1 - \phi_1 B - \ldots - \phi^p B^p)\nabla^d H_t = (1 - \theta_1 B - \ldots - \theta^q B^q)e_t \tag{3.28}$$

With ϕ_p is the p order autoregressive (AR) polynomial defined as:

$$\phi_p(B) = (1 - \phi_1 B - \ldots - \phi^p B^p) \tag{3.29}$$

θ_q is the q order moving average (MA) polynomial defined as:

$$\theta_q(B) = (1 - \theta_1 B - \ldots - \theta^q B^q) \tag{3.30}$$

And ∇^d is the differentiation operator ∇ applied d times. The advantage of the ARIMA process is its ease to predict and its relevance when good coefficients p, d and q are applied. Next sub chapter develops their determination.

3.6.3 Selection of the final forecasting method

We pre-selected two interesting forecasting methods. We could implement both on the agent to increase the number of predictions and then to make an average. However, it could be time consuming for the algorithm of the agent and thus irrelevant for the purpose of its real-time utilization. That is why for this master thesis we prefer to select one method.

3.6.3.1 Selection strategy

Regarding the selection process, we launch a comparative test using data from wells already used in chapter 3.5.1. We recall that Chapter 4 details data selection. Appendix III presents Matlab code used, with one code for the Holt method and one code for the ARIMA method. Figure 3-1 graphically presents the selection strategy, including the selection of proper μ, λ, p, d and q coefficients.

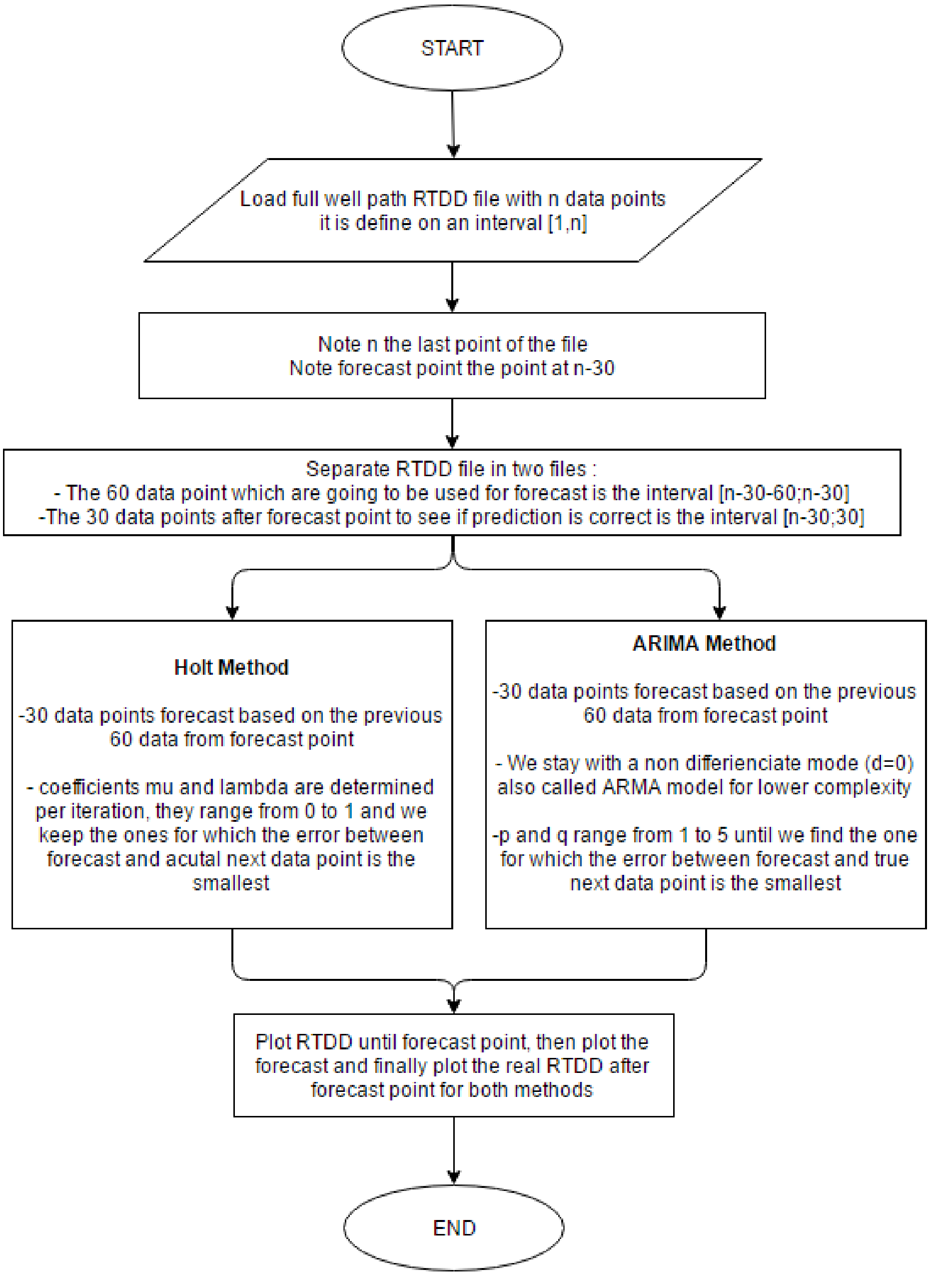


Figure 3-1 Strategy of the forecasting method selection

3.6.3.2 *Results of comparison*

We repeat the process from figure 3-1 for each of the ten wells, frequently changing the forecast point. Table 3-8 details for each well which forecast method was more precise. This table is based on the comparison of forecast from plots of both methods. Appendix III presents some of these plots for each of 10 wells.

TABLE 3-8 FORECAST ACCURACY COMPARISON BETWEEN HOLT AND ARIMA METHOD FOR 10 WELLS

	Well1	Well2	Well3	Well4	Well5	Well6	Well7	Well8	Well9	Well10
More accurate forecast method	ARIMA	ARIMA	Both inaccurate	ARIMA	ARIMA	ARIMA	ARIMA	Both inaccurate	ARIMA	Both inaccurate

For well 3, well 8 and well 10, both methods do not present reliable results (Appendix III). This issue will be discussed in Chapter 8.

It appears that the ARIMA method frequently provides more accurate prediction than the Holt method in our test panel. Thus, the strategy selected for the construction of our agent regarding the forecasting part is the ARIMA method.

4 Recovery of existing RTDD

The objective of this master thesis is to build an agent following all concepts presented in chapter 3. In order to assess its functionality, we need to be able to test it on real cases and note the number of changes in hardness it can predict compared to geological reality. The best way to proceed is to use high quality real time drilling data from real wells drilled in the North Sea. These data are being part of the Norwegian university of science and technology (NTNU) making them available to us. Statoil data were more accessible and thus were of special interest for this master thesis.

4.1 Introduction to Diskos database

In the previous project by Donne (2016), eight wells and their end of well reports were provided by Statoil and used to test the first implementation of the agent. For this master thesis, we have full access to the Diskos database thanks to the agreement between NTNU and the Norwegian National Data Repository for Petroleum data.

Diskos (2017), is an online database administered by the Norwegian Petroleum Directorate (NPD) and data are supplied by oil companies on the Norwegian Continental Shelf. The purpose of this platform is to develop, store and operate relevant petroleum data. It contains Seismic, production and well data including their end of well (EoW) reports and their RTDD.

Next sub sections of chapter 4 explains how we found, select and processed RTDD files from Diskos. However, the access to the platform and its data is limited and data are confidential. Thus, as this master thesis is public, we will not presents screenshots from the platform but only explain Diskos functionalities. Likewise, we will not directly write names of selected wells but just number them.

4.2 Wells selection inside the database

Diskos database contains hundreds of wells along the Norwegian continental shelf. For each of them there are tens of pdf, doc, excel and other types of files. The first step we made to select which data could be relevant for the master thesis was to find the largest RTDD files along the database. Indeed, a large file means small or no compression and large amount of data present that can be the sign of a high quality data file. Hopefully the platform offers the possibility to sort data per size. According to this first sorting out, we downloaded 157 RTDD files and available reports mainly from Statoil Petroleum but also from Total Exploration and Production Norge, Talisman Energy, Statoil Hydro, Norsk Hydro, Esso Exploration and production and British Petroleum (BP) Norge. All file sizes were above a hundred megabyte.

4.3 Cleaning of raw files

The 157 files downloaded were directly uploaded in Diskos by operator and service companies. These companies use different software and methods to record real time data. Thus, there is no consistency along them and we consider these data as raw data files we have to clean.

4.3.1 Different types of files present

This raw files database initially contained six different types of files:

- .ASC: ASCII Armored File is an encryption program used for secure communication. It contains a digitally signed message and may store plain-text written information, as well as binary information encoded as text (fileinfo[1], 2017)
- .TIF: Tagged Image File is an image file saved in a high-quality graphics format. (Fileinfo [2], 2017). The problem is that we need a specific Schlumberger software to remove the encryption and open it
- .TXT: Text File is a common standard text document that contains unformatted text. Openable by most editing software programs (fileinfo[3], 2017)
- .LAS: LIDAR Data Exchange Files are stored in a binary format, collected by optical remote sensors and created to exchange data between data providers and consumers (fileinfo [4], 2017).
- .LIS: SQR Output File is a Report or output file generated by a Structured Query Reporting (SQR) program (fileinfo [5], 2017).
- .CSV: Comma Separated Values File is file commonly used by spreadsheet programs such as Microsoft Excel. It contains plain text data sets separated by commas. It is often used for transferring data between databases (fileinfo [6], 2017).

Due to the large amount of RTDD files present, their diversity, and the fixed schedule of the master thesis, we decide to focus only on the most represented type of file: ASC type. It is the one used in 123 of the downloaded files. The objective is to make these files readable by the log viewer coded in Matlab by Anisa Noor Corina, another NTNU student.

Following the governing documentation she wrote (Noor Corina, 2016), ASC files, should have the shape presented in the figure 4-1.

```
WELL : 33/XX-X-XX
COMP : COMPX
FIELD : FIELD X
COUNTRY : NORWAY

  "DATE","TIME","LOGTIME","ACTC","BDIA","DBTM","DBTV","DMEA
","DVER","RSU"
dd-mmm-yy,hh:mm:ss,s,,IN,M,M,M,M,M/S
"01-Jan-07","17:00:00",852138000,-9999,-9999,4516.958984375
,-9999,4595.0146484375,1191.0538890000003,-9999"01-Jan-07",
"17:00:05",852138005,-9999,-9999,4516.958984375,-9999,4595.
0146484375,1191.0538890000003,-9999"01-Jan-07","17:00:10",8
52138010,-9999,-9999,4516.958984375,-9999,4595.0146484375,1
191.0538890000003,-9999"01-Jan-07","17:00:15",852138015,-99
99,-9999,4516.958984375,-9999,4595.0146484375,1191.05388900
00003,-9999"01-Jan-07","17:00:20",852138020,-9999,-9999,451
6.958984375,-9999,4595.0146484375,1191.0538890000003,-9999
```

Figure 4-1 Standardized ASC-file to be readable by the log viewer (Noor Corina, 2016)

Three points characterize this standardized format:

- A header contains well information, this part is not compulsory. For instance, well, company (COMP), date of start and date of end.
- A curve information section in which both curves mnemonic names and units must be begin and ended by quotation mark ("). The delimiter between each curve mnemonic name and unit is a comma (,). The curve units are written directly one line after the mnemonic names line. The curve information must contain and begin with DATE and TIME (Noor Corina, 2016).
- The log data section composed of RTDD, separated by a comma. Only Date and Time come first and begin and end by quotation mark (").

As we have 123 different files to clean, it could be interesting to create a program to do it automatically by looking at recurrent pattern raw files.

4.3.2 Different types of pattern

After opening all 123 different .ASC RTDD files using a text editor, we notice that they are built in six different formats. Table 4-1 summarizes the difference between each pattern and the standardized format. For each of these formats, an example is given in the Appendix IV.

ASC format	Definition	Number of files concerned
0	Standardized format	41
1	Header absent, quotation mark (") in Mnemonics is absent, space is used as separator instead of coma and time and date format is missing	10
2	Header absent, a unwanted info column is present, quotation mark (") in Mnemonics is absent, Units before data is absent, space is used as separator instead of coma and time and date is expressed in UNIX	37
3	Similar at format 2, moreover, mnemonics are hidden in the unwanted info column	4
4	Similar at format 2, except that header is present but settled between two lines of stars (****)	9
5	date and time are in UNIX	22

Table 4-1 Description of different format of ASC files present in the Diskos database

4.3.3 Manual cleaning methodology

Create an automatic program to clean a certain type of file could be challenging and time consuming and thus could be irrelevant when a small amount a file is concerned. Thus, for format type 1, 3 and 4, we decide to clean raw files concerned manually. The cleaning is done using notepad application and Microsoft Excel. Table 4-2 briefly summarizes procedures involved.

TABLE 4-2 PROCEDURES TO CLEAN RAW ASC FILES WITH FORMAT 1, 3 AND 4

Format	Procedure in Excel	Procedure in Notepad
1	-Put commas instead of spaces in all rows, -Add quotation marks between time and date -Separate time and date in two columns	-Rewrite curve information with proper mnemonics -delete parenthesis between units
3	-launch code created to clean type2 (presented further in chapter 4.3.4)	-Delete the carbon information presents in the unwanted info bloc
4	-launch code created to clean type2 (presented further in chapter 4.3.4)	-Remove star (****) lines in the header -Rewrite header headlines with capital letters only

4.3.4 Automatic cleaning methodology

According to Table 4-1 in chapter 4.3.2, ASC format type 2 and 5 concerns 59 files in total. It is faster to create an automatic algorithm for this amount of data rather than to do it manually. Figure 4-2 presents flow chart of both cleaning program coded in Matlab with help of Anisa Noor Corina. Her code is presented in the Appendix V

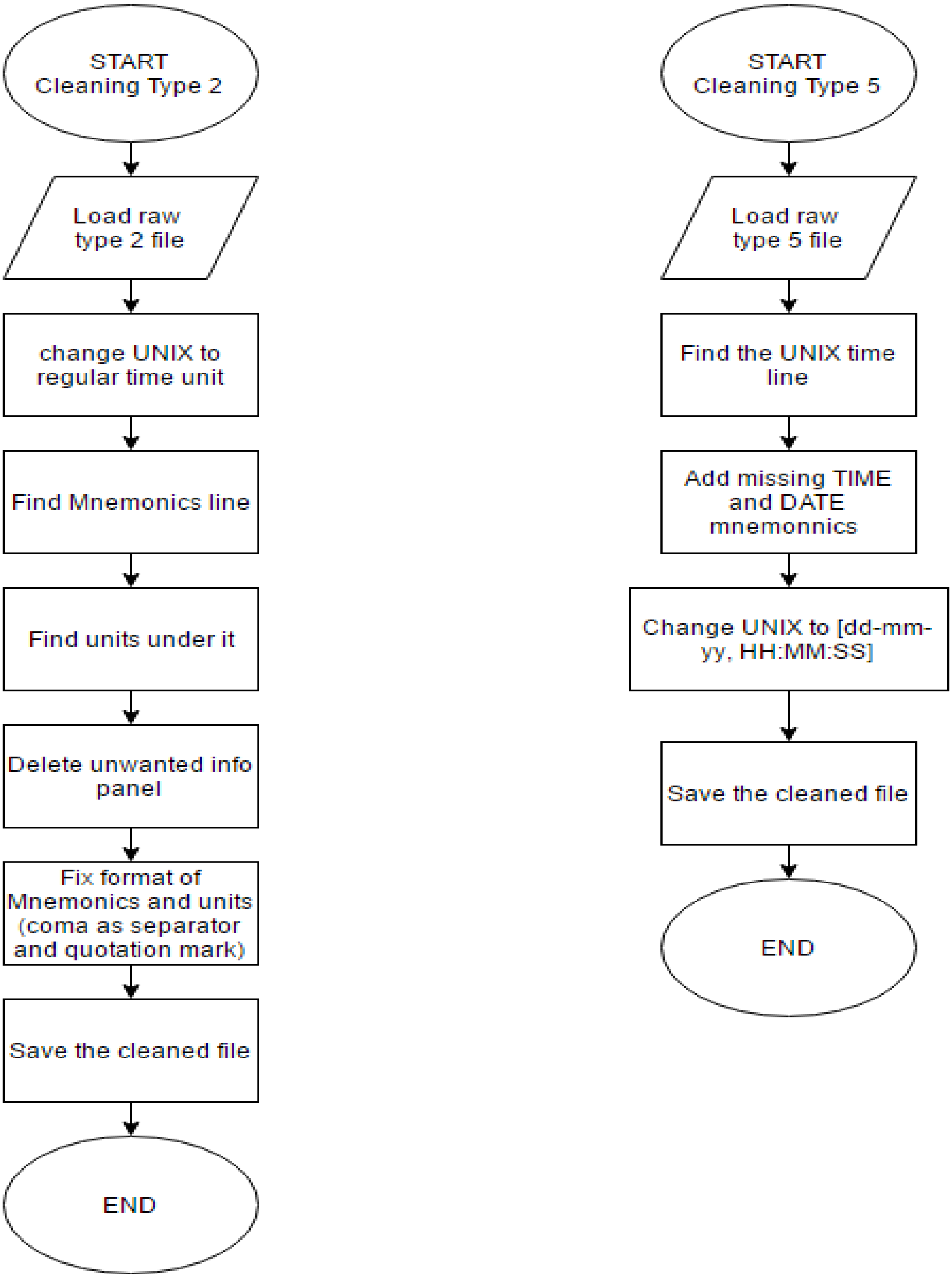


Figure 4-2 Algorithms of cleaning programs for ASC type 2 and ASC type 5

4.4 Selection of 10 wells for the study

After the cleaning process, we can properly view the 123 ASC files in the log viewer. However, the study of such a number of wells is too time consuming within the period of this master thesis. Thus, we decide to search for more relevant wells for our studies. As explained in the chapter 3.5.1, we want to focus on wells from the same surroundings. Moreover, to get as many precise information as possible, it would be preferable to use RTDD from wells having geological, end of wells or drilling reports available. We finally focused on the Statfjord field. Indeed, many high quality data are available in this area and it is easy to have access to ten surrounding wells in this sector including their geological reports. More precisely, we focused on five wells in the block 33/9-A of Statfjord (33/9-A-18, 33/9-A-23, 33/9-A-34, 33/9-A-35, 33/9-A-38) and 5 wells in the block 33/12-B of Statfjord (33/12-B-13, 33/12-B-16T2, 33/12-B-20, 33/12-B-30, 33/12-B-31). Figure 4-3 shows their emplacement on the Norwegian continental Shelf.

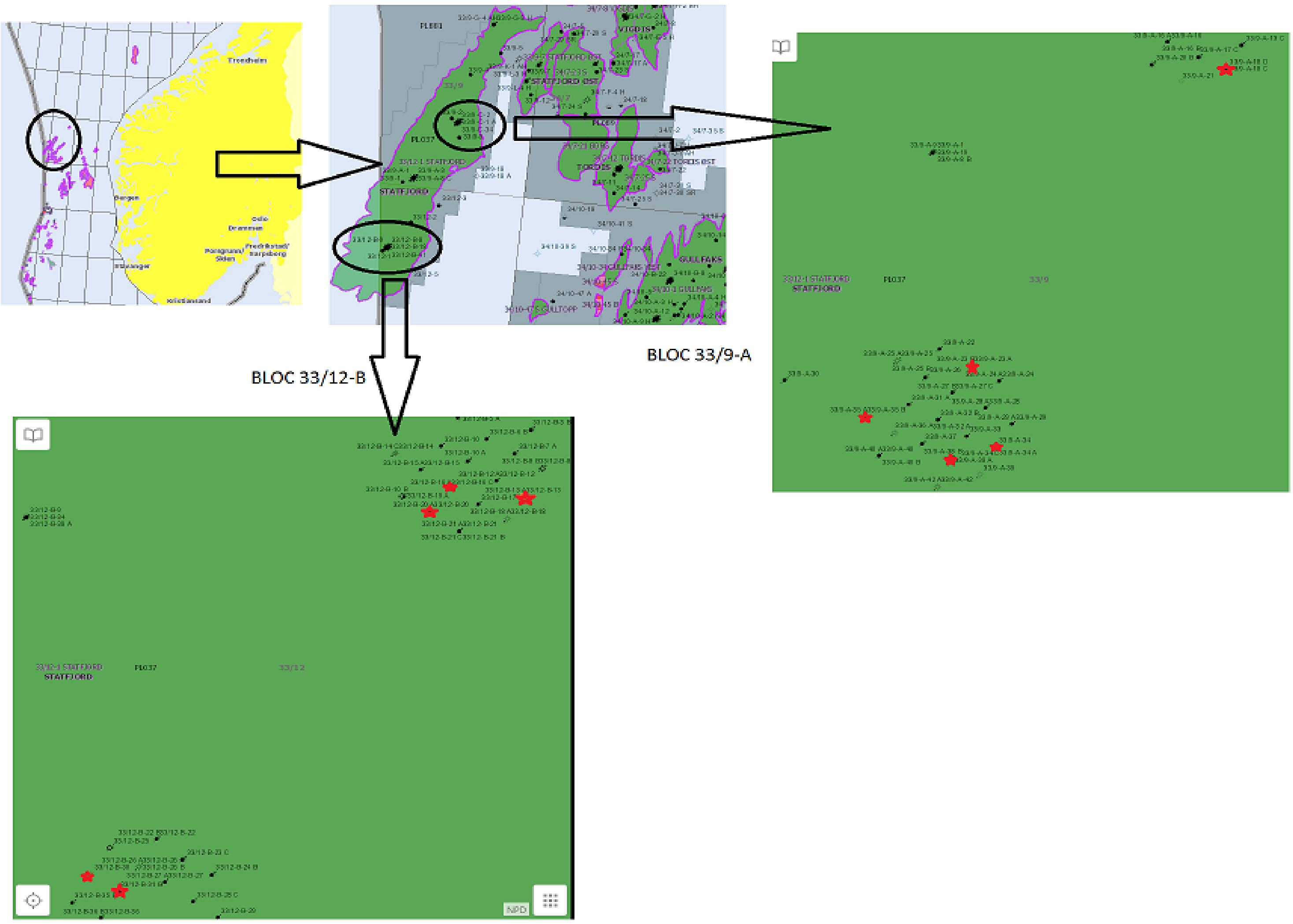


Figure 4-3 Emplacement on the Norwegian continental Shelf of the 10 selected wells, their position in each block is marked by red stars

The rest of this chapter will discuss in detail the geological parameters in all these wells. This information are public and available in factpages.npd.no, the portal from the Norwegian oil directorate containing information regarding the petroleum activities on the Norwegian continental shelf. However, to preserve confidentiality of RTDD coming from Diskos, name of wells is replaced by a random numeration from 1 to 10 in tables and plots.

All wells studied are linked to the production platform Statfjord A and B operated by Statoil petroleum A.S at 145m above the seabed. They are situated in blocks 33/9 and 33/12 of the Statfjord field discovered in 1974 (Factpages [1], 2017) located in the Tampen area in the North Sea, on the border between the Norwegian and the U.K sectors. The Statfjord reservoirs (Factpages [1], 2017) are at a depth of 2500-3000 meters goes from the Brent group formation formed in the middle Jurassic to the Lunde formation created in the late Triassic. All information on the cores we have comes from the exploration well 33/12-1, drilled in 1974 (Factpages [2], 2017).

Table 4-3 presents geological formations of each well:

TABLE 4-3 GEOLOGICAL FORMATIONS CROSSED PER EACH OF THE 10 WELLS SELECTEDIS MARKED BY X

Lithostratigraphy of the interest zone		Wells									
Top Depth	Geological formation	1	2	3	4	5	6	7	8	9	10
[m]	-	-	-	-	-	-	-	-	-	-	-
172	Nordland GP	/	/	/	/	/	/	/	/	/	/
1647	Rogaland GP	/	/	/	/	/	/	/	/	/	/
1647	Balder FM	/	/	/	/	/	/	/	/	/	/
1705	Sele FM	/	/	/	/	/	/	/	/	/	/
1876	Shetland GP	/	/	/	/	/	/	/	/	/	/
2398	Cromer Knoll GP	X	/	/	/	X	/	/	/	/	/
2402	Viking GP	/	X	X	X	/	X	X	/	/	X
2409	Brent GP	/	X	X	X	/	X	/	X	X	X
2570	Dunlin GP	/	/	X	/	/	X	X	X	/	X
2836	Statfjord GP	/	/	/	/	/	X	X	/	/	/
2960	Lunde FM	/	/	/	/	/	/	X	/	/	/

We detect six geological formations important to study using available information in Factpages website: Cromer Knoll GP, Viking GP, Brent GP, Dunlin GP, Statfjord GP and Lunde FM.

- **Cromer Knoll Group**

According the Norwegian Petroleum Directorate (Factpages [3], 2017). The thickness of the group varies considerably since the sediments were deposited in response to an active Late Jurassic tectonic phase. This group consists mainly of fine-grained, argillaceous, marine sediments with a varying content of calcareous material. Regarding deposition, marlstones become the more dominant lithology in both the upper and lower parts of the group while sandstones are more common in the upper part of the group.

- **Viking Group**

Data in Factpages [4], 2017, indicates that the group consists of dark, grey to black, marine mudstones, claystones and shales. These sediments are replaced by sandstones and occasionally conglomerates as in Intra Draupne Formation sandstone and Intra Heather Formation sandstone. Five define within the Viking group. The Heather and Draupne Formations are the most important.

- **Brent Group**

The group consists of grey to brown sandstones, siltstones and shales with subordinate coal beds and conglomerates. The group is divided into five formations. These are the Broom (base), Rannoch, Etive, Ness and Tarbert (top) formations (Factpages [5], 2017).

- **Dunlin Group**

The group consists mainly of dark to black argillaceous marine sediments, but in the marginal areas of the basin marine sandstones are well developed at several stratigraphic levels and can extend a considerable distance into the basin. The sandstones are white to light grey, very fine to medium grained and generally well sorted (Factpages [6], 2017).

- **Statfjord Group**

The group exhibits a transition from continental to shallow marine sediments. In the type well area it is a transitional "coarsening upward" sequence in the basal parts consisting of grey, green and sometimes red shale interbedded with thin siltstones, sandstones and dolomitic limestones. The top part of the group consists of thick, white to grey, fossiliferous and glauconitic sandstones (Factpages [7], 2017).

- **Lunde Formation**

The succession of the Lunde Formation is dominated by very fine to very coarse-grained sandstones, claystones, mudstones, shales and marls. The sandstones are mainly white while the fine-grained lithologies are generally red, green and grey-green. This contrasts to the pre-Alke Formation succession, in which all lithologies are generally red stained. Especially in the upper part of the succession the claystones and mudstones are non stained (Factpages [8], 2017).

5 Selection of relevant RTDD windows for agent development and testing

We managed to get ten relevant ASC files cleaned to be opened by the log viewer. These real-time drilling data are used to monitor more than sixty parameters at the same time. However, for this master thesis and the hardness computation, they are not all needed. According to the strategy used by Donne (2016), we can move from ASC format to a matlab (.mat) format, which is the database format used by Matlab. The reason is that it allows faster changes when directly uploaded in Matlab. Then we remove unwanted data from the sixty given, and just keep the ones useful for the agent according our methodology.

5.1 Creation of the .mat files for the log viewer

Chapter 2 introduced relevant information about what is happening to the drilling parameters when the drill bit encounters hard formation. As stated by Donne (2016):

- Decrease in ROP, this is read to the ROP data directly or by looking at the steepness of the block position (BPOS) curve.
- Decrease in revolution per minute (RPM)
- Increase in weight on bit (WOB)

Moreover, chapter 3 provided all equations used by the agent to provide expected results. In other words, the agent must access from the RTDD bit measured depth (DBTM), measured depth of the well (DMEA), time; ROP, RPM and WOB. This is all the information we want to keep from our starting 60 parameters monitored. The technical procedure to simplify RTDD is detailed in Appendix VI (Donne, 2016).

Figure 5-1 presents format of a standardized .mat file after cleaning process:

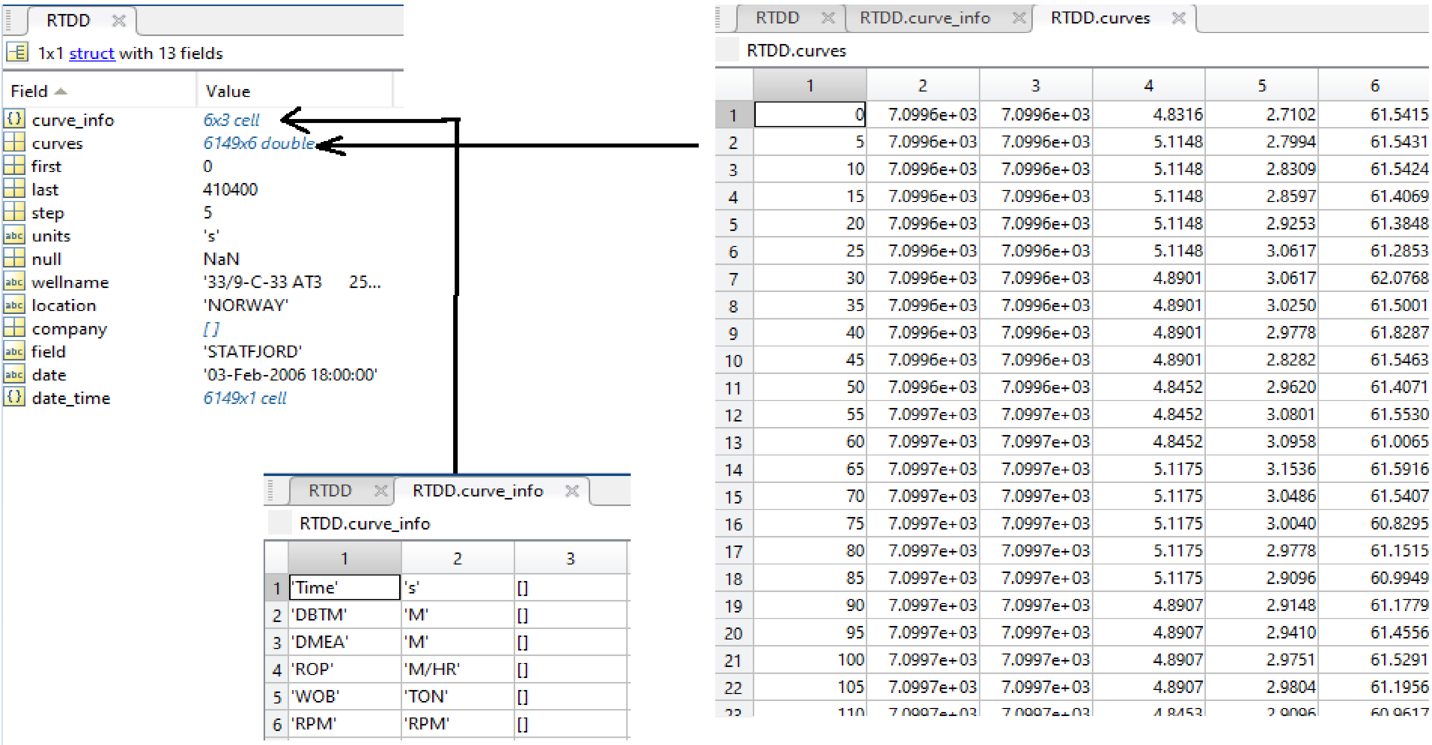


Figure 5-1 RTDD data workspace after simplification. The left window is the global workspace window and present all general information about the RTDD such as date or record, location of the well, well name, field location… The bottom window presents data available for this RTDD file, here we have Time, DBTM, DMEA, ROP, WOB and RPM. Finally, the right windows are the data. One row represents one point in time of record, each column is used per information presented in the bottom window. Column 1 is Time, column 2 is DBTM, column 3 is DMEA… (Donne, 2016)

5.2 Selection of interesting depth interval using geological reports

We found ten RTDD files to study, however each one consists in several hundreds of thousands of data points. It is interesting, instead of using the complete wells path, to find more relevant depth intervals where change in formation hardness is stated. Theses depth intervals will be our test intervals for the Matlab agent. This search for hardness is based on geological reports provided by the geological engineering department of several companies (Paulsen, 1995; Flotre and Aakvik, 1993; Flotre and Quale, 1993; Engineering dpt., 1980; Engineering dpt., 1986; Engineering dpt., 1985; Engineering dpt., 1984; Engineering dpt., 1983; Engineering dpt., 1982; Engineering dpt., 1985). As these reports coming from Diskos are confidential, we keep hiding names of wells involved but table 5-1 and table 5-2 present results of this search. Each compound of the table contains information about a small depth window in which a specific lithology was encountered and describe its hardness classification, its rock composition and in which geological formation the event occurred. Distance are written in measured depth (MD).

TABLE 5-1 SPECIFIC GEOLOGICAL FORMATIONS ENCOUNTERED DURING DRILLING FOR WELL 1 TO WELL 5

	Well 1		Well 2		Well 3		Well 4		Well 5	
Top Window	**2553**		**3550**		**3100**		**4300**		**2500**	
Top | bottom	2559	2561	3550	3557	3164	3166	4379	43881	2509	2510
Type | formation	Soft	Cromer	mod hd	-	xtrem hard	cromer	mod hd	brent-1	hard	-
lithology	mrlst, sst		lst		calcitucite		coal		clst	
Top | bottom			3557	3566	3186	3188	4389	4391	2604	2607
Type | formation			mod hd	viking	hard	cromer	mod hd	brent-1	hard	-
lithology			clst		calcitucite		lst		clst	
Top | bottom			3595	3597	3193	3196	4400	4402	2688	26992
Type | formation			mod hd	brent-1	hard	viking	soft	brent-1	xrem soft	-
lithology			clst		claystone		sst		mrlst	
Top | bottom			3650	3653	3196	3197	4458	4462	2804	2807
Type | formation			mod hd	brent-2	hard	viking	soft	brent-1	xrem sot	-
lithology			clst		claystone		sltst		mrlst	
Top | bottom			3688	3692	3207	3209	4483	4486	2873	2891
Type | formation			hd	brent-2	hard	viking	soft	brent-2	hard	-
lithology			coal		claystone		sltst		clst	
Top | bottom			3670	3680	3268	3272	4528	4532		
Type | formation			mod hd	brent-2	hard	brent-4	soft	brent-2		
lithology			clst		claystone		sltst			
Top | bottom			3725	3728	3278	3282	4550	4555		
Type | formation			hd	brent-3	hard	brent-4	soft	brent-2		
lithology			clst		claystone		clst			
Top | bottom			3728	3750	3398	3400	4555	4560		
Type | formation			mod hd	brent-3	hard	dunlin-1	soft	brent-2		
lithology			clst		claystone		sst			
Top | bottom							4603	4605		
Type | formation							xtrem soft	brent-2		
lithology							clst			
Top | bottom							4663	4666		
Type | formation							soft	brent-3		
lithology							clst			
Top | bottom							4692	4698		
Type | formation							soft	brent-3		
lithology							sltst			
Top | bottom							4760	4765		
Type | formation							soft	brent-4		
lithology							sst			
Top | bottom							4799	4800		
Type | formation							soft	brent-4		
lithology							clst			
Bottom Window	**2637**		**3777**		**3400**		**4800**		**3000**	

TABLE 5-2 SPECIFIC GEOLOGICAL FORMATIONS ENCOUNTERED DURING DRILLING FOR WELL 6 TO WELL 10

	Well 6		Well 7		Well 8		Well 9		Well 10	
Top Window	**2700**		**3000**		**2600**		**4338**		**2650**	
Top | bottom	2926	2927	3048	3053	2625	2635	4348	4353	2685	2692
Type | formation	hard	satfjord-2	sft	dunlin-1	sft	brent-5	hd	brent-1	hd	brent-2
lithology	clst		clst		sst		coal		coal	
Top | bottom	2927	2930	3062	3068	2645	2650	4370	4380	2709	2717
Type | formation	hard	statfjord-2	sft	dunlin-3	mod sft	dunlin-1	soft	brent-1	hd	brent-2
lithology	sst		clst		sh		clst		sltst	
Top | bottom	2955	2957	3185	3190	2674	2688	4400	4410	2725	2732
Type | formation	xtrem hd	statfjord-2	sft	statfjord-1	mod hd	dunlin	soft	brent-2	hd	brent-3
lithology	sst		clst		sh		clst		clst	
Top | bottom	2958	2960	3240	3253	2712	2721	4410	4415	2872	2875
Type | formation	hard	statfjord-2	hard	statfjord-2	mod hd	dunlin	hd	brent-2	hd	brent-5
lithology	clst		sltst		sh		coal		sltst	
Top | bottom	3032	3034	3325	3335	2726	2737	4430	4440	2895	2900
Type | formation	xtrem hd	statfjord-3	soft	statfjord-3	mod hd	dunlin	hd	brent-2	sft	dunlin-1
lithology	clst		clst		sh		calcitucite		clst	
Top | bottom			3360	3365	2757	2766	4445	4450		
Type | formation			hard	statfjord-3	mod hd	dunlin	mod hd	brent-2		
lithology			clst		sltst		sltst			
Top | bottom			3400	3405						
Type | formation			soft	statfjord-3						
lithology			lst							
Top | bottom			3430	3450						
Type | formation			soft	statfjord-3						
lithology			clst							
Top | bottom			3520	3528						
Type | formation			soft	statfjord-3						
lithology			sst							
Top | bottom			3530	3532						
Type | formation			mod sft	statfjord-3						
lithology			clst							
Top | bottom			3570	3575						
Type | formation			mod hd	statfjord-3						
lithology			lst							
Top | bottom										
Type | formation										
lithology										
Top | bottom										
Type | formation										
lithology										
Bottom Window	**3100**		**3640**		**2800**		**4700**		**2900**	

Table 5-3 explains colors and abbreviations used in tale 5-1 and 5-2:

TABLE 5-3 LEGEND TABLES 5-1 AND 5-2

Rocks information	Abbreviation used	Color code Geological formations	Color code hardness classification
sandstones	sst	Cromer	Xtrem Hard
marlstones	mrlst	Viking	Hard
moderate	mod	Brent	Mod hard
hard	hd	Dunlin	Mod Soft
limestone	lst	Statfjord	Soft
claystone	clst	Lunde	Xtrem Soft
silstone	sltst		
shale	sh		
coal			
calcitucite			

All these tables provide good support to establish cases to test the agent in chapter 7. Appendix VII presents two others tables presenting depths at which all these wells cross geological formations presented in the chapter 4.4.

5.3 Detection of changes in formation hardness and hard/soft stringers graphically using the log viewer

If the geological report is not available, it is also possible to use graphical results provided indirectly by ROP seen in the log viewer to detect changes in formation hardness. This technique was used during previous project (Donne, 2016). To detect hard formation, we must detect area where the RPM decreases, where block position (BPOS) curve becomes less steep, and WOB increases simultaneously. And likewise, to detect soft formations, RPM should stay constant or increase a bit while the WOB decreases a bit and the BPOS curve becomes steeper due to an increase in the penetration rate. Regarding hard of soft stringers, then we have to detect a high increase or decrease of hardness on a short depth interval. Finally, for laminated formations, we must see an alternation hard-soft-hard or hard-soft-hard in the graphical analysis. Appendix VIII presents the graphical analytical study made on the ten wells.

6 Design of the agent using Matlab

Model of the agent is ready, and test cases are ready, we have now to develop the program using the software Matlab. This chapter is an extension of Donne (2016).

6.1 Algorithm, inputs and outputs of the agent

6.1.1 Algorithm

The flow chart of the agent is given in figure 6-1. Notice that in the flow chart, the functions “launch ARMA test”, “launch warning message” and “launch hardness review” have their own flow chart given in Appendix XIX. Explanations about executing functions come in chapter 6.2.

6.1.2 Inputs

From the available real time drilling data, the agent extracts:

- Time
- Rate of penetration (ROP)
- Weight on bit (WOB)
- Rotation per minute (RPM)
- Depth of the well measured (DMEA)
- Depth of the bit measured (DBTM)

6.1.3 Outputs

In the previous specialization project (Donne, 2016), the agent provided a plot of hardness versus depth: However, previous agent was not designed to predict hardness and thus, output plot was unclear. We decide to modify outputs according to new agent tasks:

- The first output is a table providing hardness classification encountered and corresponding depth intervals in measured depth (MD).
- Second output is linked to the new forecasting tasks. It is a mild warning function to the user of the software. This function send a message to the driller when a high change in hardness is forecasted in the next three minutes. For the software, three minutes corresponds to a prediction of next 36 RTDD points, as a new point is recorded every five seconds. The user can see this message as a piece of information leading to further advice to prevent potential errors.

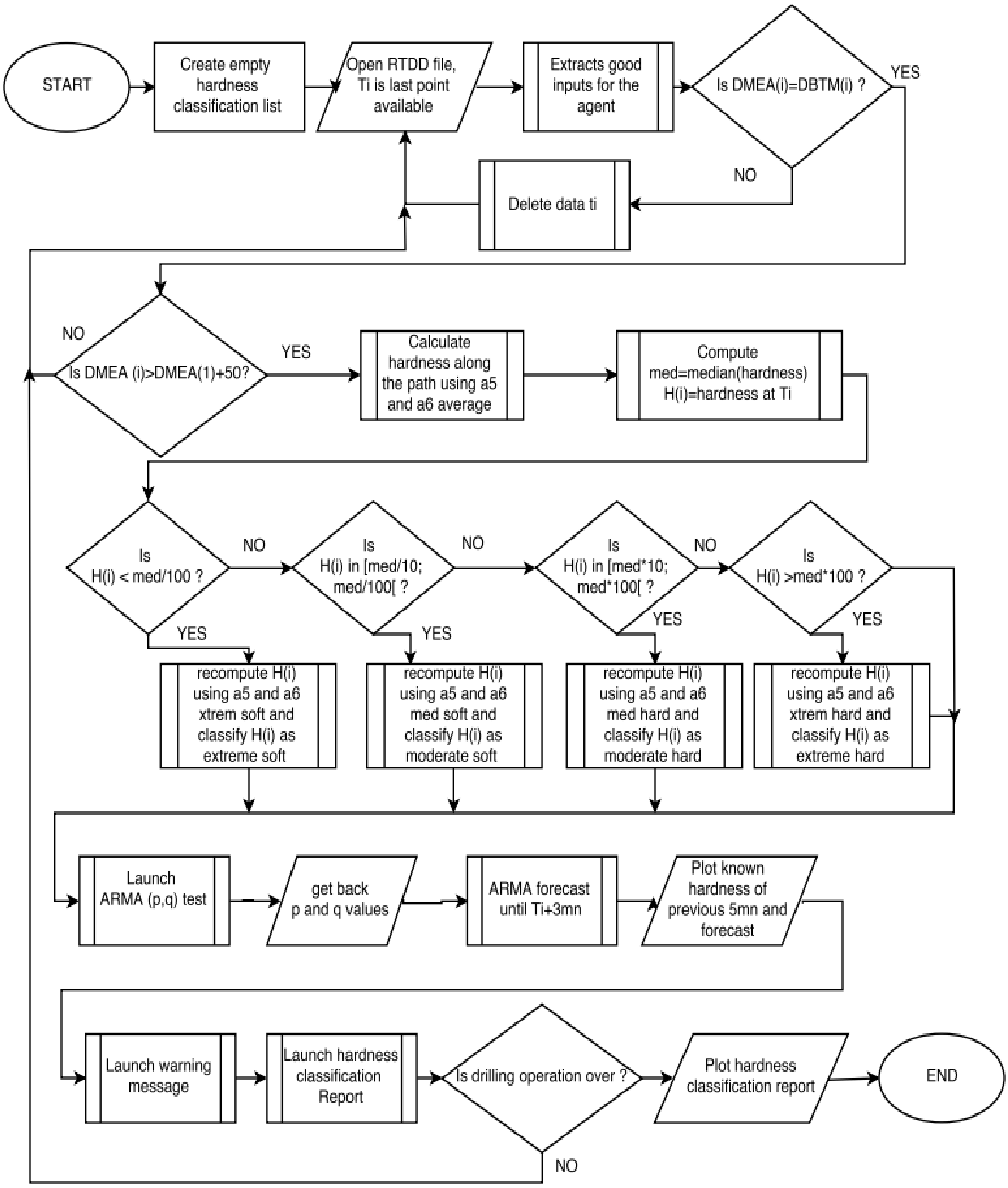


Figure 6-1 Flow chart of the agent

6.2 Implementation in Matlab

Present chapter explains how the agent should be used. It was mainly developed by Donne (2016) project.

6.2.1 File format needed to run the agent

To be able to use it, the user must upload a RTDD file in the appropriate standard .mat format. The standardized RTDD file for this agent use a workspace consisting of curve info and curve data containing all parameters needed for computation and placed in good order. For now, the agent only works if the file is pre-prepared in the format showed in figure 6-2. The column has to appear in this specific order (Time, DBTM, DEAM, ROP, WOB and RPM).

RTDD | RTDD.curve_info

RTDD.curve_info

	1	2	3
1	'Time'	'yyyy-mm-...	[]
2	'DBTM'	'm'	[]
3	'DMEA'	'm'	[]
4	'ROP'	'm/h'	[]
5	'WOB'	'MTON'	[]
6	'RPM'	'rpm'	[]

Figure 6-2 Curve info of the RTDD including all the inputs needed time, DBTM, DMEA, ROP, RPM and WOB (Donne, 2016)

To make the reading clear, the coding is done in Matlab using flow charts. Appendix X (chapter 13.10.1) contains the main code of the main algorithm of the agent, which was introduced in the previous chapter. Then Appendix X (chapter 13.10.2) contains the code of the sub program "ARMA Test" which has its flow chart in Appendix IX (chapter 13.9.1). Appendix X (chapter 13.10.3) contains code of the sub program "Warning message" which has its flow chart in Appendix IX (chapter 13.9.2). Finally, Appendix X (chapter 13.10.4) contains the algorithm of the sub program "Hardness Report" which has its flow chart in Appendix IX (chapter 13.9.3).

6.2.2 Implementation of the main code

Initially, the agent is supposed to work in real time, receiving RTDD every five seconds. However, in this master thesis, we load data directly from the full well path. To simulate the real utilization, the agent will automatically load first 90 data points (to initialize computation of the median hardness value as stated in chapter 3), then run its full algorithm and then start again taking one by one point into consideration until all data available are used. Then, to make sure we are actually studying a drilling process and not another operation on the well, ROH or RIH for example, the agent compare on the depth of the well (DMEA) and the depth of the bit (DBTM), they have to be the same while drilling. If these values are different, then the agent delete last RTDD recorded and keep going on.

6.2.3 Implementation of the “ARMA Test” sub program

Chapter 3 introduced the ARIMA method for forecasting and the three constants p, q and r needed. We assumed that due to the complexity of the model, we take a differentiation order equal to 0, thus, r is automatically equal to 0 and the method to work need a coefficient p and q. The purpose of the ARMA Test sub program is to find them. We know that p and q affect the validity of the model. To tackle this challenge, we are going to re-compute them at each new data using the last data available. To resume the method:

- We want to use data from time = [Ti-60; Ti] to forecast hardness until Ti+30, where Ti represents last time entered in the agent.
- We first use data from time = [Ti-90; Ti-30] to forecast hardness from Ti-30 to Ti using several ARMA models ranging p and q from 1 to 5.
- As we know real hardness in [Ti-30,Ti], we can find which ARMA is the most precise in its forecast
- We use p and q of best model and data from time = [Ti-60; Ti] to forecast final hardness until Ti+30, assuming that there was a local hardness trend and thus that the model is still precise enough with these coefficients.

6.2.4 Implementation of the “launch warning message” sub program

The “launch warning message” sub program is there to simulate the utilization of the agent during real drilling operations, it is the output stated in chapter 6.1.3. This sub program works by displaying messages about classification of the median hardness forecast coming in next three minutes. Of course, this is linked to the rate of penetration; if the driller increases or decreases his speed then the forecast is wrong. It is because the forecast model is based on times series and such provides time forecast and not depth forecast. We choose to stop the forecast after three minutes because of a compromise we have to do between forecast accuracy and long term prediction. We will explain this in chapter 8.

6.2.5 Implementation of the "Hardness report" sub program

The "Hardness report" sub program is there to provide an easy readable hardness classification versus depth of the well which has just been drilled. The report is a three-column table. The first and second column provides starting and ending depths of a formation in measured depth (MD) while the third column describes its hardness classification following table 6-1.

TABLE 6-1 HARDNESS CLASSIFICATION IN THE HARDNESS REPORT

Value in "Hardness Report"	Corresponding Hardness classification
1	Extreme soft
2	Moderate soft
3	Average
4	Moderate Hard
5	Extreme Hard

6.2.6 Implementation of the outputs of the agent

All implementations stated in previous sub chapter ensure a good functionality of the agent; however, it is limited in terms of test possibilities. Thus, we have to add some functionalities; they are not part of the agent itself but just to assess more precisely its efficiency.

6.2.6.1 Regarding the Hardness report

Chapter 6.2.5 explained that the agent provides a table as an output. But it is challenging to compare quickly this table with our test cases from chapter 5. Thus, we add a plot function in the sub program hardness report which plots hardness classification versus depth found by the agent and the one found by our graphical and analytical study made in chapter 5. With this update, we can quickly look at the shape of both curves and see if they diverge or correlate.

6.2.6.2 Regarding the forecasting task

For every RTDD points, the agent will provide a forecasting plot on the next 30 coming points. With several thousands of points per well and for 10 wells, it means a massive number of plots to study. This study is too time consuming for this thesis but we still can find a way to look at the way the agent operates. Instead of plotting the forecast at each iteration, we store the figure of the forecast. At the end of all the iterations that simulate the end of drilling operation, we create a movie using all these screenshot. By playing this movie we are able to see the forecast moving at the same time as the agent receive new data.

6.3 Results obtained for 10 wells

Regarding the two original outputs of the agent:

- Figure 6-3 presents the shape of hardness classification report provided by the agent for the first well and explains how it should be read. Full hardness classification reports for the 10 wells are too long to be put in this report even in the appendix, thus they will be in the .zip folder attached to this master thesis. However, chapter 6.4 will presents plots of hardness classification versus depth found by the agent and the one found by our graphical and analytical study made in chapter 5 for a comparison study.
- Regarding the forecasting output, as stated in previous chapter, it is impossible to presents here videos. But the .zip folder attached to this master thesis contains all videos records made for the 10 wells.

	1	2	3
1	2.5537e+03	2.5538e+03	5
2	2.5538e+03	2.5538e+03	3
3	2.5538e+03	2.5538e+03	4
4	2.5538e+03	2.5554e+03	5
5	2.5554e+03	2.5576e+03	3
6	2.5576e+03	2.5580e+03	4
7	2.5580e+03	2.5585e+03	3
8	2.5585e+03	2.5589e+03	2
9	2.5589e+03	2.5591e+03	1
10	2.5591e+03	2.5666e+03	2
11	2.5666e+03	2.5670e+03	3

Figure 6-3 Layout of the Hardness classification report provided by the agent, the first column give the starting depth of a certain formation, column 2 gives its end and column 3 its classification, all depths are in MD and in meters.

6.4 Comparison with the graphical study of RTDD

As explained in the chapter 6.2.6.1, we want to relate hardness classification provided by the agent to the one found graphically in chapter 5. By digitalizing our graphical study in a form understandable by Matlab we can plot both reports (from the agent and analytical) on the same figures and such have a quick understanding of the relevance of data provided by the agent. Appendix XI presents the code we add in the “hardness report” sub program to digitalize our graphical study. The agent will be assessed in chapter 7 but we can still consider it as efficient if it found the same formation hardness classification we found in chapter 5. Figure 6-4 to 6-13 presents these comparisons. For all plots, the hardness classification scale used is 1 for extreme soft, 2 for moderate soft, 3 for average, 4 for moderate hard and 5 for extreme hard formation.

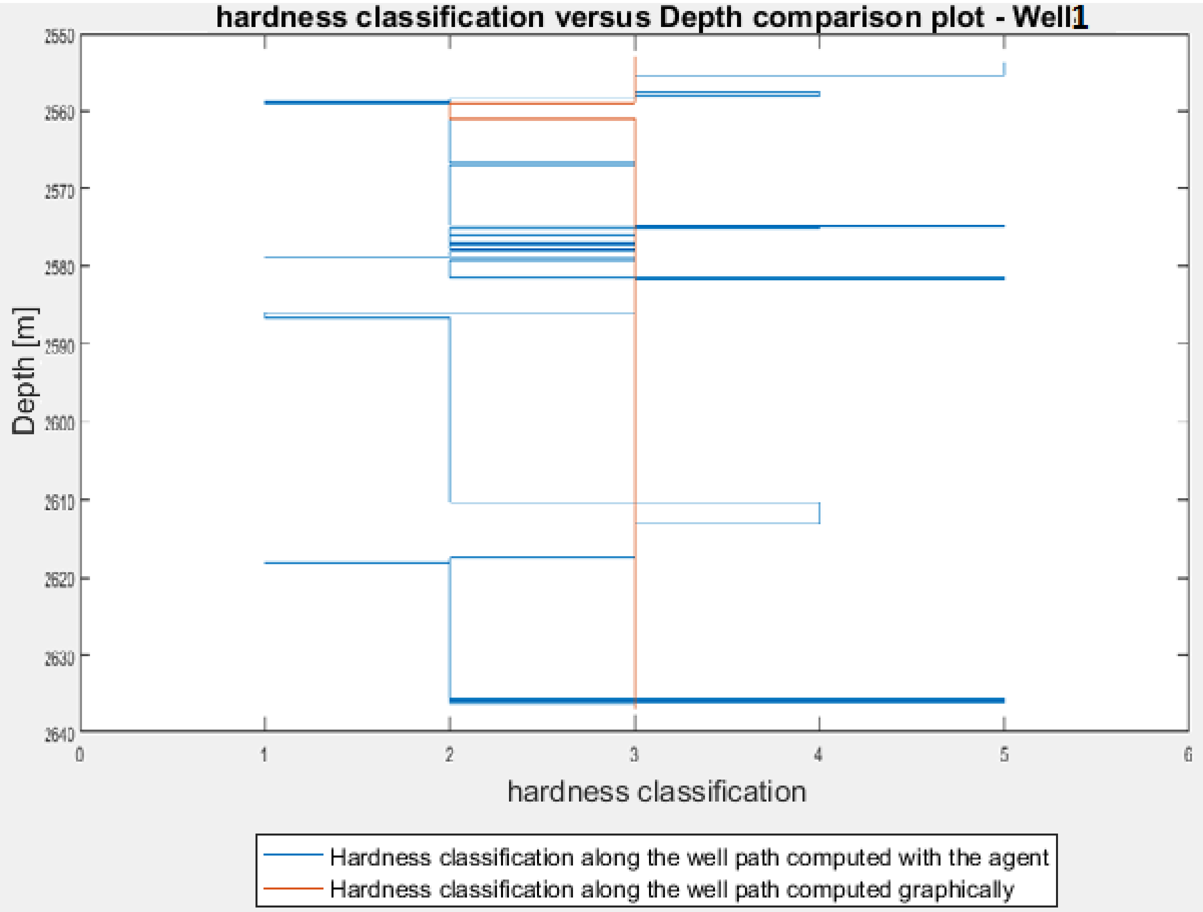


Figure 6-4 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 1

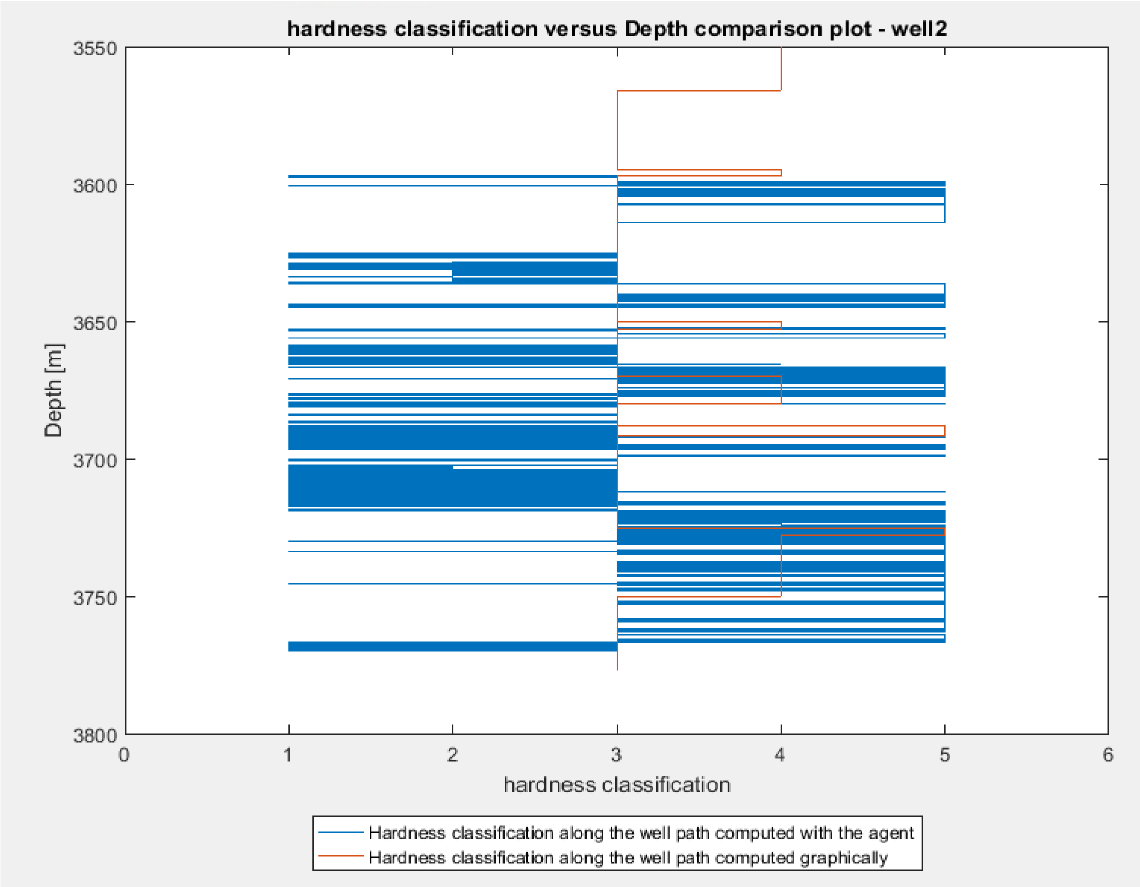


Figure 6-5 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 2

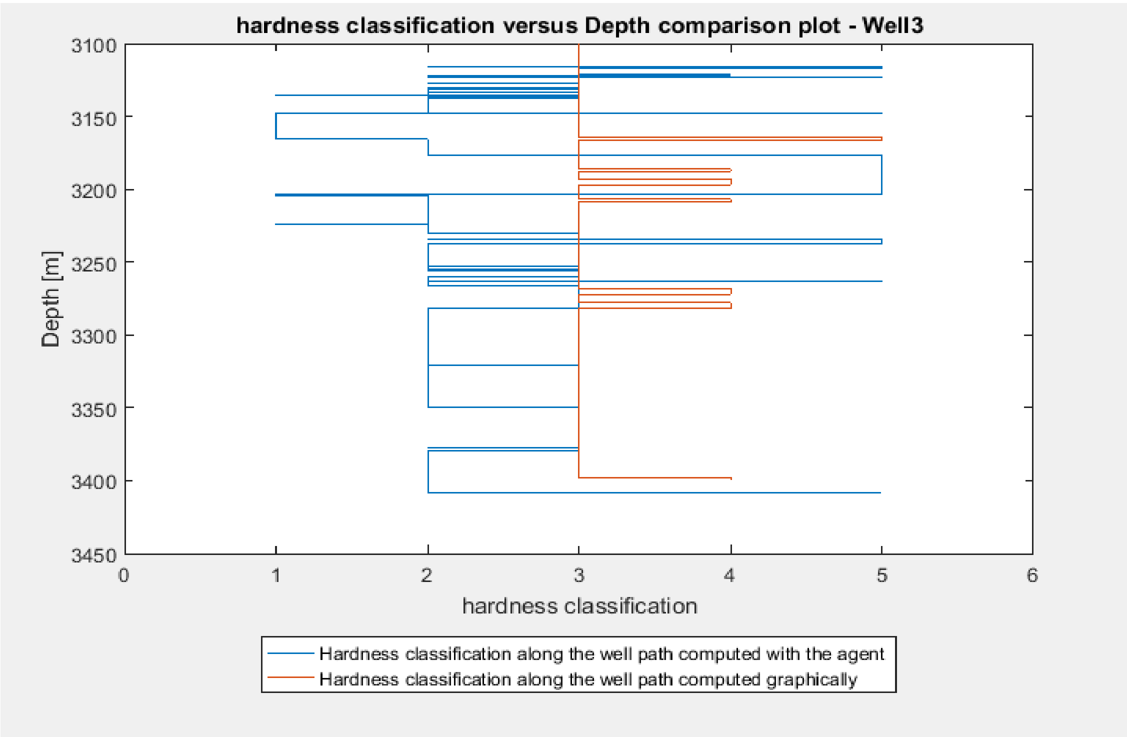


Figure 6-6 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 3

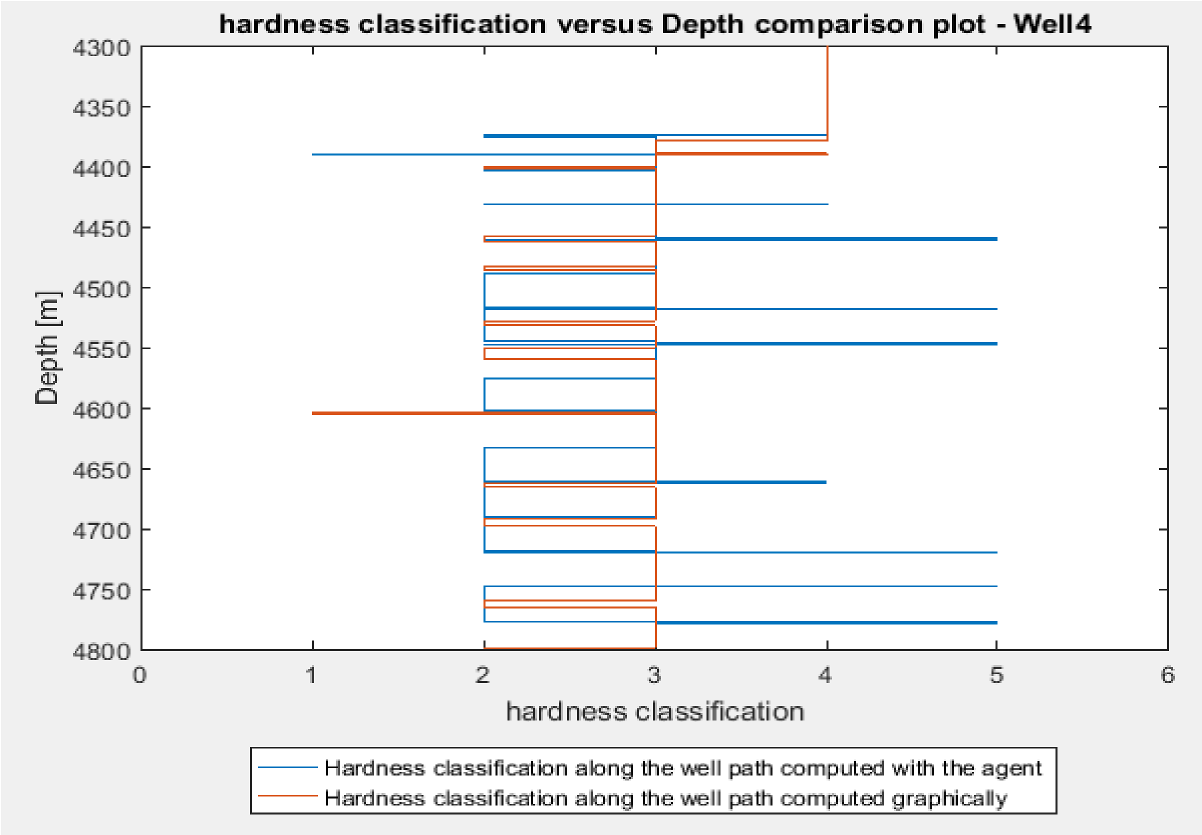


Figure 6-7 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 4

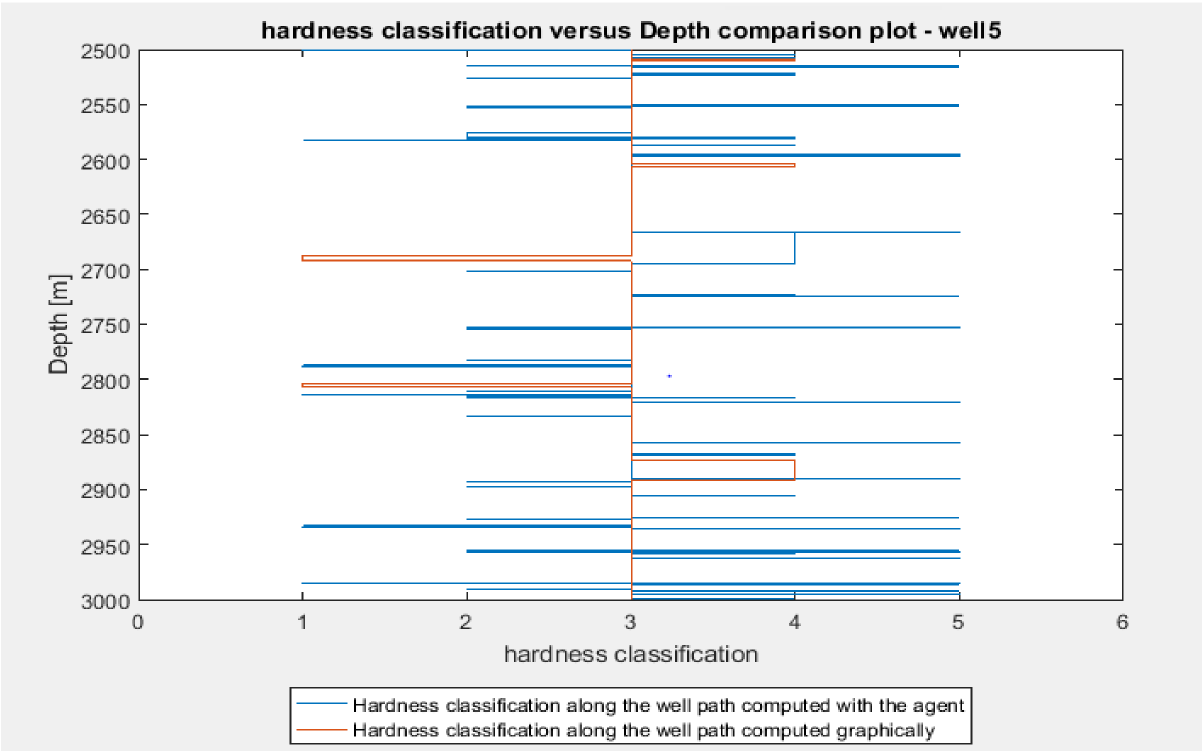


Figure 6-8 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 5

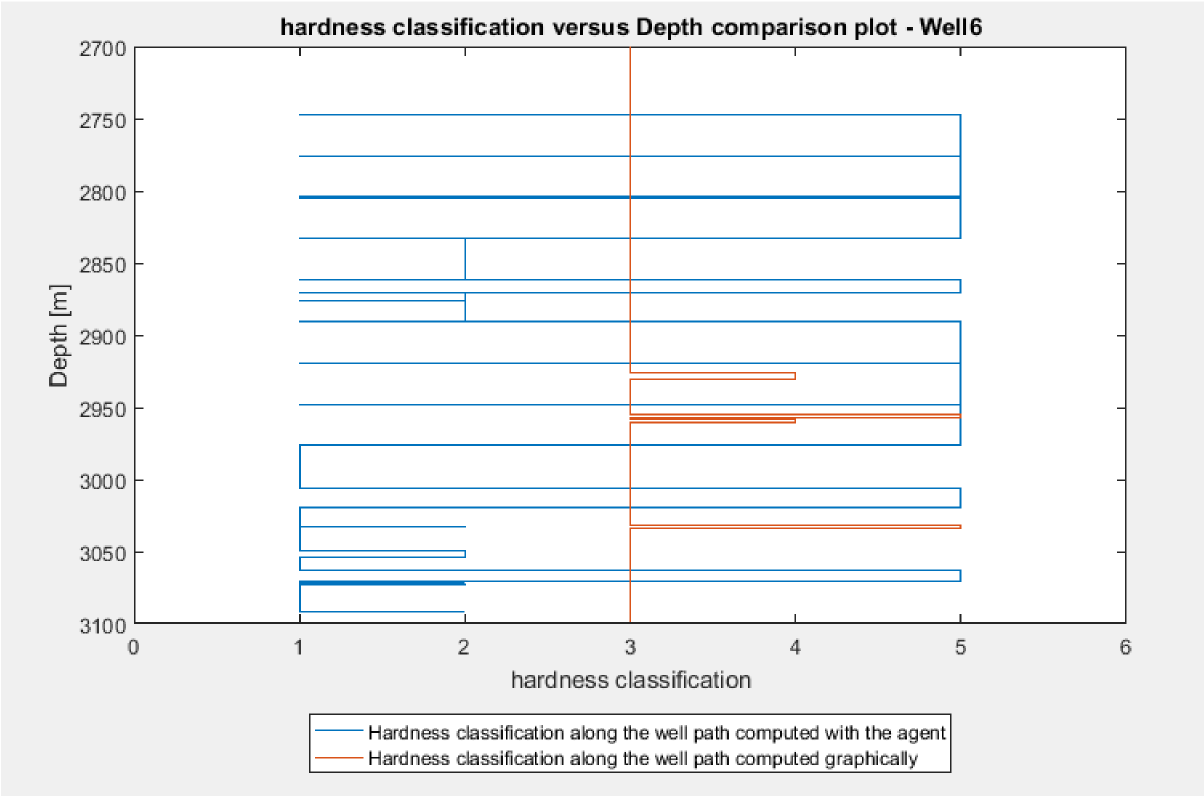


Figure 6-9 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 6

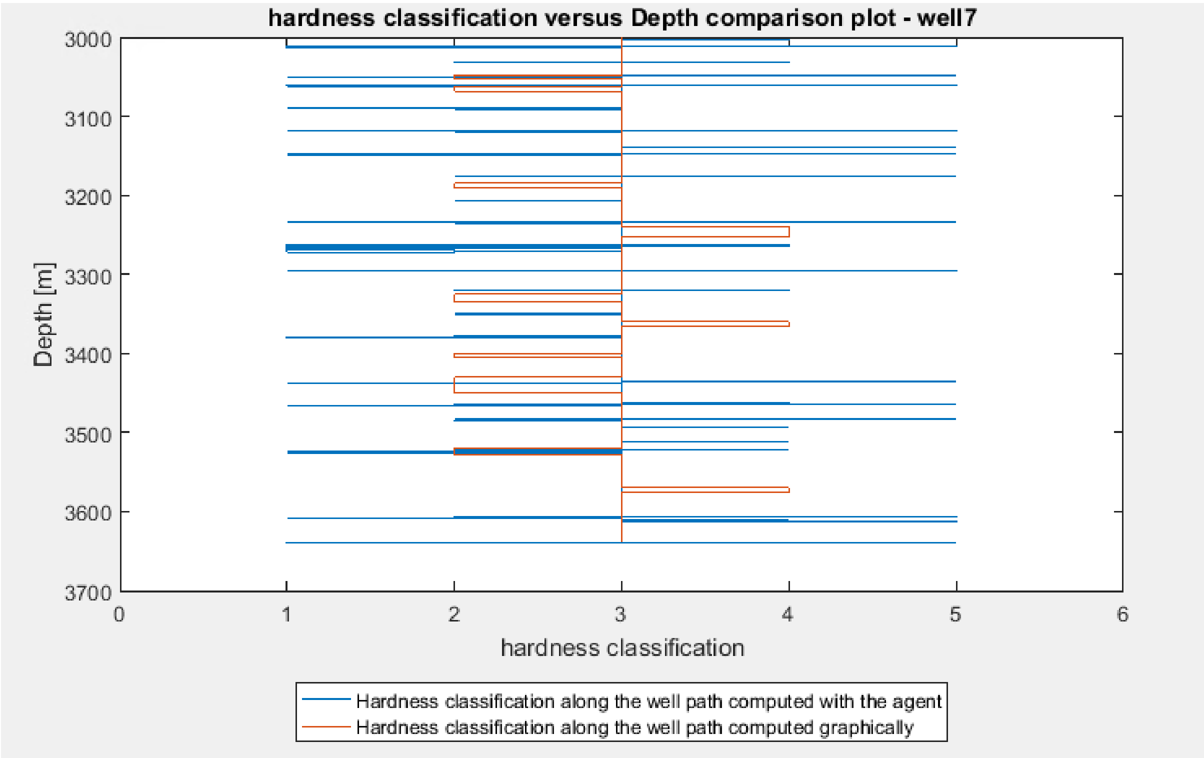


Figure 6-10 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 7

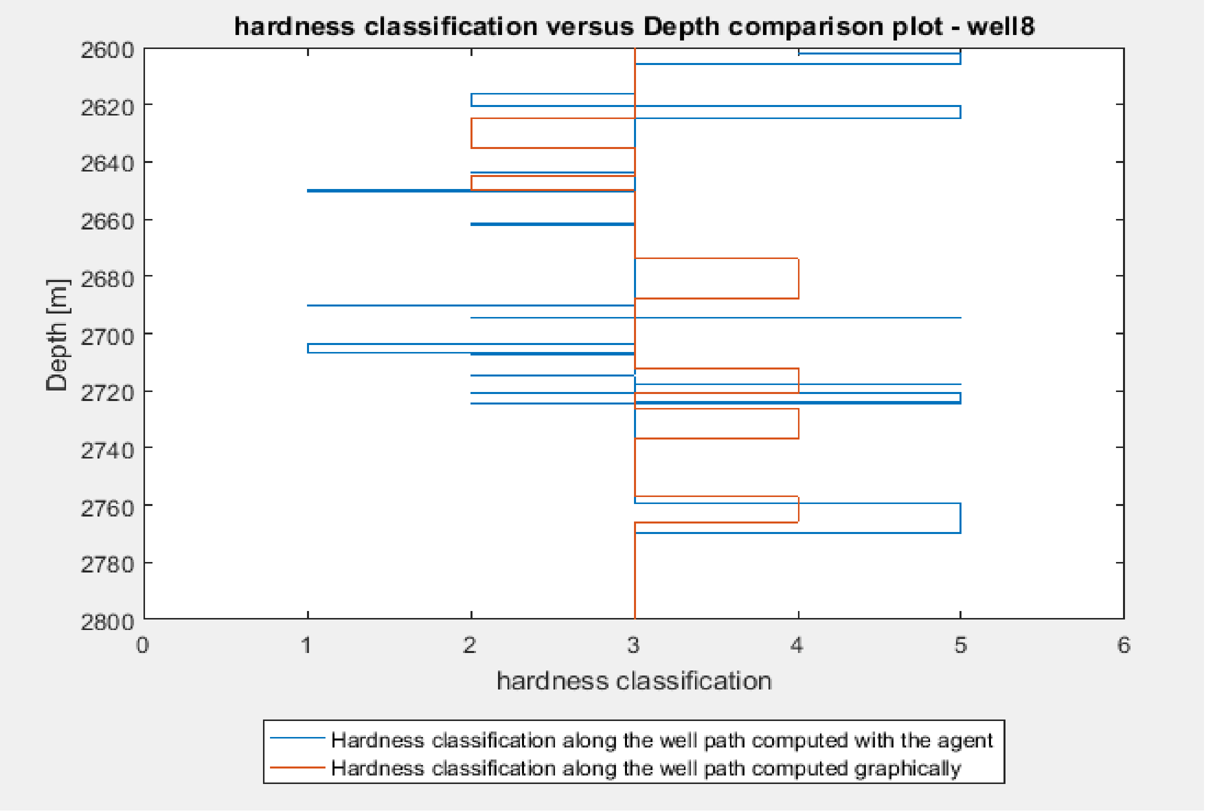


Figure 6-11 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 8

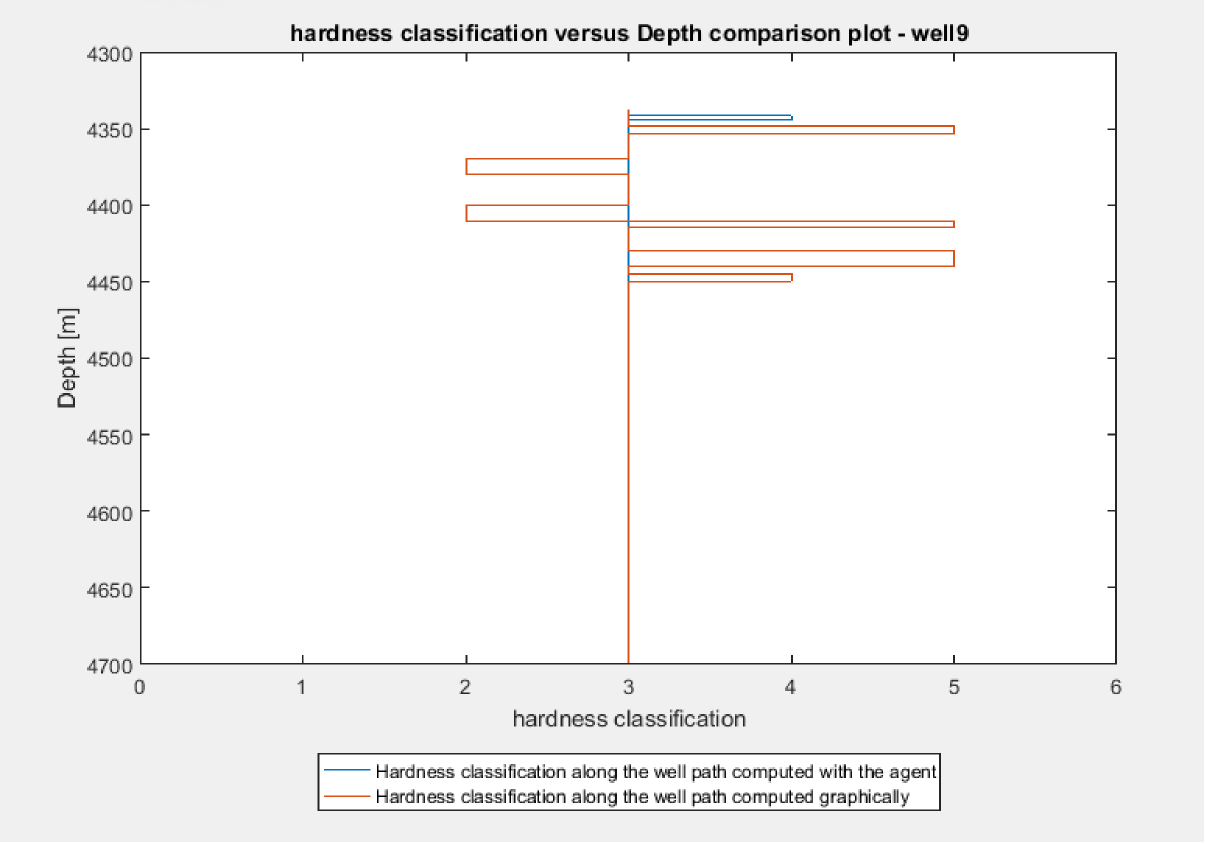


Figure 6-12 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 9

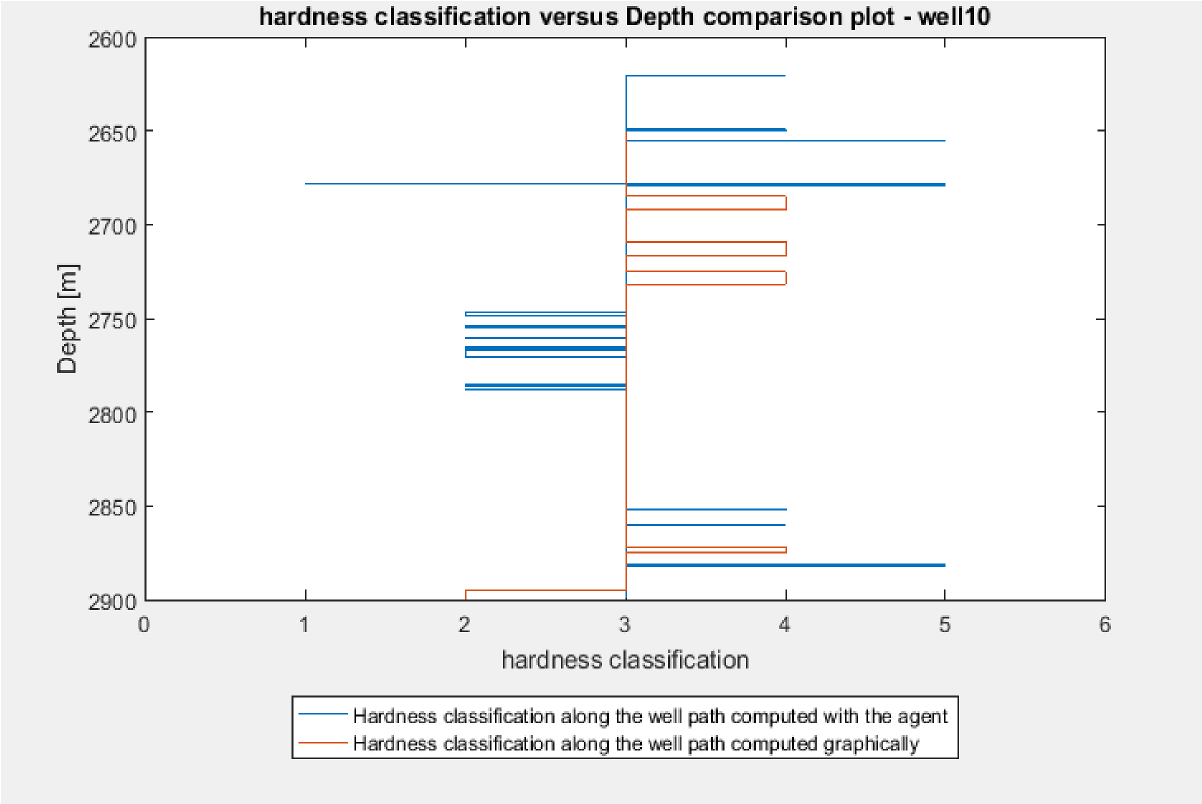


Figure 6-13 Comparison plot between hardness classification provided by the agent in blue and the one provide by the study of geological reports for the well 10

Plots are just provided, comment these plots is link with agent quality and thus it is made in chapter 7.

7 Quality assessment of the agent

We assess functionality and efficiency of the data agent according to criterions defining the quality of a quality, described in chapter 2.5.

7.1 Efficiency of the agent

The efficiency of an agent is defined by its time behavior and its resource utilization (Botella, Burgués et Al., 2004). To be a high quality software, the agent should be optimized and fast, using as few resources as possible. The best way to know what is the time behavior of our program is to use the run and time functionality of Matlab. Figure 7-1 presents the time behavior for well 3 provided by Matlab.

Start Profiling Run this code: Profile time: 31040 s

Profile Summary

Generated 06-Jun-2017 21:03:23 using performance time.

Function Name	Calls	Total Time	Self Time*	Total Time Plot (dark band = self time)
MainAgent	1	30752.970 s	8.653 s	
ARMATest	677	29493.429 s	5.071 s	
arima.estimate	17602	28707.120 s	111.934 s	
fmincon	35204	26033.965 s	18.416 s	
sqpLineSearch	17602	25920.006 s	158.775 s	
...pecifiedE0,userSpecifiedV0,trapValue)	6980295	18873.523 s	1170.052 s	
arima.estimate>nLogLike	6980295	15927.910 s	4674.338 s	
finitedifferences	573643	15640.297 s	250.421 s	
finDiffEvalAndChkErr	5071216	15347.588 s	277.406 s	
computeFinDiffGradAndJac	556041	14897.105 s	35.582 s	
LagOp.LagOp>LagOp.subsref	85103716	11113.734 s	8283.483 s	
optim\private\evalObjAndConstr	1873875	5864.407 s	182.955 s	
optim\private\computeSearchDirSQP	587040	4728.141 s	93.750 s	

Figure 7-1 Time behavior of the agent on the well 3 using the run and time functionality of Matlab. It provides time spend by the agent to treat all RTDD from well 3 in "Profile time". This Matlab's functionality also precise how many times a sub program is called during the total run of the main program in "calls". Finally, this functionality provides the running time of a sub program for a single run in "Self Time"

We see first, that the total profile time is huge, for well 3, it is 31040 seconds equivalent to 8 hours and 37 minutes. Then, we notice that most of the time is spend in the "ARMATest" sub program. It means that for every new data entered in the system (corresponding to a new iterative loop for the program), for well 3, the agent take an average of 5.071 seconds to determine p and q coefficients in order to create the ARMA model for the forecasting functionality. Table 7-1 summarizes all total profile time for the ten wells and the self-time of the "ARMATest" sub function.

TABLE 7-1 TOTAL PROFILE TIME AND "ARMATEST" SUB FUNCTION SELF-TIME FOR THE 10 WELLS

Well	1	2	3	4	5	6	7	8	9	10
Total profile time	19483 s 5 h 25 mn	134596 s 37 h 23 mn	31040 s 8 h 37 mn	17026 s 4 h 44 mn	75253 s 20 h 54 mn	13931 s 3 h 52 mn	83651 s 23 h 14 mn	18321 s 5 h 05 mn	877 s 15 mn	16082 s 4 hr 28 mn
ARMA Test self time	4.25 s	5.67 s	5.07 s	4.91 s	4.86 s	5.78 s	4.68 s	4.169 s	3.64 s	4.935 s

According to this time behavior study, the agent maintains an average of five seconds to run its "ARMATest" sub function. The agent needs a bit more than 5 seconds to process each data points. However, in a normal drilling operation, a new RTDD is implemented every 5 seconds, which is faster than the agent running time for one data point. It means that the program will provide a delayed information and struggle to operate in real conditions, as it does not process a data point fast enough. Thus we conclude that the agent have a low efficiency according ISO/IEC 9126 criterions (Botella, Burgués et Al., 2004). Chapter 8.4 details improvements which can help to increase this efficiency.

7.2 Functionality of the agent

The functionality of an agent can be defined by its time accuracy (Botella, Burgués et Al., 2004). To be characterized as a high quality software, the agent should provide precise and relevant results.

7.2.1 Factors needed to assess the functionality of the agent

In our case, we assess the accuracy of the table providing hardness classification encountered and corresponding depth intervals. The best way to know if agent outputs are relevant is to confront them to real cases found in chapter 5. This is done looking at comparison plots in chapter 6. For each comparison plots, we classify the results into four information categories provided by the agent:

- Positive true (PT): The agent provide the same hardness as the geological reports
- Negative true (NT): The agent classify hardness as average or blank as the same time as geological reports
- Positive false (PF): The agent classify the formation with a different hardness than the one stated by geological reports
- Negative false (NF): The agent classify formation with an average hardness or blank while geological reports classify it as hard or soft.

The objective is to create a confusion matrix table to describe the performance of the agent. Indeed, such matrix helps to compute accuracy, specificity, true positive rate, false positive rate, misclassification rate and precision of the agent (Santra and Christy, 2012).

7.2.2 Hits determination

We determine PT, NT, PF and NT thanks to comparison plots presented in chapter 6. We consider area where formation hardness is constant as one case and assess it. Figure 7-2 presents hits determination for the well 4 for a better understanding of the methodology. Then, table 7-2 provides results for all ten wells. PT, NT, PF and NF in table 7-2 come from Appendix XII, which presents hits determination figures for the other wells. In figures of the Appendix XII, a green arrow represents a PT, a blue arrow represents a NT, a red arrow represents a PF and a brown arrow is a NF.

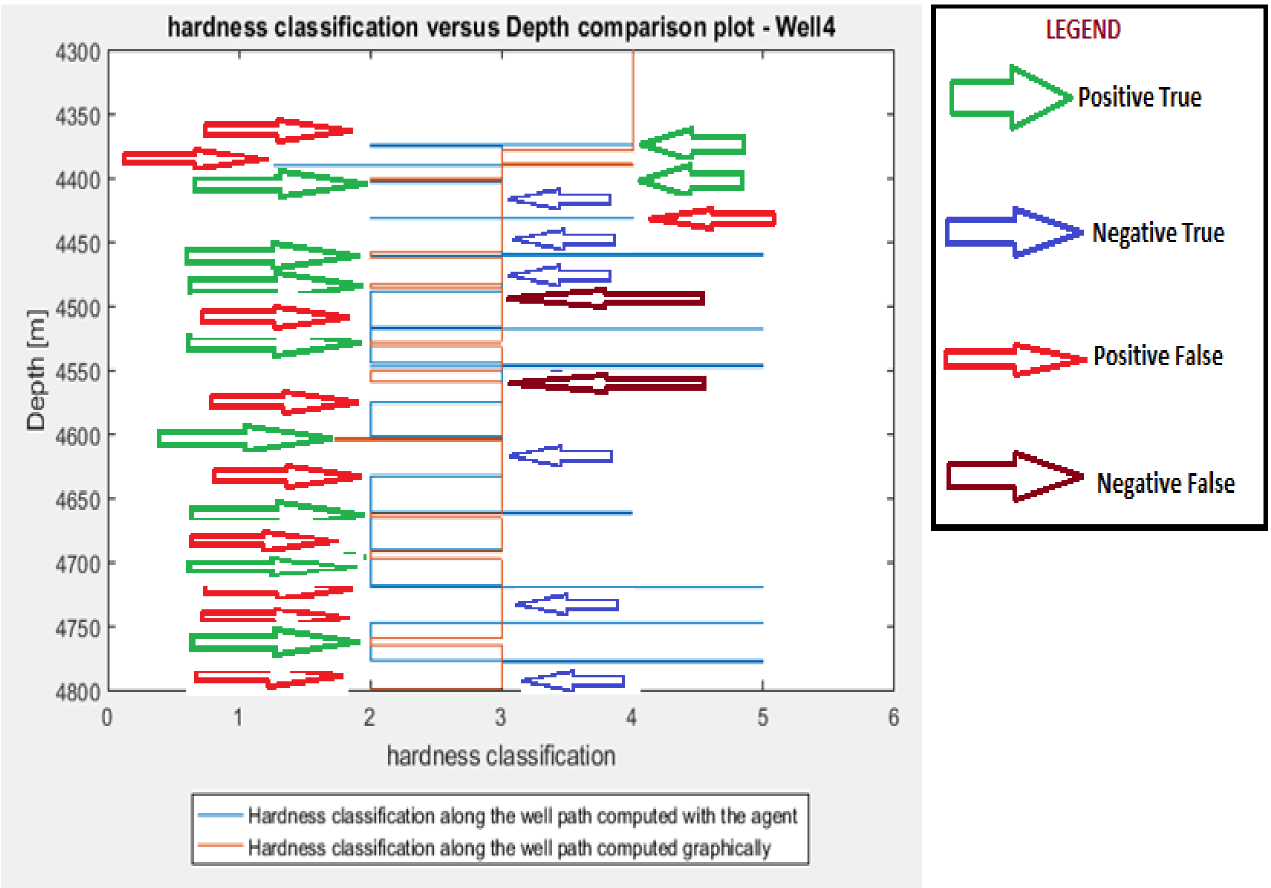


Figure 7-2 Hits determination for well 4, a green arrow is for positive true, a blue arrow is for negative true, a red arrow is for positive false and a brown arrow is for negative false. We count 10 PT represented by 10 green arrows, 6 NT represented by 6 blue arrows, 10 PF represented by 10 red arrows and 2 NF represented by 2 brown arrows

TABLE 7-2 HITS DETERMINATION FOR THE 10 WELLS

Well	1	2	3	4	5	6	7	8	9	10	Total
PT	1	5	3	10	3	2	7	3	0	0	34
NT	2	4	3	6	18	3	13	7	6	10	72
PF	11	14	9	10	22	10	13	5	1	7	102
NF	0	1	4	2	3	1	3	3	6	5	28
Total	14	24	19	28	46	16	36	18	13	22	236

We note that this graphical hits determination can be imprecise or subjective depending on how we define cases for PT, NT, PF or NF determination. However, we discuss this aspect in chapter 8.

7.2.3 Functionality assessment

Now that hits determination is available, we can build the confusion matrix following the work of Santra and Christy (2012). Table 7-3 is the confusion matrix of the agent.

TABLE 7-3 CONFUSION MATRIX OF THE AGENT

n=236	Predicted: NO	Predicted: YES	Total:
Actual: NO	NT=72	PF=102	174
Actual: YES	NF=28	PT=34	62
Total:	100	136	236

Thanks to this matrix and equations present in Santra and Christy (2012) work and Dataschool (2017), we are able to assess totally the functionality of the agent in the table 7-4 using six criterions. This table is made mixing results from all wells.

TABLE 7-4 FUNCTIONALITY ASSESSMENT OF THE AGENT THROUGH SIX CRITERIONS

Criterion	Formula	Value for the agent	What does it represent?
Accuracy	$\frac{\text{NT} + \text{PT}}{\text{n}}$	$\frac{72 + 34}{236} = 45.8\ \%$	How often the agent is right
Specificity	$\frac{\text{NT}}{\text{Actual NO}}$	$\frac{72}{174} = 41.4\ \%$	How often the agent find average when it's really average hardness
Sensitivity	$\frac{\text{PT}}{\text{Actual YES}}$	$\frac{34}{62} = 54.8\ \%$	How often the agent find hard or soft when it's really hard or soft
False positive rate	$\frac{\text{PF}}{\text{Actual NO}}$	$\frac{102}{174} = 58.6\ \%$	How often the agent find hard or soft when it's really average hardness
Error Rate	$1 - \text{Accuracy}$	$1 - 0.458 = 54.2\ \%$	How often the agent is wrong
Precision	$\frac{\text{PT}}{\text{Predicted YES}}$	$\frac{34}{136} = 25\ \%$	How often is it hard or soft when the agent says so

According to elements of table 7-4, the agent is not so accurate, precise or sensible. Chapter 8.4 details factors that could have led to such a low accuracy and provides improvements that can help to increase the functionality of the agent.

8 Self-Assessment

8.1 Quality of Data

Data from Diskos database are of great quality. File are large and often furnished with several reports (geological or drilling reports). Moreover, geological reports contain a composite log or a sidewall core description, which was helpful to establish test cases for chapter 7. However, some data were a bit scattered, for instance well 9 which made the analysis more challenging. Moreover, well 1 did not had a proper composite log, making the implement of the test case harder. All geological reports are relatively old and as such, some imprecisions were noticed in their log and sidewall core description affecting the consistency of the test cases.

8.2 Quality of the mathematical model

The mathematical model behind the agent is composed of three main parts: The hardness computation, the sorting of hardness and the forecasting model. We assess them separately.

8.2.1 Mathematical model chose for the hardness computation

Present part is an extended version of Donne (2016) as it was also the model chosen to compute hardness during previous project. Chapter 7 presented a quite low functionality according to the hardness classification and the chosen model could be one factor explaining this lack of accuracy. Indeed, this model could be improved to fit reality better, providing more accurate geology information using fewer assumptions. Some assumptions are acceptable as we removed factors changing slowly with depth form the Bourgoyne and Young model, while some assumptions need some changes. An idea could be to take into account the dc-exponent as it can vary quickly and try to vary more often a_5 and a_6 exponents as they are supposed to change constantly. The problem with these two exponents is that they change constantly and they depend of the formation drilled itself, thus they are impossible to update during a run. That is why we need to take an average representative exponent. In addition, if we want to make the agent work on long intervals, the model should include assumptions linked to bit wear and changes in mud density. These properties were not handled in this master thesis and could be another reason for bad agent outputs in long drilling intervals (Donne, 2016).

8.2.2 Model used to sort hardness

In this thesis, we used the median of the first 50 m drilled to sort our hardness as we were dealing with small drilling intervals. However we noticed, in table 3-7 of chapter 3.5.1, that the median computed using 50 m drilled was off by 50 % from reality for 30 % of the wells in our test panel. Thus, the 50 m interval is maybe too small to compute the median hardness and we better use instead 50 m at beginning and then update at 100 m. However, we have to be careful to not increase too much this interval, as our mathematical model is more precise on small window interval. Finding the good compromise could be the next objective of such a project.

8.2.3 Forecasting model

We chose to use an ARIMA forecasting model for this thesis, as it was more precise in its forecast compared to the Holt method. However, ARIMA is more complex to implement. We tried to simplify it by using a zero integration degree and varying only p and q from 1 to 5. However, the method is still quite long to run, as stated in chapter 7.1 and it may lead to difficulties to use it in real drilling operations. Moreover, according to videos in the .zip folder, this forecasting method will be relevant only for wells with lots of high quality data as for well 2, well 5 and well 7. For the others, the forecast is pretty far from reality and always overestimate the upcoming hardness. It is better than underestimate it from a safety point of view, but not from the economical one. Indeed, the driller, thinking that the coming formation is harder could reduce his WOB, RPM and ROP more than he should and thus reduce his drilling efficiency. To increase the quality of the forecast, we could, when selecting an ARIMA model, try to go higher than 5 for the p and q coefficients. The issue is that it is going to increase the complexity of the model as much as its computational time. Here again, we have to make a compromise between a high quality forecast and a short time behavior. Finally, we presented in chapter 3.6.3 that the time forecast window we choose has an effect on the forecast precision. Indeed, ARIMA model is made for short forecasts. The three minutes window we chose is a small window from a drilling operation point of view but it is still already a low quality forecast for most of 70 % of our tested wells. This is another challenge we have to tackle to create a better agent: Finding the compromise between accuracy and long term forecast.

8.3 Quality of the test methodology

All computation regarding the functionality of the agent is based on hits determination. However, this hits determination of PT, NT, PF and NF is made graphically and thus is imprecise. Moreover, definition of cases is quite subjective as this methodology is made normally for programs that just have to answer yes and no. Here we had to consider a case as an area where the hardness stays constant. It would have been better to consider for example one data point as one case and check if the classification provided by the agent is good or not for this data point. However, regarding to the important number of data points (several hundreds) per well for each wells, it was too time consuming to be done in the semester. Finally, test cases themselves has to be assessed. They all come from geological reports, composite logs and sidewall core descriptions. We have to remember that they are not perfect and thus pieces of information could miss, which directly affect the accuracy provided by chapter 7.

8.4 Future improvements

This master thesis is an initial work on creating a data agent using Matlab forecasting changes in formation hardness while drilling. Objectively, in more than 70 % of our test cases, this agent is inaccurate and slow. Nevertheless, it still provide hardness reports and forecasts of hard stingers while drilling which can be relevant depending on the quality of the RTDD implemented. That is why we can think about some future improvements to increase the relevance of this master thesis and its agent:

- Change the computation of the median hardness by re-calculating after 100 m drilled instead of only at 50 m drilled to improve the quality of the hardness sorting as stated in the chapter 8.2.2.
- Reduce the assumptions used in the mathematical agent model. The first step could be to take into account changes in formation pressure using dc-exponent and use a new iterative loop to compute relevant drillability values. It is challenging as parameters involved in dc-exponent computations are similar to the ones used for hardness computation. However, it can help improving the accuracy in large drilling intervals.
- Following the idea raised by Solberg (2012) make the agent use the variation of the block position through time instead of using the given ROP. In her work, she presents the fact that this ROP could be quite imprecise when measured by services companies and so in order to increase the accuracy of the agent we could try to compute ROP ourselves.

- Present agent only works for a specific format type presented in chapter 6. It is time consuming for the user and far from reality as RTDD are monitored following different formats depending on companies. Thus, the next step could be the improvement of the Matlab code to allow all types of RTDD file to be directly uploaded without specific pre-work on it.
- In order to improve the time efficiency of the agent an improvement has to be done in the "ARMATest" sub program. An idea is to reduce number of model creations by ranging p and q only to 4 or to 3. But, this will reduce the agent forecast accuracy, which is not already high enough. The other idea is to use more resources. Test were made on the same old personal computer to ensure same external factors to all wells during the time behavior test. Improving hardware's quality of the machine, for instance its processor power, could ensure a significant reduction of the time behavior of the program. Moreover, the implementation of the agent itself is naive as made by a petroleum engineering student with limited skills in computer science. Thus, work on the optimization of the algorithm itself would also lead to a better time behavior.
- As stated in the introduction, the motivation for this agent is to prevent incidents and failures linked to change in formation hardness. To reach this goal, this master thesis focused on the creation of a data agent. But another approach could be complementary: the use of the ontology as a tool to detect failures induced by formation hardness. Ontology is defined by Gruber (1993) as an explicit specification of a conceptualization. And using Skalle, Aamodt and Gundersen's work (2013) we are able to link it to the thesis as they developed ontology as a method that helps reveal the most probable cause of a drilling-process failure immediately after occurrence. Thus, the implementation of Skalle et al. (2013) ontology model in the future represents a good strategy to test the agent and assess its efficiency.
- At this point, an output of the agent is the classification of the hardness encountered while drilling. Some improvements can also be done there. First, the classification should only contain either if the formation is hard or soft but not if it is average. This average class was needed for coding plots but an improvement of the main code of the agent could lead to deleting this class in the output report, as it is not necessary. Moreover, the classification should also include laminated formation. It is challenging to detect them just looking at hardness plots but seeing quick alternation between soft

and hard stringers could be the sign of a laminated formation and thus a condition in the hardness classification report of the agent should be added.

- Regarding the forecasting function, it provides a plot of hardness versus time and thus give a time information to the driller, as the mathematical model used is time series. However, the agent should instead be able to forecast information linked with depth. As ROP links time and depth, an improvement could be to forecast the depth of the incoming hard or soft formation using present ROP with a re-computation of the forecast in case of a sudden modification of the rate of penetration.
- Finally, in all RTDD files, extreme hardness values were observed as stated in chapter 3. They are not realistic, it has to do with reporting and especially time between observations. Therefore, only average values over at least 3 time steps can be regarded as trustfully and it should be in the next version of the agent.

9 Conclusion

Forecasting and detecting changes in formation hardness while drilling can prevent important and expensive failures of happening. It has a direct relation to decreasing failures and thus NPT and to increasing drilling efficiency. To tackle this challenge, a Matlab agent has been created. This program compute hardness from drillability using information found in real time drilling data, then it forecast incoming hardness and provide a classification of hardness encountered while drilling with corresponding depth intervals.

On basis of the work done in this master thesis, the following statement can be made:

- The mathematical model behind this program is a simplified equation of the ROP model proposed by Bourgoyne and Young (1986). Assumptions made on this model allow the agent to work on small drilling intervals but still present uncertainties due to the simplification.
- The forecasting method is based on ARMA (p, q) model. Assumptions made on p and q selection does not allow the agent to have fast enough accurate forecast on most of the test cases. However, for some high quality and large RTDD it appears to work with significate fidelity to reality.
- Bourgoyne and Young exponents change constantly versus the formation drilled. In order to create a useful model, we simplified this issue sorting the hardness using the median hardness RTDD corresponding to the first 50 m drilled.
- The agent was tested on ten existing wells against a manually detected hardness type, supported by geological reports, composite log and sidewall core description coming from Diskos, a large governmental database.
- Assessment of the agent's quality is made using a confusion matrix and a time behavior study.

10 References

Aadnoy, Bernt S., and Bernt Sigve Aadn²y. Modern Well Design : Second Edition. Modern Well Design: Second Edition. 2nd ed. ed. Hoboken: CRC Press, 2010.

Bahari, Mohamad Hasan, Aboozar Bahari, Farzad Nejatimoharrami, Ramin Rajaei, and Bijan Vosoughi Vahdat. "Forecasting the Rate of Penetration Using Bourgoyne and Young Model Associated with Genetic Algorithm." Journal of Applied Sciences (2008): 5.

Botella, P, X Burgués, JP Carvallo, X Franch, G Grau, J Marco, and C Quer. "Iso/Iec 9126 in Practice: What Do We Need to Know." Paper presented at the Software Measurement European Forum, 2004.

Bourgoyne Jr, AT, and FS Young Jr. "A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection." Society of Petroleum Engineers Journal 14, no. 04 (1974): 371-84.

Bourgoyne, Adam T. Applied Drilling Engineering. Spe Textbook Series. Vol. Vol. 2, Richardson, Tex: Society of Petroleum Engineers, 1986.

Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015.

Brown, ET, SJ Green, and KP Sinha. "The Influence of Rock Anisotropy on Hole Deviation in Rotary Drilling—a Review." Paper presented at the International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 1981.

Corina, Anisa Noor. "User Guide : Rtdd Log Viewer." 37: Norwegian University of Science and Technology, 2016.

Dataschool. Confusion matrix, available at http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/, Cited May 2017

Deartment, Engineering. "Geological Report Development Well 33/12-B13 Statfjord Oil Production Well." Mobil Exploration Norway Inc., 1985.

Department, Engineering. "Geological Report Development Well 33/12-B20 Lower Brent Oil Production Well." Mobil Exploration Norway Inc, 1986.

Department, Engineering. "Geological Report Development Well 33/12-B30 Upper Brent Water Injector." Mobil Exploration Norway Inc., 1984.

Department, Engineering. "Geological Report Development Well 33/12-B31 Lower Brent Oil Production Well." Mobil Exploration Norway Inc., 1985.

Department, Engineering. "Geological Report Development Well 33/9-A18." Mobil Exploration Norway INC., 1982.

Department, Engineering. "Geological Report Development Well 33/9-A34." Mobil Exploration Norway Inc, 1983.

Diskos. Norway's national data repository for petroleum data, Available at http://www.diskos.no, Cited May 2017

Donne, Clement. "Parameter Detection in Real Time Drilling Data, Create a Matlab Agent to Distinct Changes in Formation Hardness." NTNU: Norwegian University of Science and Technology, 2016.

Eckel, John R. "Microbit Studies of the Effect of Fluid Properties and Hydraulics on Drilling Rate." Journal of Petroleum Technology 19, no. 04 (1967): 541-46.

Engineering, Development. "Geological Report Development Well 33/9-A38." Mobil Exploration Norway Inc., 1980.

Factpages [1]. Statfjord field, Available at http://factpages.npd.no/ReportServer?/FactPages/PageView/field&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=43658&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [2]. Exploration Well 33/12-1, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/wellbore_exploration&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=417&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [3]. Cromer Knoll Group, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/strat_Litho_level1_group_formation&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=23&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [4]. Viking Group, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/strat_Litho_level1_group_formation&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=188&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [5]. Brent Group, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/strat_Litho_level1_group_formation&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=16&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [6]. Dunlin Group, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/strat_Litho_level1_group_formation&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=29&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [7]. Statfjord Group, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/strat_Litho_level1_group_formation&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=157&IpAddress=129.241.94.119&CultureCode=en, Cited May 2017

Factpages [8]. Lunde formation, available at http://factpages.npd.no/ReportServer?/FactPages/PageView/strat_Litho_level1_group_formation&rs:Command=Render&rc:Toolbar=false&rc:Parameters=f&NpdId=97&IpAddress=129.241.94.119&CultureCode=nb-no, Cited May 2017

Fileinfo [1]. ASC file, available at https://fileinfo.com/extension/asc, Cited May 2017

Fileinfo [2]. TIF file, available at https://fileinfo.com/extension/TIF, Cited May 2017

Fileinfo [3]. TXT file, available at https://fileinfo.com/extension/TXT, Cited May 2017

Fileinfo [4]. LAS file, available at https://fileinfo.com/extension/las, Cited May 2017

Fileinfo [5]. LIS file, available at https://fileinfo.com/extension/lis, Cited May 2017

Fileinfo [6]. CSV file, available at https://fileinfo.com/extension/csv, Cited May 2017

Gardner, Everette S. "Exponential Smoothing: The State of the Art—Part Ii." International Journal of Forecasting 22, no. 4 (2006): 637-66.

Gruber, Thomas R. "A Translation Approach to Portable Ontology Specifications." Knowledge acquisition 5, no. 2 (1993): 199-220.

H., Idrissi. "Prévisions Par L'utilisation De Série Chronologique Et La Régression Linéaire Simple." FACULTE DES SCIENCES ET TECHNIQUES FES, 2014.

Heidelberg. Dictionary of Gems and Gemology. Berlin, Heidelberg: Springer Berlin Heidelberg: Berlin, Heidelberg, 2005. doi:10.1007/b138842.

Helgestad, Dag-Erik. "The Drilling Process: A Plantwide Control Approach." June, 2010.

Holt, Charles C. "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages." International Journal of Forecasting 20, no. 1 (2004): 5-10.

Hoseinie, S. H., M. Ataei, and R. Mikaiel. "Comparison of Some Rock Hardness Scales Applied in Drillability Studies." Arabian Journal for Science and Engineering 37, no. 5 (2012): 1451-58.

J.O.Fløtre, E.C.Quale. "Geological Report Development Well 33/9-A23a." Statoil Protek Operations Stavanger, 1993.

J.O.Fløtre, R.E.Aakvik. "Statoil Driftsdivisjon Stavanger Geological Report Development Well 33/9-A35 Upper Brent Oil Production Well." Statoil driftsdivisjon stavanger, 1993.

Jorden, JR, and OJ Shirley. "Application of Drilling Performance Data to Overpressure Detection." Journal of Petroleum Technology 18, no. 11 (1966): 1,387-1,94.

Jovanović, Irena. "Software Testing Methods and Techniques." The IPSI BgD Transactions on Internet Research 30 (2006).

Karfakis, MG, and JF Evers. "Effects of Rock Lamination Anisotropy on Drilling Penetration and Deviation." Paper presented at the International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 1987.

Manutchehr-Danai, Mohsen. Dictionary of Gems and Gemology. Springer Science & Business Media, 2013.

Mélard, Guy. "Initiation À L'analyse Des Séries Temporelles Et À La Prévision." Revue MODULAD 82, no. 35 (2006).

Mélard, Guy. "Méthodes De Prévision À Court Terme." ULB--Universite Libre de Bruxelles, 2007.

Nascimento, Andreas, David Tamas Kutas, Asad Elmgerbi, Gerhard Thonhauser, and Mauro Hugo Mathias. "Mathematical Modeling Applied to Drilling Engineering: An Application of

Bourgoyne and Young Rop Model to a Presalt Case Study." Mathematical Problems in Engineering 2015 (2015).

NDT. Iowa state university and nondestructive testing resources center. Hardness, available at
https://www.nde-ed.org/EducationResources/CommunityCollege/Materials/Mechanical/Hardness.htm, Cited May 2017

Oilprice. Historic of crude Brent oil price, available at http://oilprice.com/commodity-price-charts?1&page=chart&sym=CB*1, Cited May 2017

Pål Skalle, Isak Swahn. "Best Practice for Construction of Data Agents." 7. Trondheim: NTNU, 2016.

Paulsen, åge. "Geological Report Statfjord Field 33/12-B16 33/12-B-16a 33/12-B16at2." Statoil, 1995.

Rahimzadeh, H, M Mostofi, and A Hashemi. "A New Method for Determining Bourgoyne and Young Penetration Rate Model Constants." Petroleum Science and Technology 29, no. 9 (2011): 886-97.

Rajabov, Vusal, Stefan Z Miska, Laurence Mortimer, Mengjiao Yu, and Mehmet Evren Ozbayoglu. "The Effects of Back Rake and Side Rake Angles on Mechanical Specific Energy of Single Pdc Cutters with Selected Rocks at Varying Depth of Cuts and Confining Pressures." Paper presented at the IADC/SPE Drilling Conference and Exhibition, 2012.

Rakhleev, G. "Method and Results of an Experiment to Measure the Mechanical Drilling Speed and Durability of the 6l-145-Ok Bit in Relation to Drilling Conditions and Rock Hardness." Soviet Mining 3, no. 2 (1967): 156-62.

Rashidi, Behrad, Geir Hareland, and Zebing Wu. "Performance, Simulation and Field Application Modeling of Rollercone Bits." Journal of Petroleum Science and Engineering 133 (2015): 507-17.

Rehm, Bill, and Ray McClendon. "Measurement of Formation Pressure from Drilling Data." Paper presented at the Fall Meeting of the Society of Petroleum Engineers of AIME, 1971.

Santra, AK, and C Josephine Christy. "Genetic Algorithm and Confusion Matrix for Document Clustering." International Journal of Computer Science, no. 2012 (2012): 322-8.

Schlumberger [1]. Hard rock, Available at http://www.glossary.oilfield.slb.com/Terms/h/hard_rock.aspx, Cited May 2017.

Schlumberger [2]. Dogleg, Available at http://www.glossary.oilfield.slb.com/Terms/d/dogleg.aspx Cited May 2017.

Schlumberger [3]. Washout, Available at http://www.glossary.oilfield.slb.com/Terms/w/washout.aspx, Cited May 2017.

Schlumberger [4]. Real Time Data, Available at http://www.glossary.oilfield.slb.com/Terms/r/real_time_data.aspx, Cited May 2017.

Skalle, Pål, Agnar Aamodt, and Odd Erik Gundersen. "Detection of Symptoms for Revealing Causes Leading to Drilling Failures." SPE Drilling & Completion 28, no. 02 (2013): 182-93.

Skalle, Pål. "Compendium Tpg 4242 - Pressure Control and Chemical Stability in Wellbores." 122. Trondheim, Norway: NTNU, 2015.

Solberg, Silje Marie. "Improved Drilling Process through the Determination of Hardness and Lithology Boundaries." (2012).

Standardization, International Organization for, and International Electrotechnical Commission. Software Engineering--Product Quality: Quality Model. Vol. 1: ISO/IEC, 2001.

Stunes, Sindre. "Methods of Pore Pressure Detection from Real-Time Drilling Data." (2012).

Wu, A, G Hareland, and B Rashidi. "The Effect of Different Rock Types and Roller Cone Insert Types and Wear on Rop (Rate of Penetration)." Paper presented at the 44th US Rock Mechanics Symposium and 5th US-Canada Rock Mechanics Symposium, 2010.

11 Nomenclature

$(\frac{W}{d_b})_t$	Threshold bit weight per inch of bit diameter at which the bit begins to drill
$\widehat{H}_T$	Hardness at horizon h
$\hat{a}_1$	Level for Holt method
$\hat{a}_2$	Slope for Holt method
F_j	Hydraulic impact force beneath the bit in [lbf]
H_t	Hardness time series
e_T	Forecast error
g_p	Pore pressure gradient in [lbm/gal],
θ_q	MA polynomial from order q
ρ_c	Equivalent circulating density,
ϕ_p	AR polynomial from order p
∇	Differentiation operator
B	delay operator
D	True vertical well depth in [ft]
d_e	d-exponent
d	Differentiation order of ARIMA model
d_b	Bit Diameter in [in]
d_c	dc exponent
h	Horizon of the forecast
h	Fractional tooth dullness,
H(t)	hardness value at time t
MW1	Normal mud weight gradient for the area
MW2	Equivalent circulating density or mud weight in use
N	RPM in [rev/min]
p	Number of integration in ARIMA model
q	Polynomial order of the moving average in ARIMA model

R	ROP in [ft/hr]
t	Time
W	WOB in [lbs]
W	Weight on bit [1000 lbf]
α	Exponential smoothing parameter
λ	Holt parameter for level
μ	Holt parameter for slope

12 Abbreviations

AR	AutoRegressive
ARIMA	AutoRegressive Integrated Moving Average
BP	British Petroleum
BPOS	Block Position
CLST	Claystone
COMP	Company
DBTM	Measured depth of the Drilling Bit
DMEA	Measured depth of the well
DRILLEX	Drilling Expenses
EOW	End of Well Report
ES	Exponential Smoothing
FM	Formation
GP	Group
HD	Hard
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
LST	Limestone
MA	Moving Average
MD	Measured Depth
Mod	moderate
MRLST	Marlstones
NF	Negative False
NPD	Norwegian Petroleum Directorate
NPT	Non Productive Time
NT	Negative True
NTNU	Norges Teknisk-Naturvitenskapelige Universitet

PF	Positive False
PT	Positive True
ROP	Rate Of Penetration
RPM	Rotation Per Minute
RTDD	Real Time Drilling Data
SH	Shale
SLTST	Silstone
SQR	Structured Query Reporting
SST	Sandstones
TD	True Depth
UK	United Kingdom
WOB	Weight On Bit

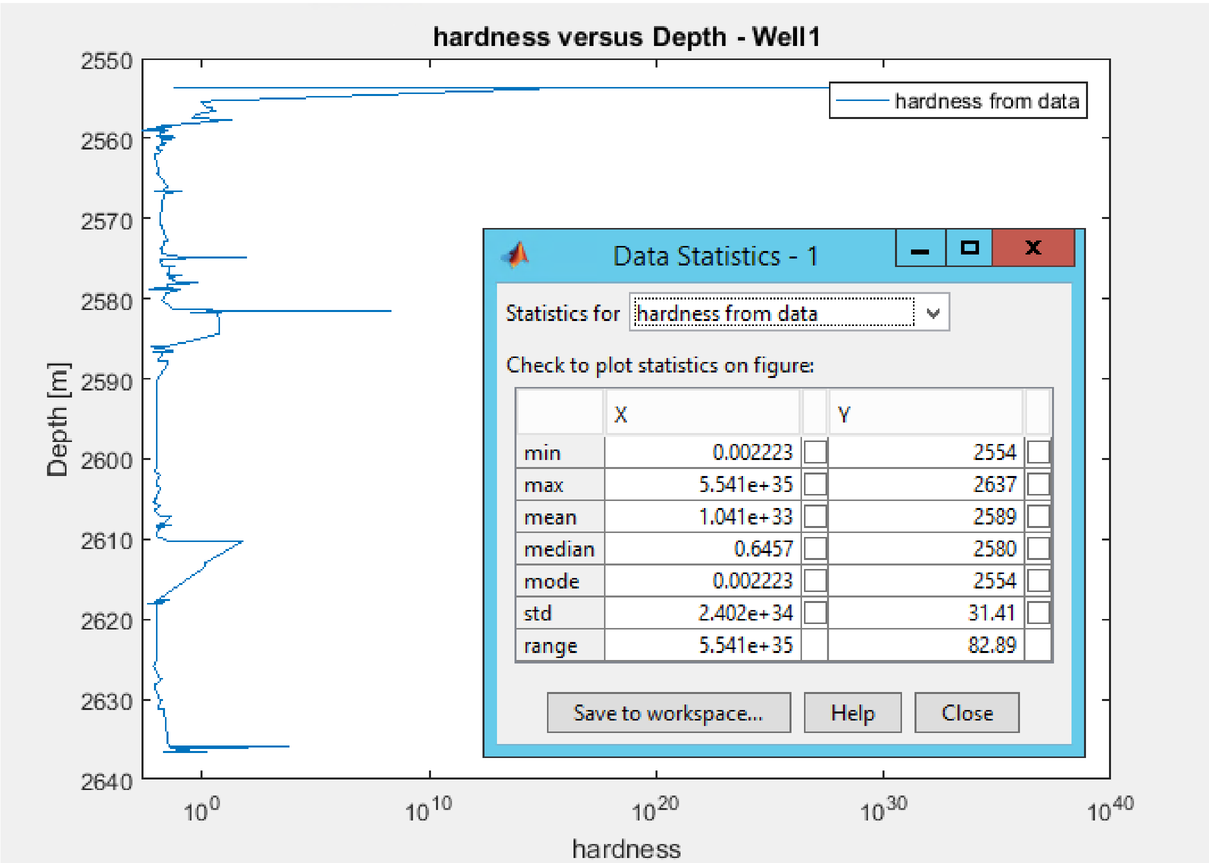
13 Appendices

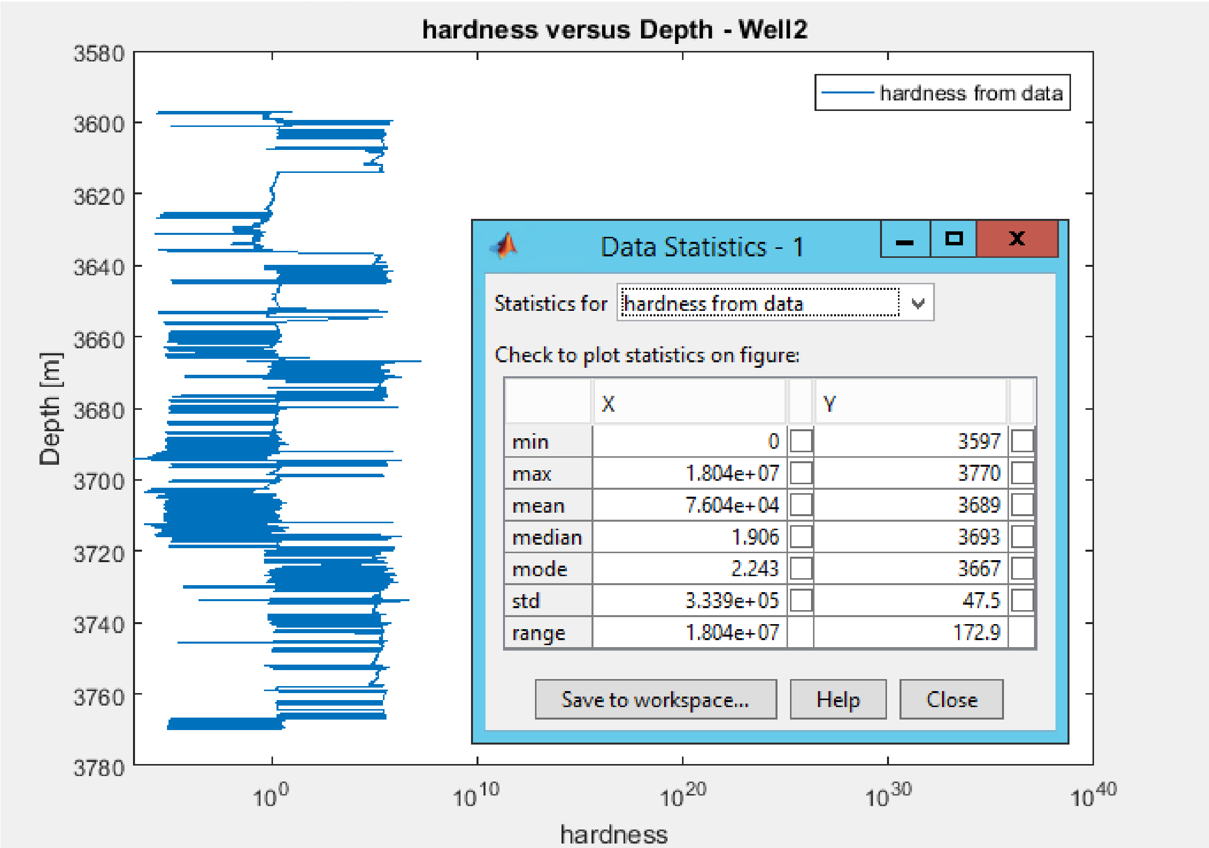
13.1 Appendix I – Matlab code to compute hardness along full well path and output plot of hardness versus depth for 10 wells

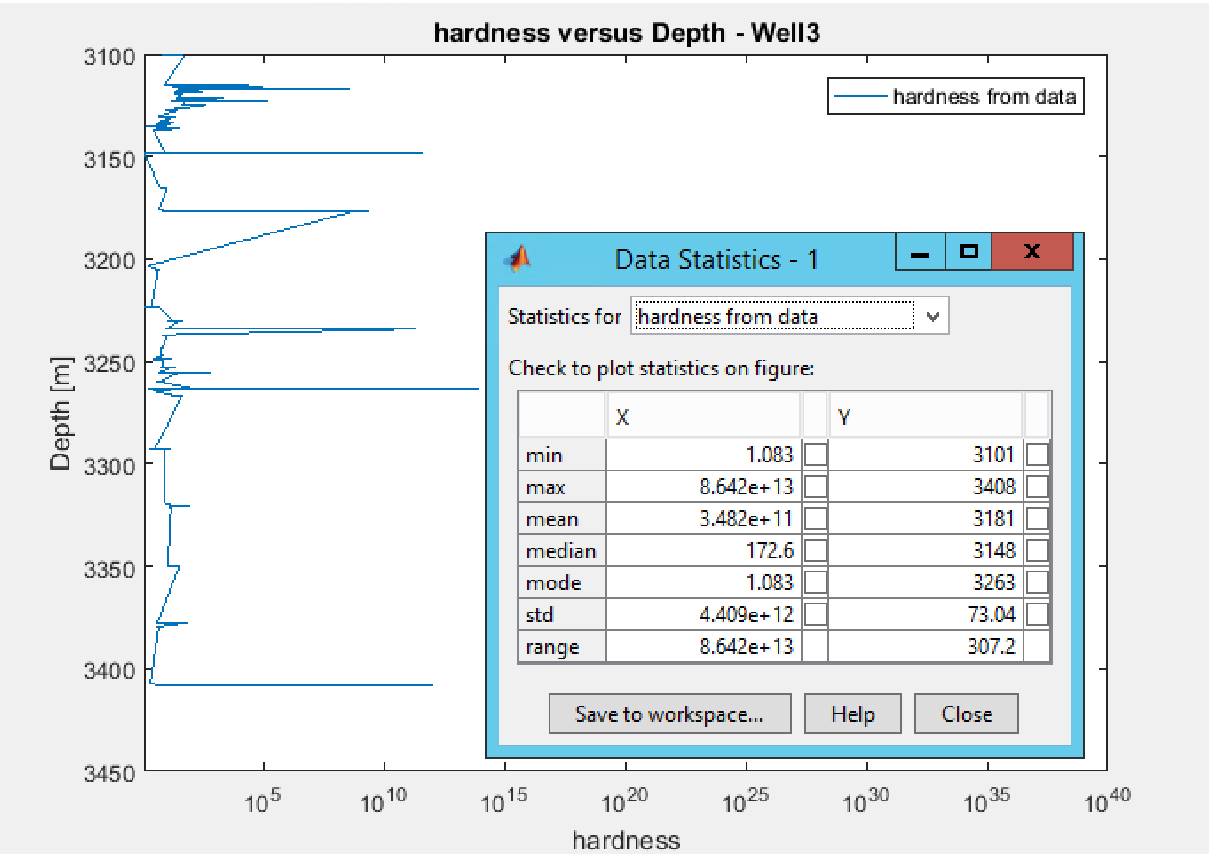
```
%Matlab code to compute hardness along full wellpath
%and plot it versus depth
%
%This example uses well 1 but you just have to change well name to use it
%
%delete previous operation made in Matlab
clear
clc
%load the workspace in this example the file name is well1.mat
load Well1.mat
%WARNING : The program only works if the first column of RTDD curves
%are TIME,DBTM,DMEA,ROP,WOB, RPM in the same order!
%Be sure to have the appropriate file .mat before launch
%Test if the file have the proper shape
test1=strcmp(RTDD.curve_info(1,1),'Time');
test2=strcmp(RTDD.curve_info(2,1),'DBTM');
test3=strcmp(RTDD.curve_info(3,1),'DMEA');
test4=strcmp(RTDD.curve_info(4,1),'ROP');
test5=strcmp(RTDD.curve_info(5,1),'WOB');
test6=strcmp(RTDD.curve_info(6,1),'RPM');
Test= [test1 test2 test3 test4 test5 test6];
if Test==[1 1 1 1 1 1]
    disp(['the file uploaded is on the good format']);
else disp(['the file uploaded is not on the good format']);
end
clear test1 test2 test3 test4 test5 test6 Test
%Assign ROP,RPM,WOB,DMEA,DBTM,TIME into separated lists
LengthRTDD = structfun(@(field) length(field),RTDD);
NumberData=LengthRTDD(length(LengthRTDD));
Time1=[RTDD.curves(1:NumberData)];
DBTM1=[RTDD.curves(NumberData+1:2*NumberData)];
DMEA1=[RTDD.curves(2*NumberData+1:3*NumberData)];
ROP1=[RTDD.curves(3*NumberData+1:4*NumberData)];
WOB1=[RTDD.curves(4*NumberData+1:5*NumberData)];
RPM1=[RTDD.curves(5*NumberData+1:6*NumberData)];
clear RTDD LengthRTDD
%Remove data from other operations than drilling creating clean lists
Time2=[];DBTM2=[];DMEA2=[];ROP2=[];WOB2=[];RPM2=[];
for i = 1:NumberData
    if DBTM1(i)==DMEA1(i)
        Time2=[Time2 Time1(i)];
        DBTM2=[DBTM2 DBTM1(i)];
        DMEA2=[DMEA2 DMEA1(i)];
        ROP2=[ROP2 ROP1(i)];
        WOB2=[WOB2 WOB1(i)];
        RPM2=[RPM2 RPM1(i)];
    end
end
clear Time1 DBTM1 DMEA1 ROP1 WOB1 RPM1
%Create Bourgoyne and Young coefficients
a5_average=1.1;
a6_average=0.65;
%Compute hardness along the path
HardnessFromData=[];
```

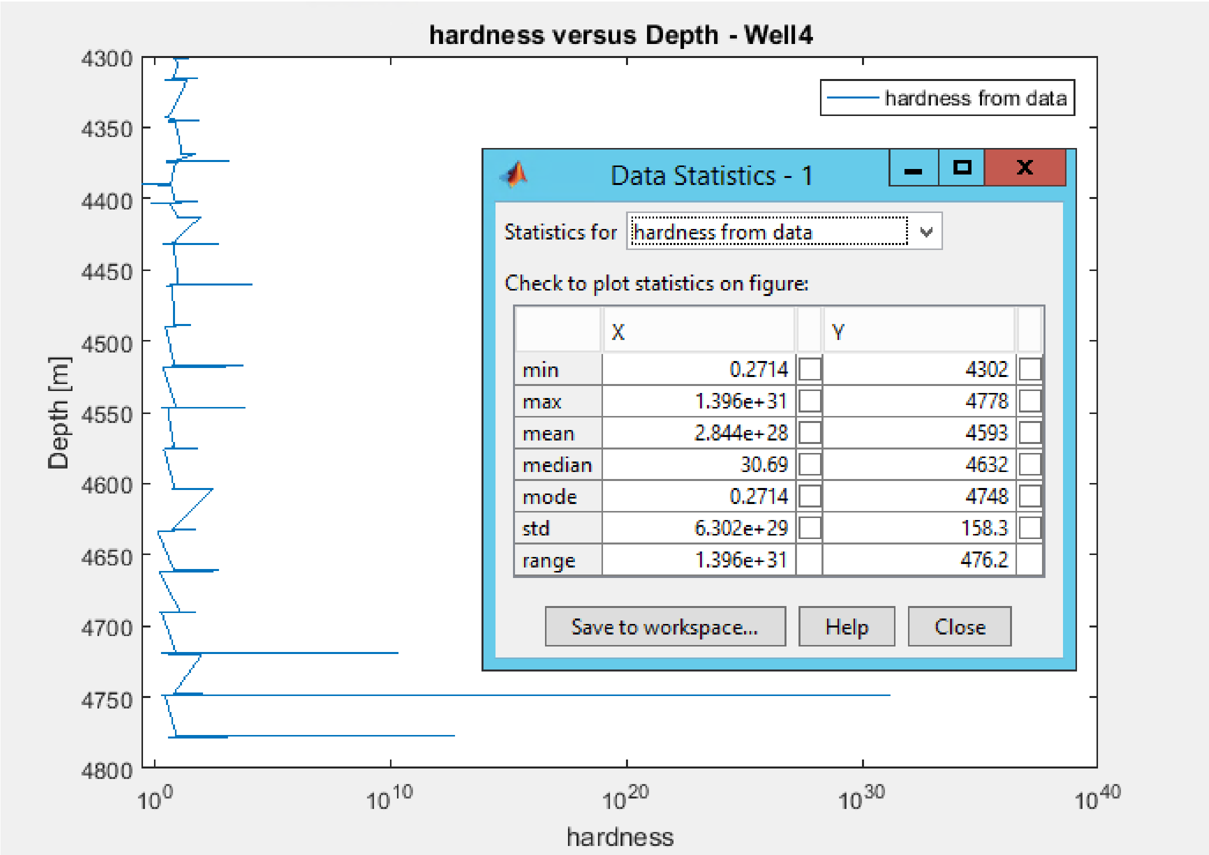
```
DepthHardness=[];
TimeHardness=[];
for i=1:length(Time2)
    hardness=abs(((WOB2(i)^a5_average)*(RPM2(i)^a6_average))/(ROP2(i)));
    if hardness < Inf
    HardnessFromData=[HardnessFromData hardness];
    DepthHardness=[DepthHardness DMEA2(i)];
    TimeHardness=[TimeHardness Time2(i)];
    end
end
figure (1)
semilogx(HardnessFromData,DepthHardness)
title('hardness versus Depth - Well1')
set(gca,'YDir','reverse')
xlabel('hardness')
xlim([0 10^40])
ylabel('Depth [m]')
legend('hardness from data')
%Provide hardness boundaries
maxhardness=max(HardnessFromData)
minhardness=min(HardnessFromData)
medhardness=median(HardnessFromData)
logmaxhardness=log(max(HardnessFromData))
logminhardness=log(min(HardnessFromData))
logmedhardness=log(median(HardnessFromData))
```

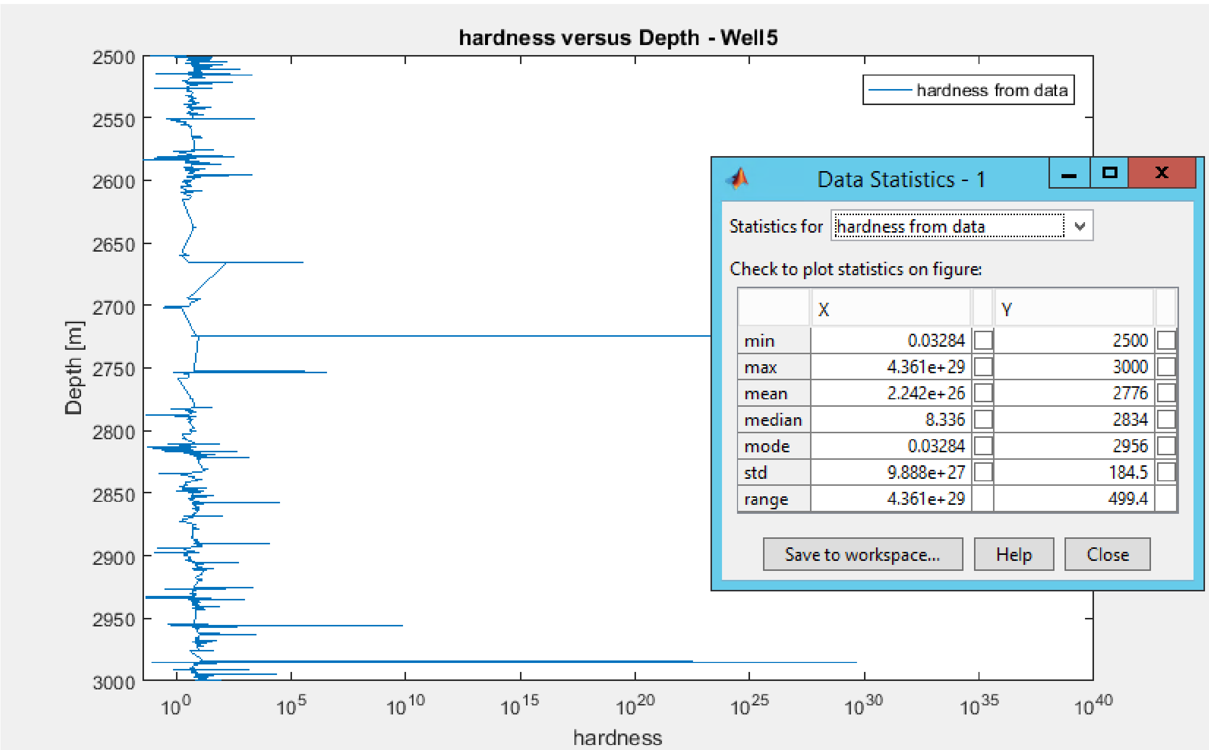
Outputs for the 10 Wells:

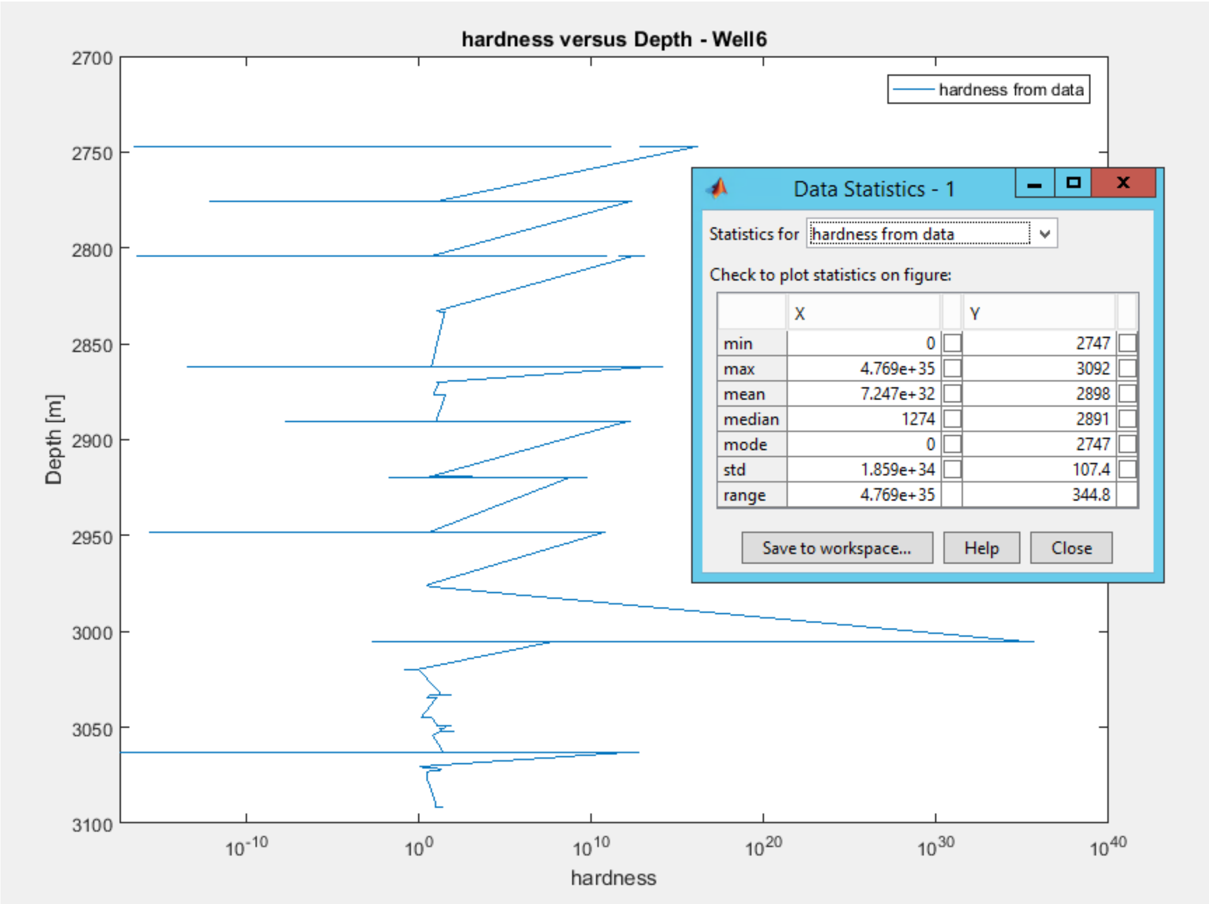


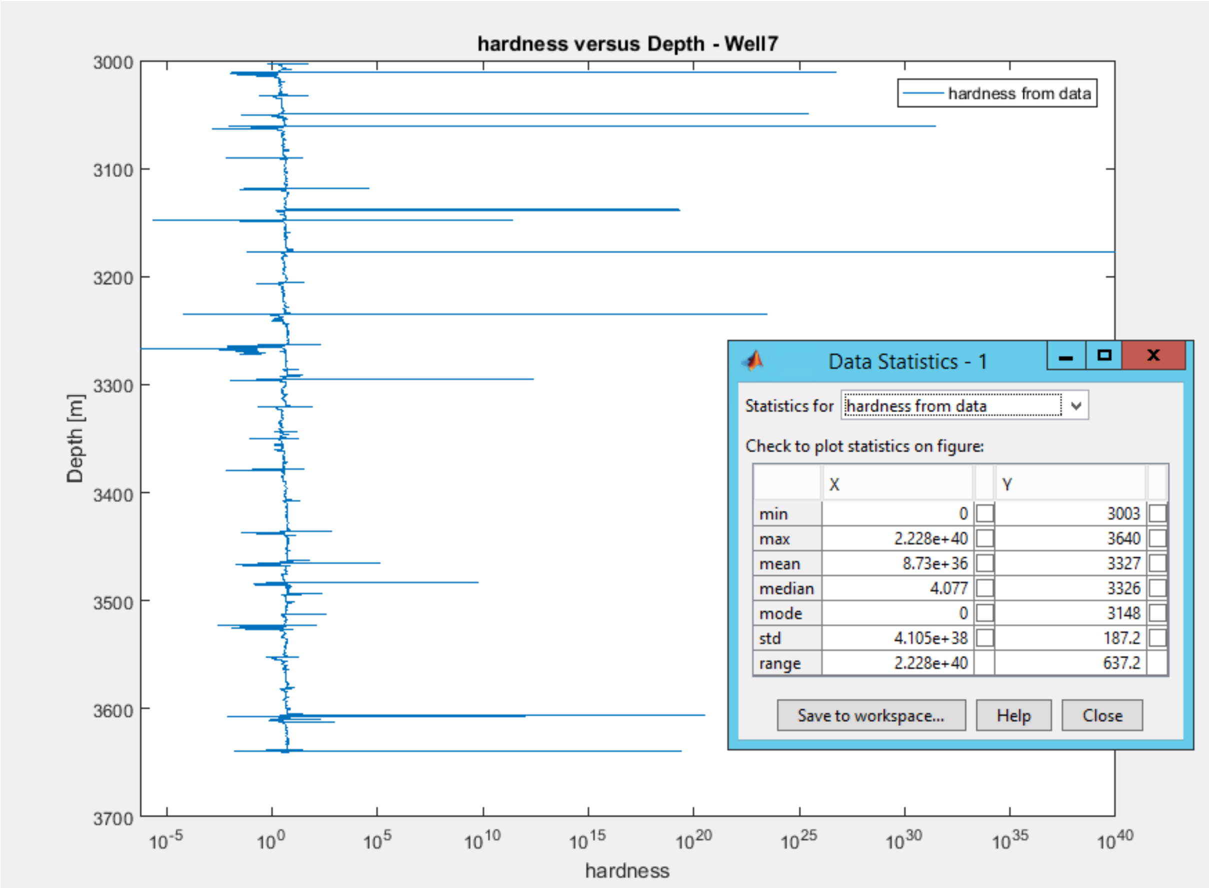

hardness versus Depth - Well2
hardness from data
Depth [m]
hardness
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X
Y
min 0 3597
max 1.804e+07 3770
mean 7.604e+04 3689
median 1.906 3693
mode 2.243 3667
std 3.339e+05 47.5
range 1.804e+07 172.9
Save to workspace...
Help
Close

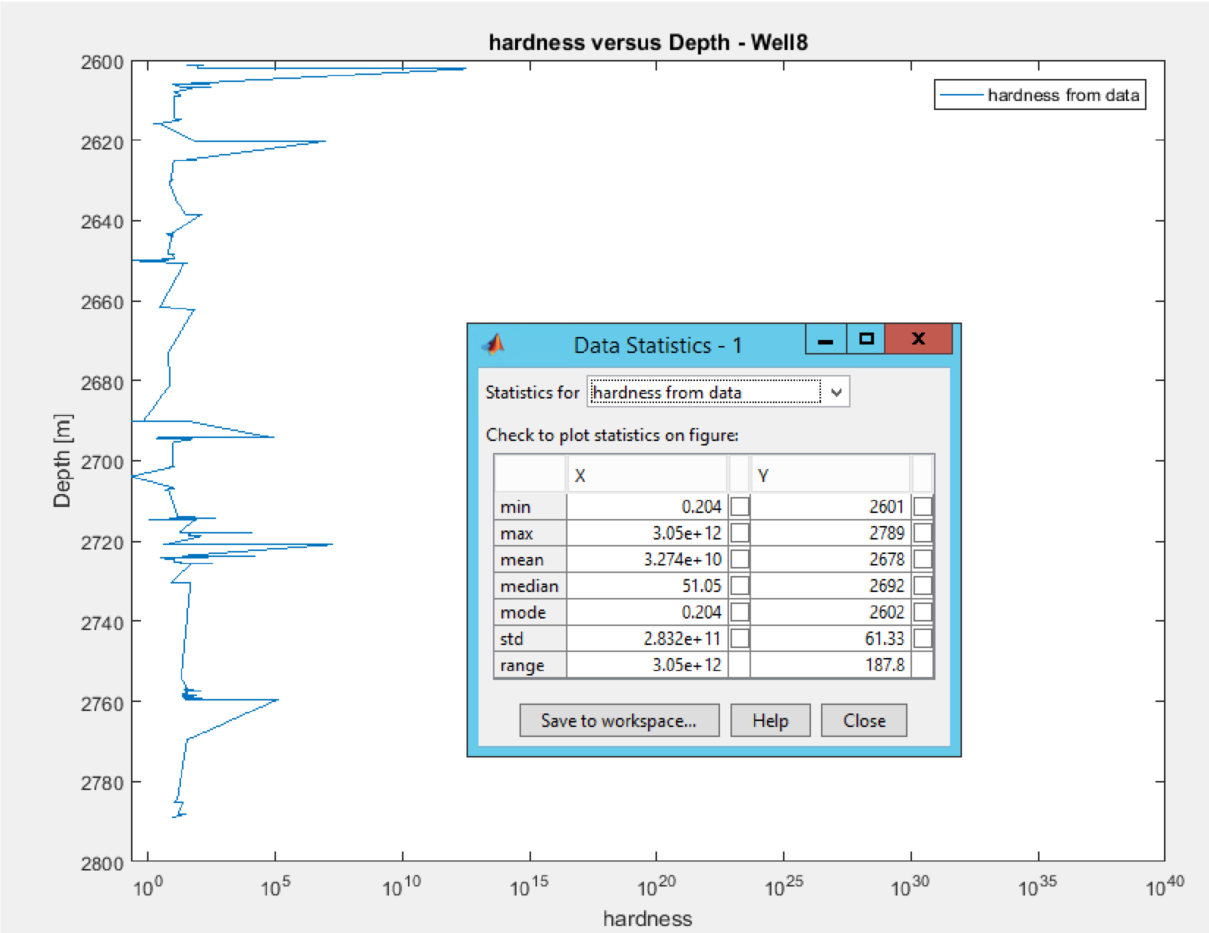

hardness versus Depth - Well3
hardness from data
Depth [m]
hardness
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X
Y
min 1.083 3101
max 8.642e+13 3408
mean 3.482e+11 3181
median 172.6 3148
mode 1.083 3263
std 4.409e+12 73.04
range 8.642e+13 307.2
Save to workspace...
Help
Close

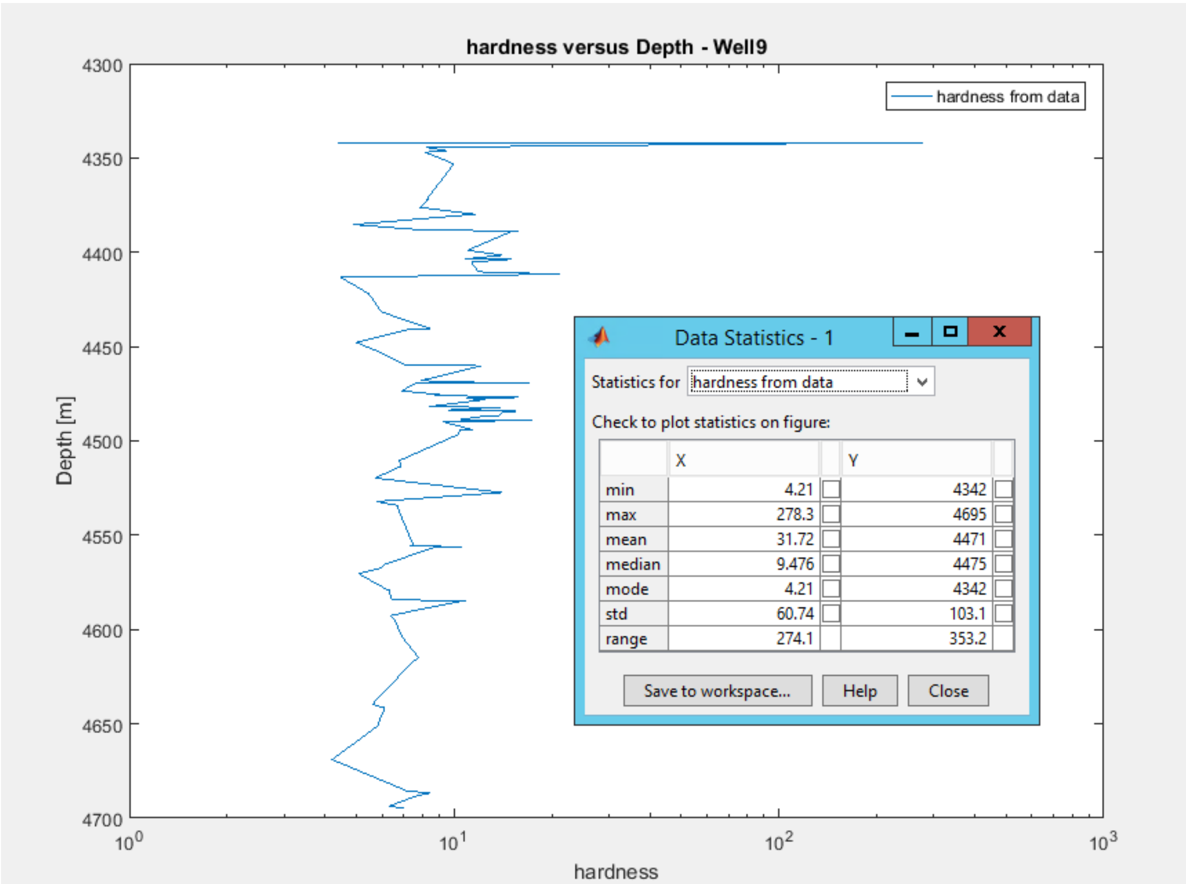

hardness versus Depth - Well4
hardness from data
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X Y
min 0.2714 4302
max 1.396e+31 4778
mean 2.844e+28 4593
median 30.69 4632
mode 0.2714 4748
std 6.302e+29 158.3
range 1.396e+31 476.2
Save to workspace... Help Close
Depth [m]
4300 4350 4400 4450 4500 4550 4600 4650 4700 4750 4800
10^{0} 10^{10} 10^{20} 10^{30} 10^{40}
hardness

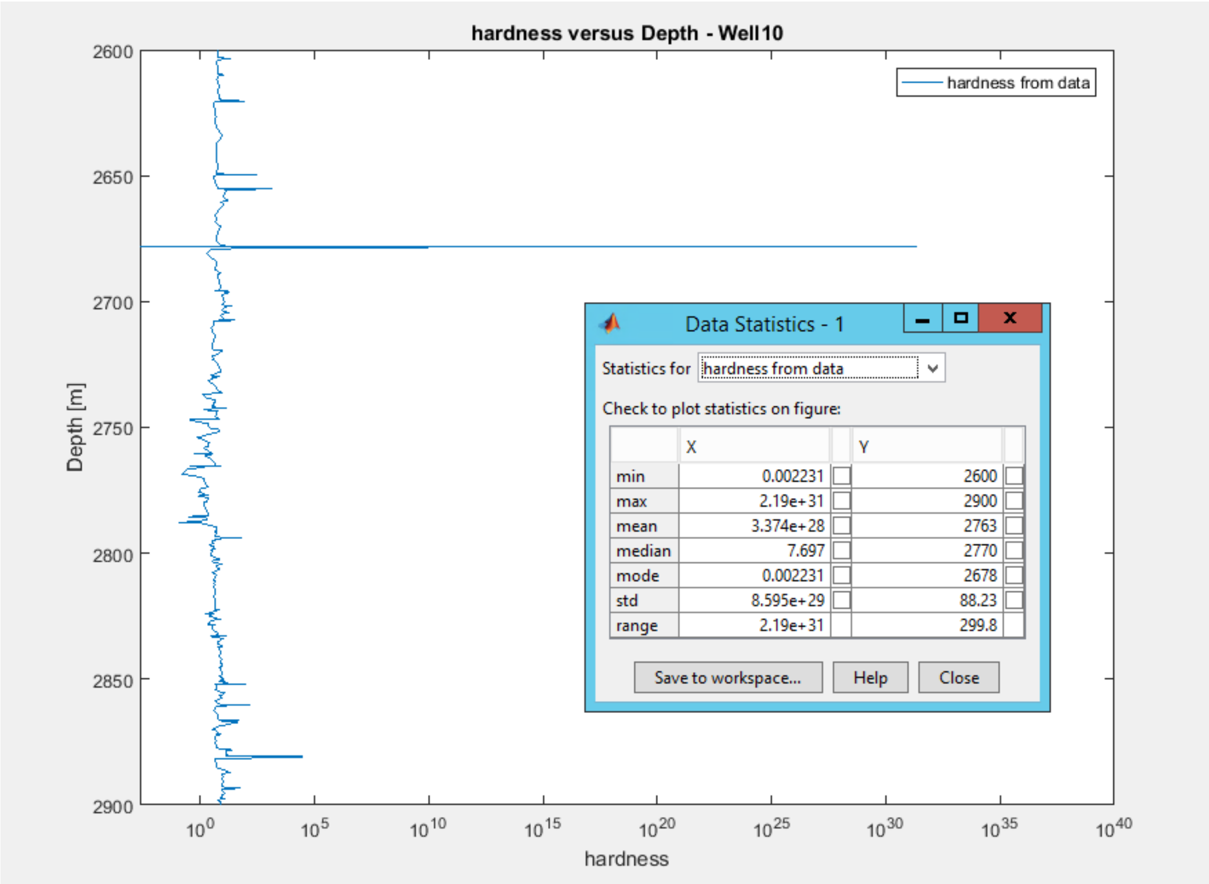

hardness versus Depth - Well5
hardness from data
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X Y
min 0.03284 2500
max 4.361e+29 3000
mean 2.242e+26 2776
median 8.336 2834
mode 0.03284 2956
std 9.888e+27 184.5
range 4.361e+29 499.4
Save to workspace... Help Close
Depth [m]
2500 2550 2600 2650 2700 2750 2800 2850 2900 2950 3000
10^{0} 10^{5} 10^{10} 10^{15} 10^{20} 10^{25} 10^{30} 10^{35} 10^{40}
hardness


hardness versus Depth - Well6
hardness from data
Depth [m]
hardness
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X
Y
min 0 2747
max 4.769e+35 3092
mean 7.247e+32 2898
median 1274 2891
mode 0 2747
std 1.859e+34 107.4
range 4.769e+35 344.8
Save to workspace...
Help
Close


hardness versus Depth - Well7
hardness from data
Depth [m]
hardness
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X
Y
min 0 3003
max 2.228e+40 3640
mean 8.73e+36 3327
median 4.077 3326
mode 0 3148
std 4.105e+38 187.2
range 2.228e+40 637.2
Save to workspace...
Help
Close


hardness versus Depth - Well8
hardness from data
Depth [m]
hardness
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X Y
min 0.204 2601
max 3.05e+12 2789
mean 3.274e+10 2678
median 51.05 2692
mode 0.204 2602
std 2.832e+11 61.33
range 3.05e+12 187.8
Save to workspace...
Help
Close


hardness versus Depth - Well9
hardness from data
Depth [m]
hardness
Data Statistics - 1
Statistics for hardness from data
Check to plot statistics on figure:
X Y
min 4.21 4342
max 278.3 4695
mean 31.72 4471
median 9.476 4475
mode 4.21 4342
std 60.74 103.1
range 274.1 353.2
Save to workspace...
Help
Close



13.2 Appendix II – Matlab code to compute median hardness using only first fifty meters drilled

```
%Matlab code to compute median hardness using first 50 m drilled
%delete previous operation made in Matlab
clear
clc
%load the workspace in this example the file name is well48A.mat
load Well1.mat
%WARNING : The program only works if the first column of RTDD curves
%are TIME,DBTM,DMEA,ROP,WOB, RPM in the same order!
%Be sure to have the appropriate file .mat before launch
%Test if the file have the proper shape
test1=strcmp(RTDD.curve_info(1,1),'Time');
test2=strcmp(RTDD.curve_info(2,1),'DBTM');
test3=strcmp(RTDD.curve_info(3,1),'DMEA');
test4=strcmp(RTDD.curve_info(4,1),'ROP');
test5=strcmp(RTDD.curve_info(5,1),'WOB');
test6=strcmp(RTDD.curve_info(6,1),'RPM');
Test= [test1 test2 test3 test4 test5 test6];
if Test==[1 1 1 1 1 1]
    disp(['the file uploaded is on the good format']);
else disp(['the file uploaded is not on the good format']);
end
clear test1 test2 test3 test4 test5 test6 Test
```

```
%Assign ROP,RPM,WOB,DMEA,DBTM,TIME into separated lists
LengthRTDD = structfun(@(field) length(field),RTDD);
NumberData=LengthRTDD(length(LengthRTDD));
Time1=[RTDD.curves(1:NumberData)];
DBTM1=[RTDD.curves(NumberData+1:2*NumberData)];
DMEA1=[RTDD.curves(2*NumberData+1:3*NumberData)];
ROP1=[RTDD.curves(3*NumberData+1:4*NumberData)];
WOB1=[RTDD.curves(4*NumberData+1:5*NumberData)];
RPM1=[RTDD.curves(5*NumberData+1:6*NumberData)];
clear RTDD LengthRTDD
%Remove data from other operations than drilling creating clean lists
Time2=[];DBTM2=[];DMEA2=[];ROP2=[];WOB2=[];RPM2=[];
for i = 1:NumberData
    if DBTM1(i)==DMEA1(i)
        Time2=[Time2 Time1(i)];
        DBTM2=[DBTM2 DBTM1(i)];
        DMEA2=[DMEA2 DMEA1(i)];
        ROP2=[ROP2 ROP1(i)];
        WOB2=[WOB2 WOB1(i)];
        RPM2=[RPM2 RPM1(i)];
    end
end
clear Time1 DBTM1 DMEA1 ROP1 WOB1 RPM1
%Create Bourgoyne and Young coefficients
a5_average=1.1;
a6_average=0.65;
%Compute hardness along the path
HardnessFromData=[];
DepthHardness=[];
TimeHardness=[];
for i=1:length(Time2)
    hardness=abs(((WOB2(i)^a5_average)*(RPM2(i)^a6_average))/(ROP2(i)));
    if hardness < Inf
    HardnessFromData=[HardnessFromData hardness];
    DepthHardness=[DepthHardness DMEA2(i)];
    TimeHardness=[TimeHardness Time2(i)];
    end
end
%Keep data from first 50m drilled
HardnessFirst50m=[ ];
DepthFirst50m=[ ];
TimeFirst50m=[ ];
i=1;
while DepthHardness(i)< DMEA2(1)+50
    HardnessFirst50m=[HardnessFirst50m HardnessFromData(i)];
    DepthFirst50m=[DepthFirst50m DepthHardness(i)];
    TimeFirst50m=[TimeFirst50m TimeHardness(i)];
    i=i+1;
end
%compute median hardness of first 50m drilled and give it back to user
MedHardnessfirst50m=median(HardnessFirst50m)
```

13.3 Appendix III – Matlab code to compare forecasting accuracy using Holt method and ARIMA method for ten wells and comparison plots

13.3.1 Matlab code – forecast using Holt method

```
clear
clc
%First need to compute hardness using code present in Appendix I, save it
%in a file named HardnessFromData.mat
load HardnessFromData.mat
Hardness2=HardnessFromData;
N=length(Hardness2);
prevision=30; %number of the data points in the forecast
datawindow=60;%number of data points used in the RTDD file to make the
%forecast
Hardnessforecastwindow=zeros(datawindow,1);
Timeforecastwindow=zeros(datawindow,1);
TrueHardness=zeros(prevision,1);
TrueTime=zeros(prevision,1);
for i=1:datawindow
    Hardnessforecastwindow(i)=Hardness2(N-prevision-datawindow+i);
    Timeforecastwindow(i)=Time2(N-prevision-datawindow+i);
end
for i=1:prevision
    TrueHardness(i)=Hardness2(N-prevision+i);
    TrueTime(i)=Time2(N-prevision+i);
end
%initialization of a1 and a2 coefficients
a=0;
b=0;
y=Hardnessforecastwindow; %y is the known RTDD data
n=length(y);
yy=zeros(1,n); %yy is the hardness predicted
p=zeros(n,1);
%p and m represents mu and lambda and range from 0 to 1
for p=0:0.1:0.9
    for m=0:0.1:0.9
        for i=2:n
        a1(1)=y(1);
        a2(1)=y(2)-y(1);
        a1(i)=p*y(i)+(1-p)*(a1(i-1)+a2(i-1));
        a2(i)=m*(a1(i)-a1(i-1))+(1-m)*a2(i-1);
        yy(i)=a1(i-1)+a2(i-1);
        e(i)=y(i)-yy(i);
        b(i)=abs(e(i));
        s(i)=sum(b(i))/n;
        end
    s;
    min=s(1);
    %look into the minimum error link to best mu and lambda choice
    for k=1:n
        if s(k)<min
        min=s(k);
        end
    end
    end
end
yy;
```

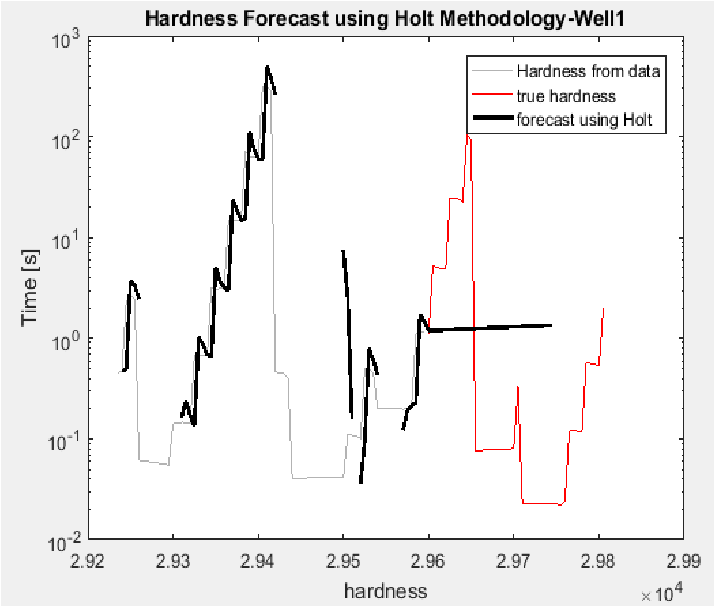
```
%d is the forecast length
d=n+prevision;
k=d-n;
Timeyy=zeros(d,1);
%creation of the time linked to prediction following the 5 seconds time
%step of the RTDD
for i=1:n
    Timeyy(i)=Timeforecastwindow(i);
end
for j=1:k
    Timeyy(j+n)=Timeforecastwindow(n)+j*5;
end
for j=1:k
yy(j+n)=a1(n)+a2(n)*j;
end
yy;
%output is a plot of hardness known before forecast point, followed by a
%plot of planned hardness and hardness known after forecast point
figure(1)
semilogy(Timeforecastwindow,y,'Color',[.7,.7,.7]);
hold on
semilogy(TrueTime,TrueHardness,'r');
hold on
semilogy(Timeyy,yy,'k','LineWidth',2)
legend('Hardness from data','true hardness','forecast using Holt')
title('Hardness Forecast using Holt Methodology - Well1')
xlabel('hardness')
ylabel('Time [s]')
hold off
```

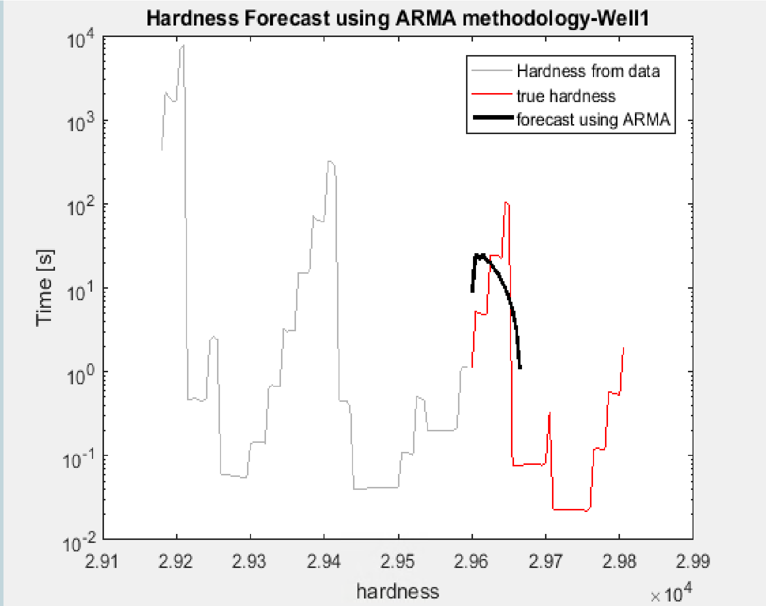
13.3.2 Matlab code – forecast using ARIMA model

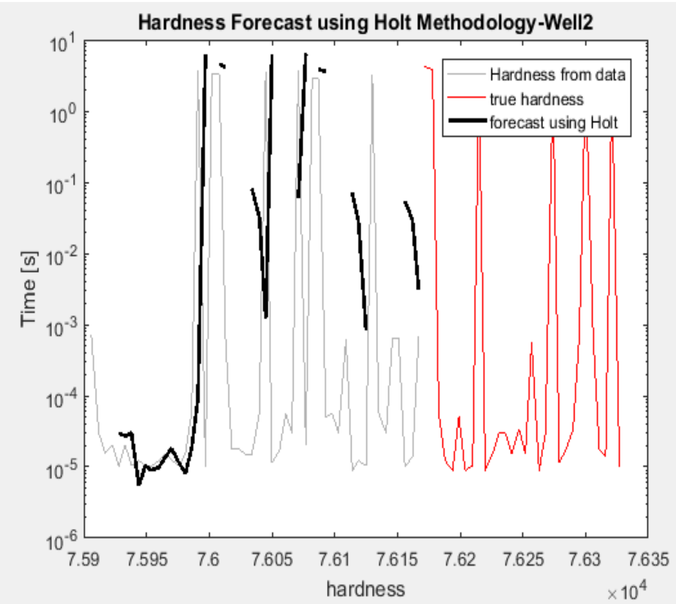
```
clear
clc
%First need to compute hardness using code present in Appendix I, save it
%in a file named HardnessFromData.mat
load HardnessFromData.mat
Hardness2=HardnessFromData;
N=length(Hardness2);
prevision=30; %length of forecast
datawindow=60; %number of data points used to establish the forecast
%creation of different windows, step 3 of the flow chart in figure 3-1
Hardnessforecastwindow=zeros(datawindow,1);
Timeforecastwindow=zeros(datawindow,1);
TrueHardness=zeros(prevision,1);
TrueTime=zeros(prevision,1);
for i=1:datawindow
    Hardnessforecastwindow(i)=Hardness2(N-prevision-datawindow+i);
    Timeforecastwindow(i)=Time2(N-prevision-datawindow+i);
end
for i=1:prevision
    TrueHardness(i)=Hardness2(N-prevision+i);
    TrueTime(i)=Time2(N-prevision+i);
End
%Determination of the good p and q couple
%Range from 1 to 5 and compute for each couple the error with known
hardness
MatrixDeterminepq=zeros(5,5);
for p=1:5
    for q=1:5
```

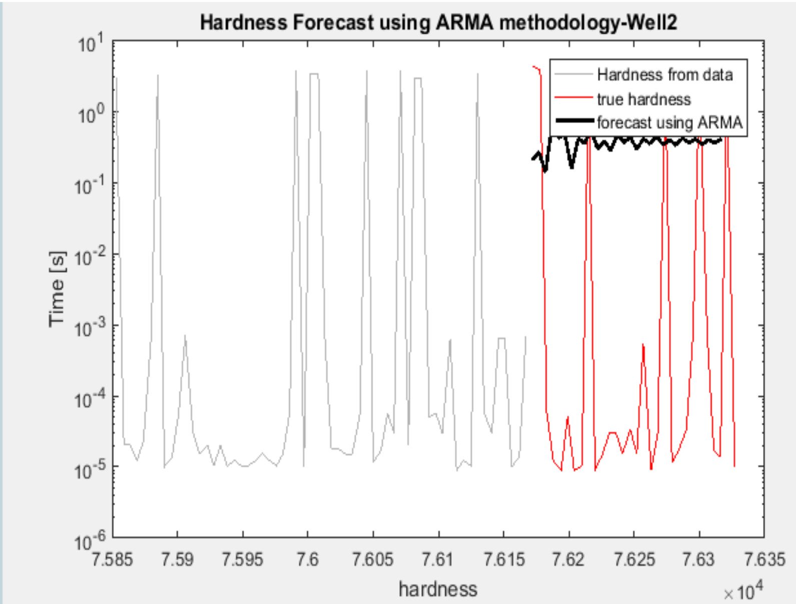
```
        try
y=Hardnessforecastwindow;
x=Timeforecastwindow;
n=length(x);
Model=arima(p,0,q);
Fit=estimate(Model,y);
[yy,ymse] = forecast(Fit,prevision,'Y0',y);
Error=zeros(prevision,1);
for i=1:prevision
    Error(i)=sqrt(abs((yy(i)).^2-(TrueHardness(i)).^2));
end
MatrixDeterminepq(p,q)=mean(Error);
        catch
            MatrixDeterminepq(p,q)=Inf;
        end
    end
end
for i=1:5
    for j=1:5
        if MatrixDeterminepq(i,j)== Inf
            MatrixDeterminepq(i,j)=999999999999999;
        end
    end
end
%find the couple of p and q predicting the smallest error
[p,q] = find(MatrixDeterminepq == min(abs(MatrixDeterminepq(:))));
y=Hardnessforecastwindow;
x=Timeforecastwindow;
n=length(x);
%Recompute the ARIMA model using the good p and q choice
Model=arima(p,0,q);
Fit=estimate(Model,y);
[yy,ymse] = forecast(Fit,prevision,'Y0',y);
Timeyy=zeros(prevision,1);
for j=1:prevision
    Timeyy(j)=x(n)+j*5;
End
%plot the final prediction
figure(1)
semilogy(x,y,'Color',[.7,.7,.7]);
hold on
semilogy(TrueTime,TrueHardness,'r');
hold on
semilogy(Timeyy,yy,'k','LineWidth',2);
legend('Hardness from data','true hardness','forecast using ARMA')
title('Hardness Forecast using ARMA methodology - Well1')
xlabel('hardness')
ylabel('Time [s]')
hold off
```

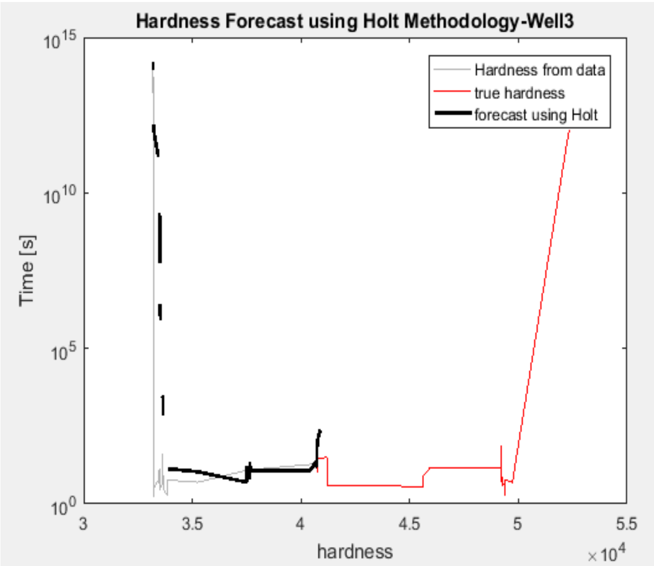
13.3.3 Plots of forecast on 10 wells using Holt and ARIMA method

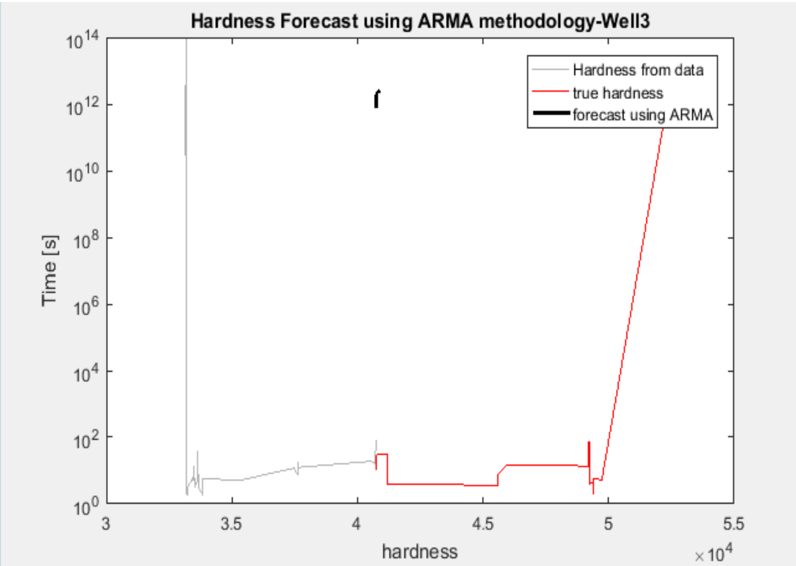


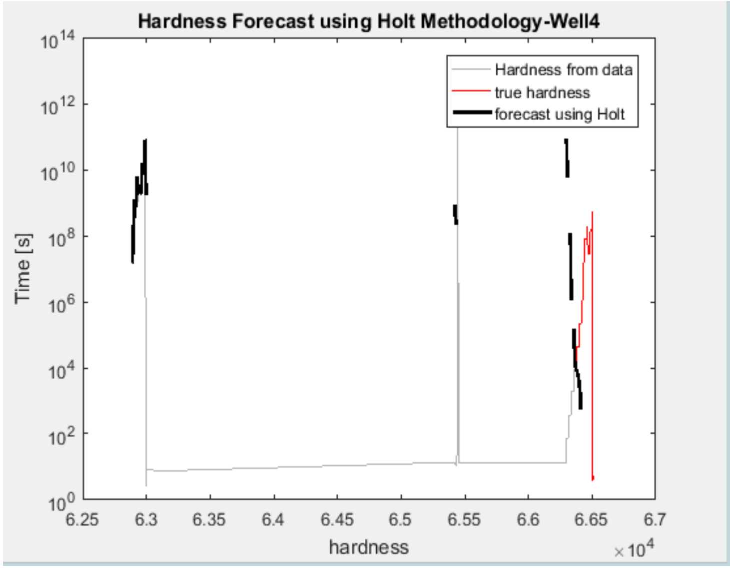


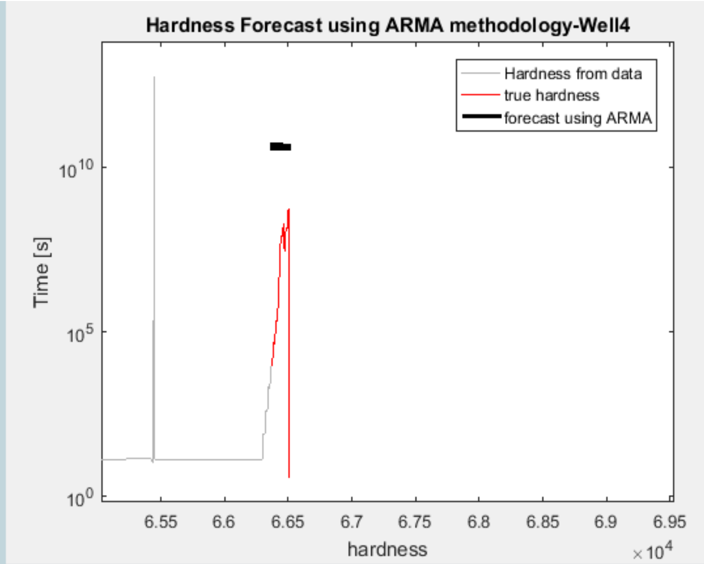


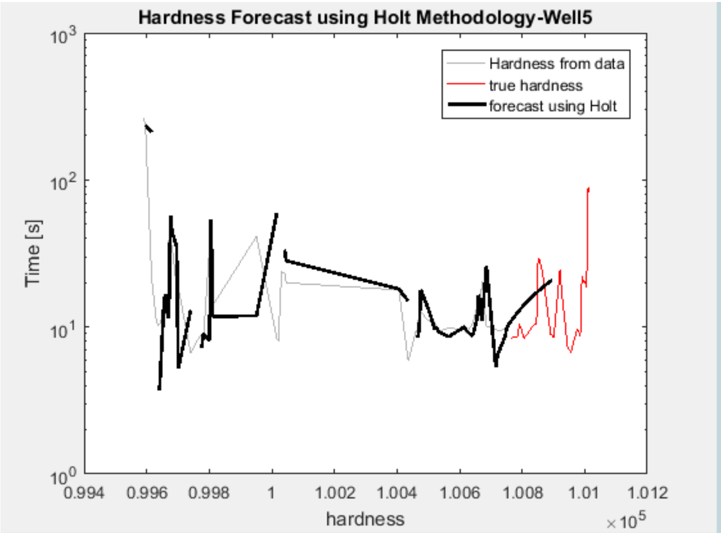


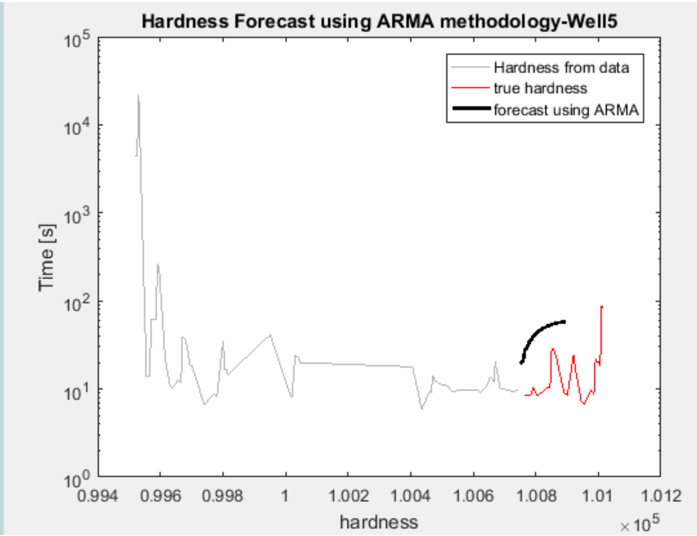


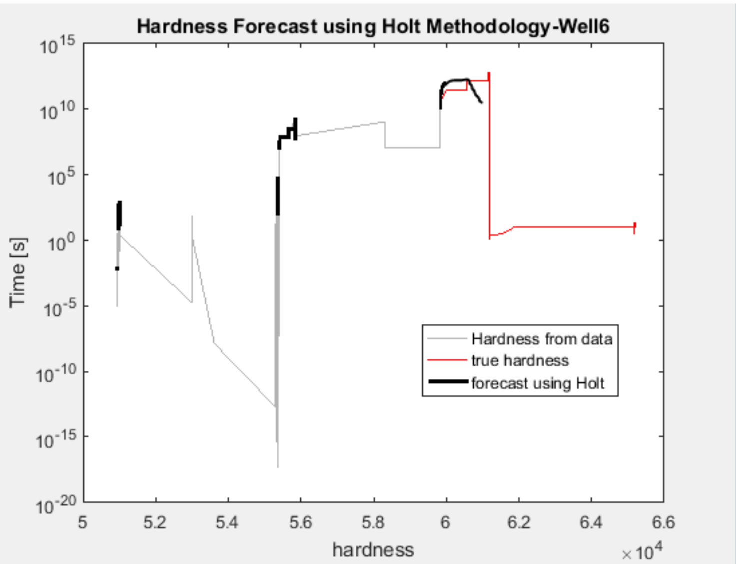


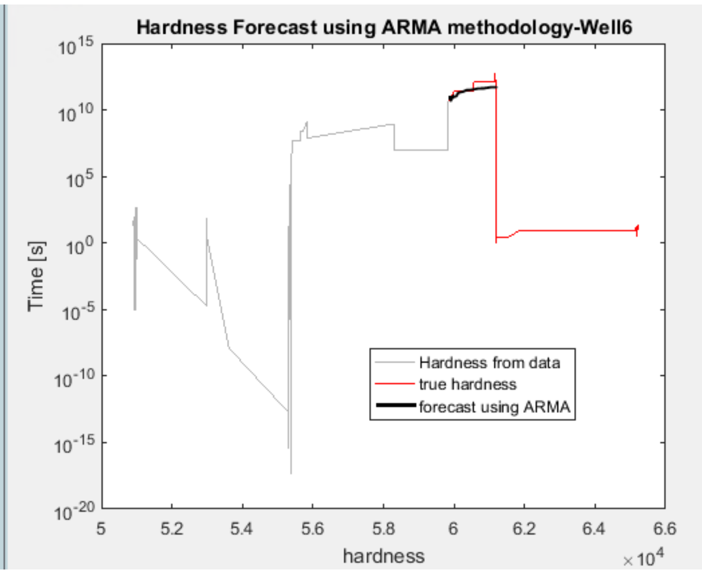

Hardness Forecast using Holt Methodology-Well4
Hardness from data
true hardness
forecast using Holt
Time [s]
hardness

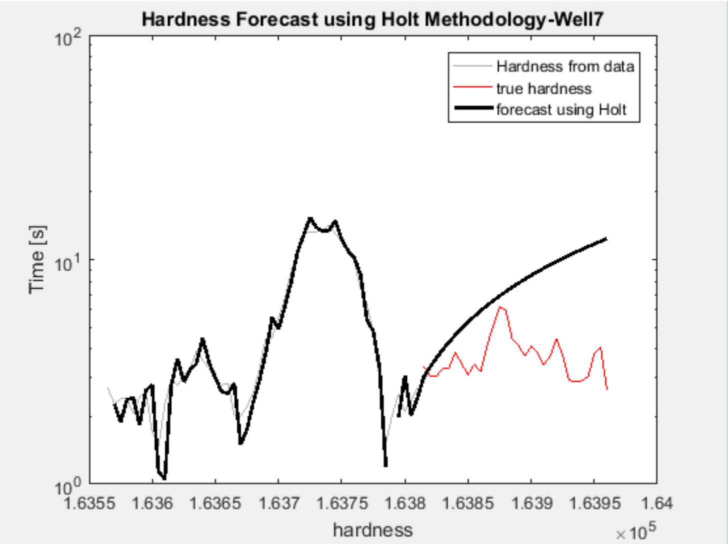

Hardness Forecast using ARMA methodology-Well4
Hardness from data
true hardness
forecast using ARMA
Time [s]
hardness

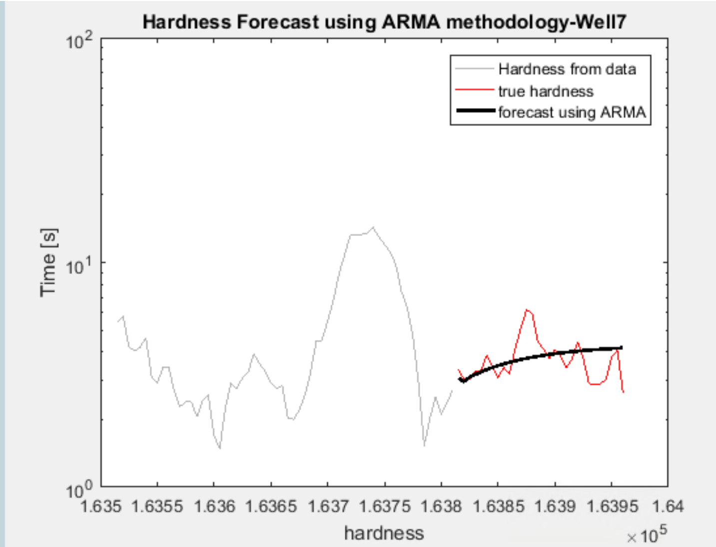

Hardness Forecast using Holt Methodology-Well5
Hardness from data
true hardness
forecast using Holt
Time [s]
hardness

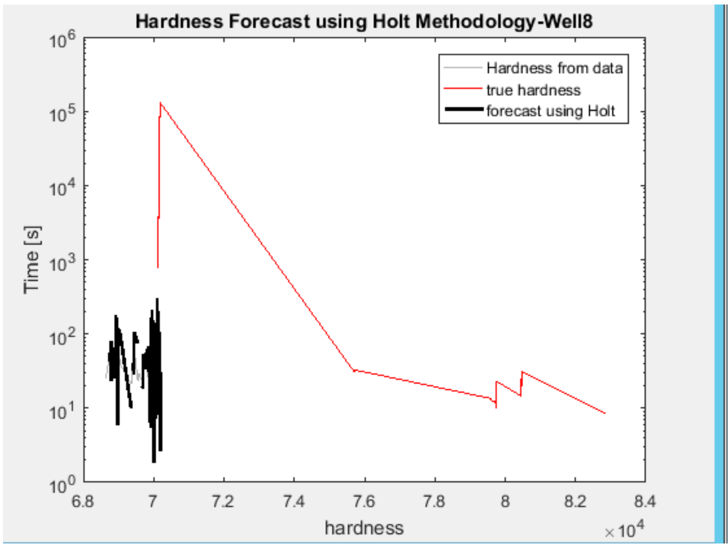

Hardness Forecast using ARMA methodology-Well5
Hardness from data
true hardness
forecast using ARMA
Time [s]
hardness

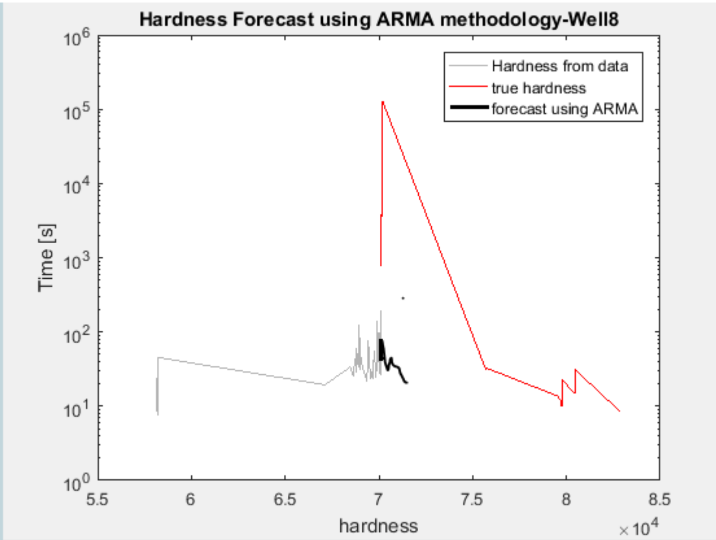

Hardness Forecast using Holt Methodology-Well6
Hardness from data
true hardness
forecast using Holt
Time [s]
hardness

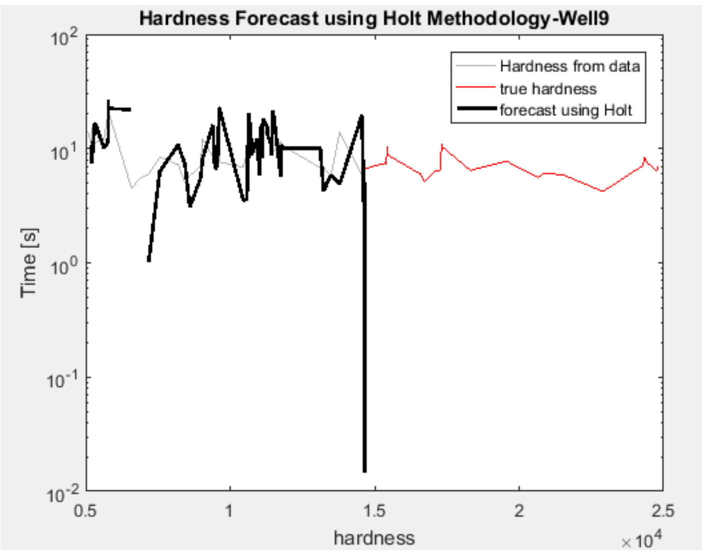

Hardness Forecast using ARMA methodology-Well6
Hardness from data
true hardness
forecast using ARMA
Time [s]
hardness

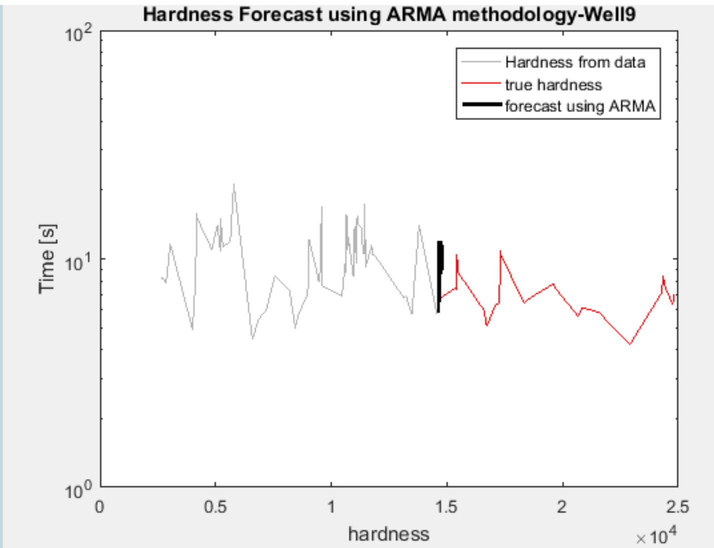

Hardness Forecast using Holt Methodology-Well7
Hardness from data
true hardness
forecast using Holt
Time [s]
hardness

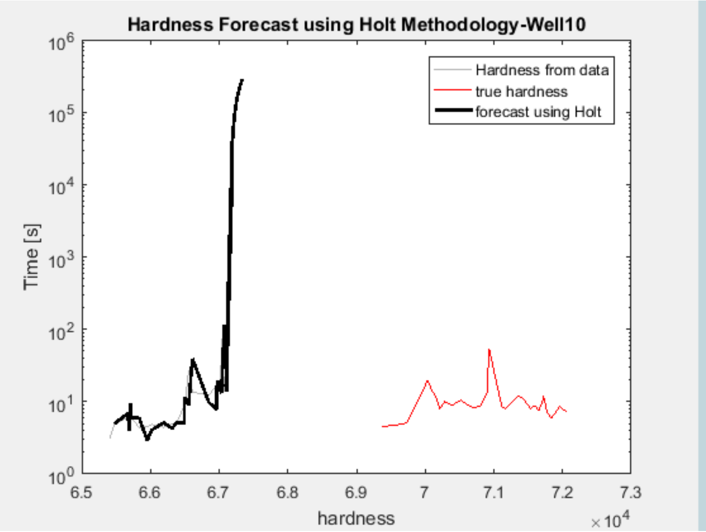

Hardness Forecast using ARMA methodology-Well7
Hardness from data
true hardness
forecast using ARMA
Time [s]
hardness

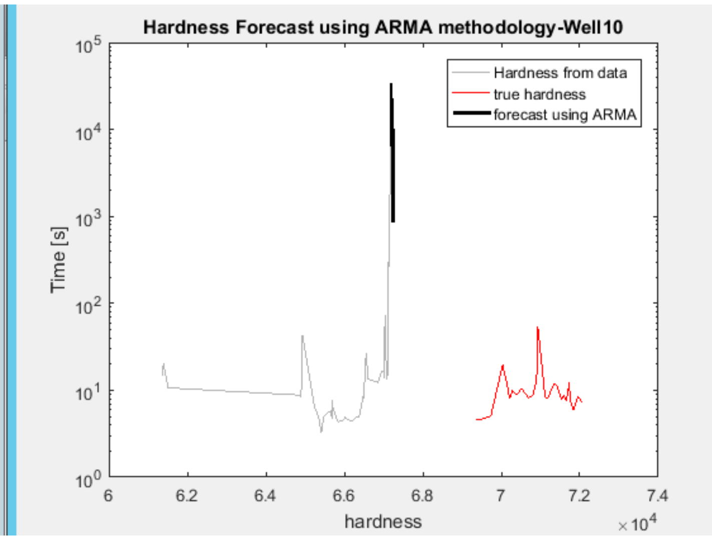

Hardness Forecast using Holt Methodology-Well8
Hardness from data
true hardness
forecast using Holt
Time [s]
hardness


Hardness Forecast using ARMA methodology-Well8
Hardness from data
true hardness
forecast using ARMA
Time [s]
hardness


Hardness Forecast using Holt Methodology-Well9
Hardness from data
true hardness
forecast using Holt
Time [s]
hardness


Hardness Forecast using ARMA methodology-Well9
Hardness from data
true hardness
forecast using ARMA
Time [s]
hardness





13.4 Appendix IV – Examples of the six different formats of ASC files present in the Diskos database

13.4.1 ASC format 0: Standardized format

```
"DATE","TIME","LOGTIME","ACTC","BDIA","DBTM","DBTV","DMEA","DVER","RSU","RSD","ROP","BPOS","HKL","WOB","TRQ","RPMA",
"RPMB","BROT","BDTI","TBR","SPP","CEPP","WHP","KLP","CHP","CCVL","CFO","CFI","CDI","CDO","CTVL","TVA","TPVT","ETPT",
"MFO","MFI","MDO","MDI","MTO","MTI","ECDB","ECDM","GAS"
dd-mmm-yy, hh:mm:ss, s, , IN, M, M, M, M, M/S, M/S, M/HR, M, TON, TON, KNM, RPM, RPM, HR, HR, REV,
 BAR, BAR, BAR, BAR, BAR, M3, LPM, LPM, G/C3, G/C3, M3, M3, M3, M3, %, LPM, G/C3, G/C3, DEGC, DEGC,
G/C3, G/C3, %
"23-Nov-03"," 10:00:00     ",754048800,-9999,-9999,2404.2770507812502,-9999,2406.95654296875,2126.2375859999997,
-9999,-0.64208275627461264,0,13.303145122528075,96.026315307617196,-0.60605773925781248,8578.8706975448422,
59.545126721249609,59.38132972418753,-9999,-9999,-9999,225.98673400878909,226.16076660156253,-9999,0.55999922752380371,
0.48890455067157751,-9999,-9999,1.0550464153289798,-9999,-9999,75.626690673828122,102.50142898559571,8.5950116157531742,
-9999,60.740158843994138,2007.3708618164062,0.18143595606088639,1.5654279708862306,43.427998352050849,18.600000381469783,
1.6821726000000001,2.7100000381469727,0.15999999642372131
"23-Nov-03"," 10:00:05     ",754048805,-9999,-9999,2404.3155273437501,-9999,2406.95654296875,2126.2375859999997,-9999,
-1.2653549283742904,0,13.223429584503174,96.204715728759766,-0.78445816040039063,8588.9639451249532,59.989501568639376,
60.219175717187412,-9999,-9999,-9999,226.1962844848633,226.36309204101565,-9999,0.55783452987670901,0.49005224406719211,
-9999,-9999,1.055016064643860 2,-9999,-9999,75.794700622558594,102.41070327758788,8.5954158782958991,-9999,61.218560791015626,
2008.0048583984376,0.1815094262361 5265,1.5665871381759646,43.471999359130905,18.592000198364303,1.6828889999999999,2.7100000381469727,
0.15199999809265136
```

13.4.2 ASC format 1: Header absent, quotation mark (") in Mnemonics is absent, space is used as separator instead of coma and time and date format is missing

```
      Time            Hole     ROP      Flow      WOB      RPM       Torque     SPP
     & Date           Depth    Inst   In Pum Avg  Avg   Surface Av Abs Avg     Avg
      ( )             (m)     (m/hr)    (gpm)    (klb)    (rpm)     (f-p)     (psig)
 21:09:00 07-Sep-01   46.397   0.000      0.0      0.0         0     350.4          0
 21:09:05 07-Sep-01   60.475   0.000      0.0      0.0         0     350.4          0
 21:09:10 07-Sep-01   73.866   0.000      0.0      0.0         0     350.4          0
 21:09:15 07-Sep-01   84.295   0.000      0.0      0.0         0     350.4          0
 21:09:20 07-Sep-01   92.417   0.000      0.0      0.0         0     350.5          0
 21:09:25 07-Sep-01   98.743   0.000      0.0      0.0         0     350.4          0
 21:09:30 07-Sep-01  103.669   0.000      0.0      0.0         0     350.4          0
 21:09:35 07-Sep-01  107.506   0.000      0.0      0.0         0     350.4          0
 21:09:40 07-Sep-01  110.494   0.000      0.0      0.0         0     350.4          0
 21:09:45 07-Sep-01  112.821   0.000      0.0      0.0         0     350.5          0
 21:09:50 07-Sep-01  114.633   0.000      0.0      0.0         0     350.5          0
 21:09:55 07-Sep-01  116.045   0.000      0.0      0.0         0     350.5          0
 21:10:00 07-Sep-01  117.144   0.000      0.0      0.0         0     350.5          0
 21:10:05 07-Sep-01  118.000   0.000      0.0      0.0         0     350.5          0
```

13.4.3 ASC format 2: Header absent, a unwanted info column is present, quotation mark (") in Mnemonics is absent, Units before data is absent, space is used as separator instead of coma and time and date is expressed in UNIX

```
UTIM Unix Time sec
DATE Date ddmmyy
TIME Time hhmmss
WAC Wits Activity Code unitless
BDIA Bit Diameter inch
DBTM Bit Measured Depth m
DBTV Bit Vertical Depth m
DMEA Hole Measured Depth m
DVER Hole Vertical Depth m
RSU Pulling Speed m/sec
RSD Running Speed m/sec
SWAB Swab Pressure Gradient g/cc
SURG Surge Pressure Gradient g/cc
ROP ROP m/hr
BPOS Block Position m
HKL Hookload tons
WOB Weight on Bit tons
TRQ Torque kNm
RPMA String RPM rpm
RPMB Bit RPM rpm
BROT Rotation Time hr
BDTI Bit Drilled Time hr
TBR Total Bit Revolutions unitless
SPP Pump Pressure bar
COPP Completion Pump Pressure bar
CEPP Cement Pump Pressure bar


UTIM DATE TIME WAC BDIA DBTM DBTV DMEA DVER RSU RSD SWAB SURG ROP BPOS HKL WOB TRQ RPMA RPMB BROT BDTI TBR SPP COPP CEPP
1060025758 04Aug03 20-35-58 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 33.08 0.00 0.00 0 0 0.00 0.00 0 1.4
1060040881 05Aug03 00-48-01 3 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.53 32.62 0.00 0.00 -0 0 8597.00 0.00 0 1.6
1060071348 05Aug03 09-15-48 3 0.00 334.43 334.43 600.00 600.00 0.38 0.00 0.00 0.00 0.00 10.97 33.99 0.00 0.00 -0 0 0.00 0.00
1060071353 05Aug03 09-15-53 3 0.00 334.43 334.43 600.00 600.00 0.38 0.00 0.00 0.00 0.00 9.72 32.16 0.00 0.00 -0 0 0.00 0.00
1060071358 05Aug03 09-15-58 3 0.00 334.43 334.43 600.00 600.00 0.38 0.00 0.00 0.00 0.00 9.30 32.62 0.00 0.00 -0 0 0.00 0.00
1060071363 05Aug03 09-16-03 3 0.00 334.43 334.43 600.00 600.00 0.38 0.00 0.00 0.00 0.00 9.31 32.71 0.00 0.00 -0 0 0.00 0.00
1060071368 05Aug03 09-16-08 3 0.00 334.43 334.43 600.00 600.00 0.38 0.00 0.00 0.00 0.00 9.27 33.26 0.00 0.00 -0 0 0.00 0.00
```

13.4.4 ASC format 3: Similar at format 2, moreover, mnemonics are hidden in the unwanted info column

```
UTIM Unix Time sec
DATE Date ddmmyy
TIME Time hhmmss
WAC Wits Activity Code un
BDIA Bit Diameter inch
DBTM Bit Measured Depth m
DBTV Bit Vertical Depth m
DMEA Hole Measured Depth m
DVER Hole Vertical Depth m
RSU Pulling Speed m/sec
RSD Running Speed m/sec
SWAB Swab Pressure Gradient g/cc
SURG Surge Pressure Gradient g/cc
ROP ROP m/hr
BPOS Block Position m
HKL Hookload tons
WOB Weight on Bit tons
TRQ Torque kNm
RPMA String RPM rpm
RPMB Bit RPM rpm
BROT Rotation Time hr
BDTI Bit Drilled Time hr
TBR Total Bit Revolutions unitless
SPP Pump Pressure bar
COPP Completion Pump Pressure bar
CEPP Cement Pump Pressure bar
WHP Wellhead Pressure bar
KLP Kill Line Pressure bar
UTIM DATE TIME WAC BDIA DBTM DBTV DMEA DVER RSU RSD SWAB SURG ROP BPOS HKL WOB TRQ RPMA RPMB BROT BDTI TBR SPP COPP CEPP |
c1                                      ppm
c2                                      ppm
c3                                      ppm
ic4                                     ppm
nc4                                     ppm
ic5                                     ppm
nc5                                     ppm
nec5                                    ppm
          959306472 26May00  04-01-12     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306477 26May00  04-01-17     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306482 26May00  04-01-22     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306487 26May00  04-01-27     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306493 26May00  04-01-33     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306498 26May00  04-01-38     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306503 26May00  04-01-43     3         16.000000       418.880000       418.830000       418.880000       418.830000
          959306508 26May00  04-01-49     3         16.000000       418.880000       418.830000       418.880000       418.830000
```

13.4.5 ASC format 4: Similar at format 2, except that header is present but settled between two lines of stars (****)

```
|*****************************************************************************************

Well:           25/8-C-14
Path:           25/8-C-14 A
Hole:           25/8-C-14 A
Start:  Wednesday, January 04, 2006 /03:44:20
End:            Tuesday, January 10, 2006 /22:00:00
Date/Time:      Thursday, January 12, 2006 /01:12:11
*****************************************************************************************


TIME                 ACTINNX                  ACTHTFX                  ACTECDR                  APRESR                   TCDM
DD/MM/YY-hh:mm:ss  deg                      deg                      sg                       bar                      degC

04/01/06-03:44:20 -999.25                  -999.25                  -999.25                  -999.25                  33.238
04/01/06-03:44:21 -999.25                  -999.25                  -999.25                  -999.25                  -999.25
04/01/06-03:44:21 -999.25                  -999.25                  -999.25                  -999.25                  -999.25
04/01/06-03:44:33 -999.25                  -999.25                  -999.25                  -999.25                  -999.25
04/01/06-03:44:33 -999.25                  -999.25                  -999.25                  -999.25                  -999.25
04/01/06-03:44:45 -999.25                  -999.25                  -999.25                  -999.25                  -999.25
```

13.4.6 ASC format 5: date and time are in UNIX

```
"TIME","ACTC","BDIA","DBTM","DBTV","DMEA","DVER","RSU","RSD","ROP","BPOS","HKL","WOB",
"TRQ","RPMA","RPMB","BROT","BDTI","TBR","SPP","CEPP","WHP","KLP","CHP","CCVL","CFO",
"CFI","CDI","CDO","CTVL","TVA","TPVT","ETPT","MFO","MFI","MDO","MDI","MTO","MTI","ECDB",
"ECDM","GAS"
S, , IN, M, M, M, M, M/S, M/S, M/HR, M, TON, TON, KNM, RPM, RPM, HR, HR, REV, BAR,
BAR, BAR, BAR, BAR, M3, LPM, LPM, G/C3, G/C3, M3, M3, M3, M3, %, LPM, G/C3, G/C3,
DEGC, DEGC, G/C3, G/C3, %
740522700,-9999,-9999,2472.9005371093749,-9999,2475.9990234375,2300.8725469999995,-9999,
-7.7580493450164791,0,19.459481239318848,76.133058929443365,0.6999488830566406,0,9.3093141163675813E-2,
130.63396148024893,-9999,-9999,-9999,132.02207489013674,0.83999996599943749,-9999,0.12776692733168604,
0.32687114477157597,-9999,-9999,-4514.4003295898437,-9999,-9999,49.342634201049805,97.083441162109381,
61.646697100471059,-9999,25.889782905578613,1332.0340576171875,0.94940083622932436,0.79904001951217651,
36.580591964721748,32.522991180419979,0.90471950000000001,1305.050048828125,1.9999999552965164E-2
740522705,-9999,-9999,2473.6344238281249,-9999,2475.9990234375,2300.8725469999995,-9999,
-5.3691394805908201,0,18.820430374145509,76.094559478759763,0.73844833374023444,0,
9.8662936682509306E-2,130.66111144362648,-9999,-9999,-9999,131.76077117919922,0.83999996599943749,
-9999,8.9879918843507781E-2,0.32535477280616765,-9999,-9999,-4514.4003295898437,-9999,-9999,49.444075775146487,
97.171360015869141,61.646697100471059,-9999,25.463933944702148,1332.0340332031251,0.94934335947036741,
0.79899201989173896,36.560840988159271,32.533918762207122,0.90469929999999987,1305.050048828125,1.9999999552965164E-2
740522710,-9999,-9999,2474.2527343749998,-9999,2475.9990234375,2300.8725469999995,-9999,
-3.8839585065841673,0,18.387715148925782,76.203631591796878,0.62937622070312504,0,9.6096304054197487E-2,
130.79844054518207,-9999,-9999,-9999,130.89162902832032,0.83999996599943749,-9999,6.9625290110707283E-2,
0.32328421473503116,-9999,-9999,-4514.4003295898437,-9999,-9999,49.561365127563477,97.029121398925781,
```

13.5 Appendix V: Anisa Noor Corina Matlab code to clean raw ASC files from type 2 and 5

13.5.1 Anisa Noor Corina Matlab code to clean raw ASC files from type 2

```matlab
function fix_asc_type2(file_input,file_output)
% Check this condition:
% This code is valid if the data contains UTIM as the first line
% This code is valid if the second and third line are DATE and TIME
% This code will not fix the unit of the date and time. Make sure to check
% the unit of date and time with the data!!
% Example:
% fix_asc_type2('my_file.ASC')
%    this command will replace the old file with the edited one
% fix_asc_type2('my_file.ASC','my_new_file.ASC')
%    this command will create a new file that has been edited
% Open file
fid       = fopen(file_input);
if fid < 0
    error('Cannot open file :');
    file_output = [];
    return
end
tlines=textscan(fid,'%s','delimiter','\n');
fclose(fid);
tlines=tlines{:};
% Fix tlines
dummy = find(cellfun(@isempty,tlines));
tlines(dummy) = [];
% Find the number of mnemonics available and the position
FirstLine = textscan(tlines{1},'%s','delimiter',' ');
FirstMnem = FirstLine{1}{1};
MnemPos = find(~cellfun(@isempty,regexpi(tlines,FirstMnem)));
NMnem = MnemPos(end)-1;
```

```
% Get the units
Unit = cell(1,NMnem);
for i=1:NMnem
    dummy = textscan(tlines{i},'%s','delimiter',' ');
    Unit(i)=upper(dummy{1}(end));
end
Unit = strjoin(Unit(2:end),',');
% Fix mnemonic format
Mnem=textscan(tlines{MnemPos(end)},'%s','delimiter',' ');
Mnem=Mnem{:};
NewMnem = ['"',strjoin(Mnem(2:end),'","'),'"'];
% Fixing header (mnemonics and units)
NewLines = cell(length(tlines)-NMnem+1,1);
NewLines{1} = NewMnem;
NewLines{2} = Unit;
% Fix time and date in the data
j = 3;
for i=NMnem+2 : length(tlines)
    dummy = textscan(tlines{i},'%s','delimiter',' ');
    dummy = dummy{:};
    NewLines{j} =
['"',dummy{2},'","',dummy{3},'",',strjoin(dummy(4:end),',')];
    j = j+1;
end
% Saving the file
if isequal(nargin,1)
    fid = fopen(file_input,'w');
    frewind(fid);
else
    fid = fopen(file_output,'w');
end
fprintf(fid,'%s \r\n',NewLines{:});
fclose(fid);
end
```

13.5.2 Anisa Noor Corina Matlab code to clean raw ASC files from type 5

```
function fix_asc_type5(file_input,file_output)
% Codes for fixing type 5 ASC file
% Example:
% fix_asc_type2('my_file.ASC')
%    this command will replace the old file with the edited one
% fix_asc_type2('my_file.ASC','my_new_file.ASC')
%    this command will create a new file that has been edited
% Open file
fid      = fopen(file_input);
if fid < 0
    error('Cannot open file :');
    file_output = [];
    return
end
tlines=textscan(fid,'%s','delimiter','\n');
fclose(fid);
tlines=tlines{:};
% Fix tlines
dummy = find(cellfun(@isempty,tlines));
tlines(dummy) = [];
% Fix Mnemonic Name: adding DATE and TIME
Mnem=textscan(tlines{1},'%s','delimiter',',');
Mnem=Mnem{:};
NewMnem = ['"DATE","TIME",',strjoin(Mnem(2:end),',')];
```

```matlab
% Fix Unit
Unit=textscan(tlines{2},'%s','delimiter',',');
Unit=Unit{:};
NewUnit = ['dd-mmm-yy,HH:MM:SS,',strjoin(Unit(2:end),',')];
% Fixing header (mnemonics and units)
NewLines = cell(length(tlines),1);
NewLines{1} = NewMnem;
NewLines{2} = NewUnit;
% Fix time and date in the data
for i=3 : length(tlines)
    dummy = textscan(tlines{i},'%s','delimiter',',');
    dummy = dummy{:};
    DateTime = datetime(str2num(dummy{1}),'convertfrom','posixtime');
    Date = datestr(DateTime,'dd-mmm-yy');
    Time = datestr(DateTime,'HH:MM:SS');
    NewLines{i} = ['"',Date,'","',Time,'",',strjoin(dummy(2:end),',')];
end
% Saving the file
if isequal(nargin,1)
    fid = fopen(file_input,'w');
    frewind(fid);
else
    fid = fopen(file_output,'w');
end
fprintf(fid,'%s \r\n',NewLines{:});
fclose(fid);
end
```

13.6 Appendix VI: Procedure to convert ASC files to .mat files

13.6.1 Create a .mat file from ASC format (Donne, 2016)

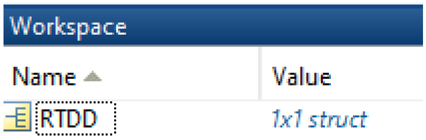
1-In Matlab, launch the log viewer using the program written by Noor Corina (2016) called main_log.

2-Once the log viewer window appears upload a file using “File”-> “Load File”.

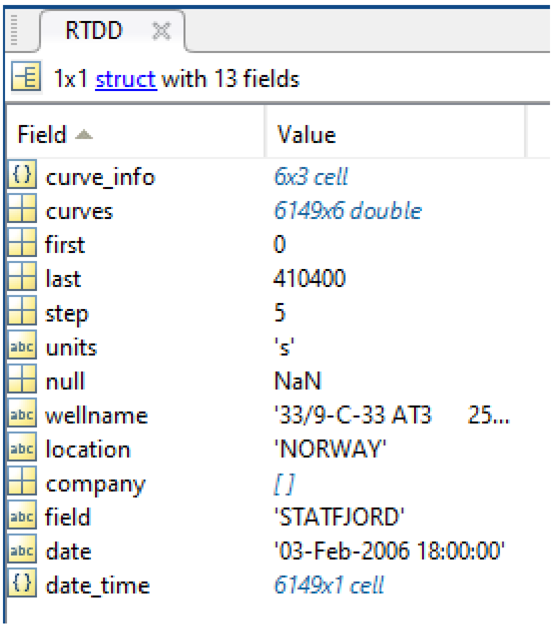
3-When the log viewer displays that the file is successfully loaded save the file using “File”-> “Save File”. The new file saved have now the format .mat wanted

13.6.2 Remove unwanted data from RTDD

1-In Matlab, open your .mat file. It should appear in the workspace



2-Double click on the file, RTDD in the example. This new window appears:



3-Now double click on the curve_info, it opens this new window:

RTDD | RTDD.curve_info

RTDD.curve_info

	1	2	3
1	'Time'	's'	[]
2	'DBTM'	'M'	[]
3	'DMEA'	'M'	[]
4	'ROP'	'M/HR'	[]
5	'WOB'	'TON'	[]
6	'RPM'	'RPM'	[]

4-Now note down the **row number** of the data you want to delete, for example if you want to delete WON and DBTM remember 5 and 2. Then click on curves :

RTDD | RTDD.curve_info | RTDD.curves

RTDD.curves

	1	2	3	4	5	6
1	0	7.0996e+03	7.0996e+03	4.8316	2.7102	61.5415
2	5	7.0996e+03	7.0996e+03	5.1148	2.7994	61.5431
3	10	7.0996e+03	7.0996e+03	5.1148	2.8309	61.5424
4	15	7.0996e+03	7.0996e+03	5.1148	2.8597	61.4069
5	20	7.0996e+03	7.0996e+03	5.1148	2.9253	61.3848
6	25	7.0996e+03	7.0996e+03	5.1148	3.0617	61.2853
7	30	7.0996e+03	7.0996e+03	4.8901	3.0617	62.0768
8	35	7.0996e+03	7.0996e+03	4.8901	3.0250	61.5001
9	40	7.0996e+03	7.0996e+03	4.8901	2.9778	61.8287
10	45	7.0996e+03	7.0996e+03	4.8901	2.8282	61.5463
11	50	7.0996e+03	7.0996e+03	4.8452	2.9620	61.4071
12	55	7.0997e+03	7.0997e+03	4.8452	3.0801	61.5530
13	60	7.0997e+03	7.0997e+03	4.8452	3.0958	61.0065
14	65	7.0997e+03	7.0997e+03	5.1175	3.1536	61.5916
15	70	7.0997e+03	7.0997e+03	5.1175	3.0486	61.5407
16	75	7.0997e+03	7.0997e+03	5.1175	3.0040	60.8295
17	80	7.0997e+03	7.0997e+03	5.1175	2.9778	61.1515
18	85	7.0997e+03	7.0997e+03	5.1175	2.9096	60.9949
19	90	7.0997e+03	7.0997e+03	4.8907	2.9148	61.1779
20	95	7.0997e+03	7.0997e+03	4.8907	2.9410	61.4556
21	100	7.0997e+03	7.0997e+03	4.8907	2.9751	61.5291
22	105	7.0997e+03	7.0997e+03	4.8907	2.9804	61.1956
23	110	7.0997e+03	7.0997e+03	4.8453	2.9096	60.9617

5-Maintains control button while selecting **column** with same number as rows of step 4, here we select column 2 and column 5. We press delete and save the new workspace. The file is simplified.

13.7 Appendix VII – Depths intervals in which the 10 study wells cross similar geological formations

13.7.1 Depths intervals with depth wrote in measured depth (MD) in which the 10 study wells cross similar geological formations

Formation	Wellbores																			
	Well 1		Well 2		Well 3		Well 4		Well 5		Well 6		Well 7		Well 8		Well 9		Well 10	
	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End
Cromer	2190	2200	-	-	3193	3193	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Viking	2200	2595	3557	3566	3193	3211	4261	4310	3423	3439	2727	2737	2996	3034	2487	2490	4252	4296	2673	2674
Brent	2861	2933	3566	3956	3211	3397	4311	4851	3439	3547	2737	2756	3034	3046	2490	2641	4296	4555	2388	2881
Dunlin	2933	3108	3956	4006	3397	3516	-	-	3547	3866	2756	2893	3046	3172	2641	2855	-	-	2881	2910
Statfjord	3018	3304	-	-	-	-	-	-	3866	4160	2893	3132	3172	3595	-	-	-	-	-	-
Lunde	-	-	-	-	-	-	-	-	-	-	-	-	3595	3642	-	-	-	-	-	-

IN THIS TABLE ALL DEPTHS ARE MD DEPTH

window studied

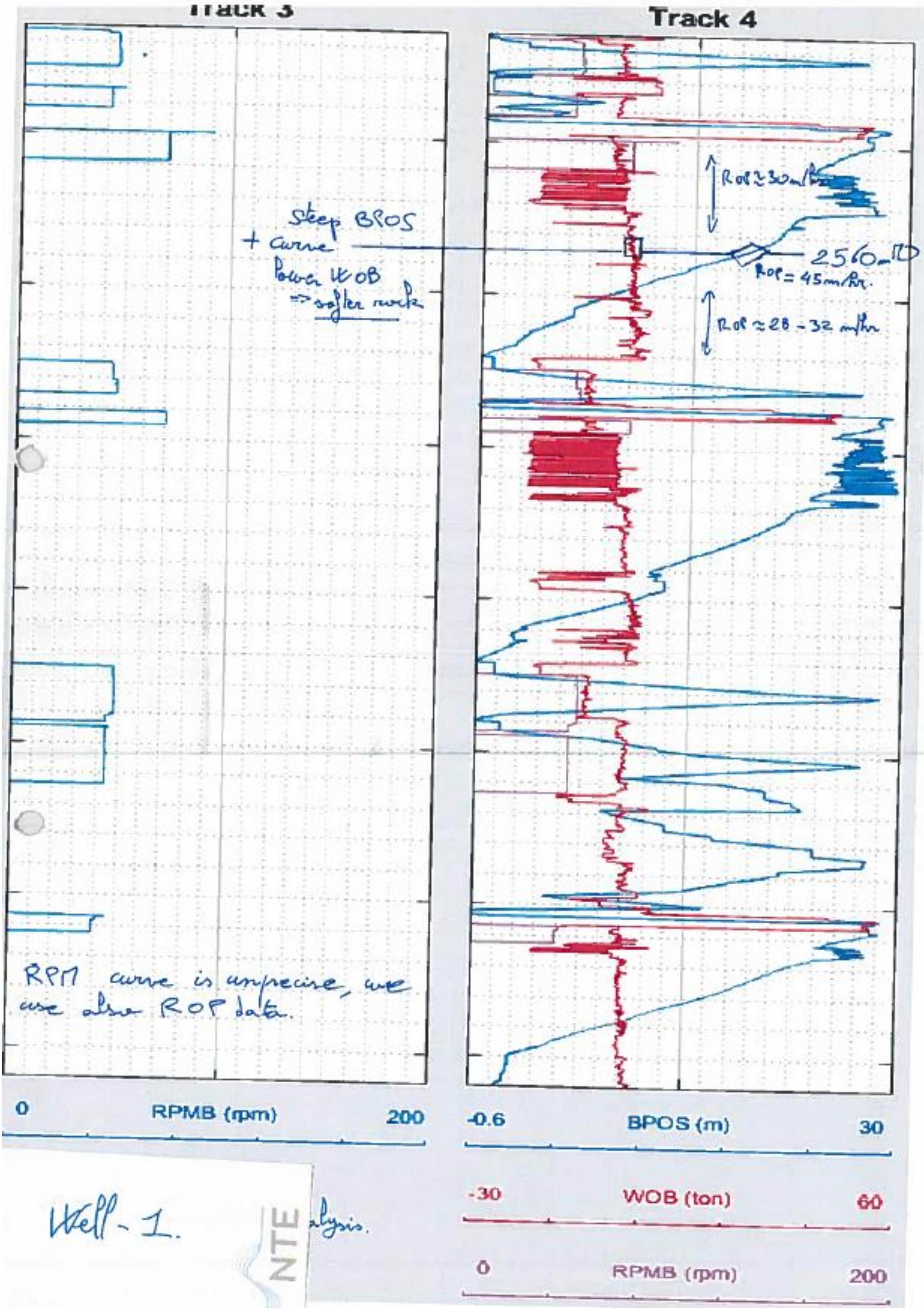
13.7.2 Depths intervals with depth wrote in true depth (TD) in which the 10 study wells cross similar geological formations

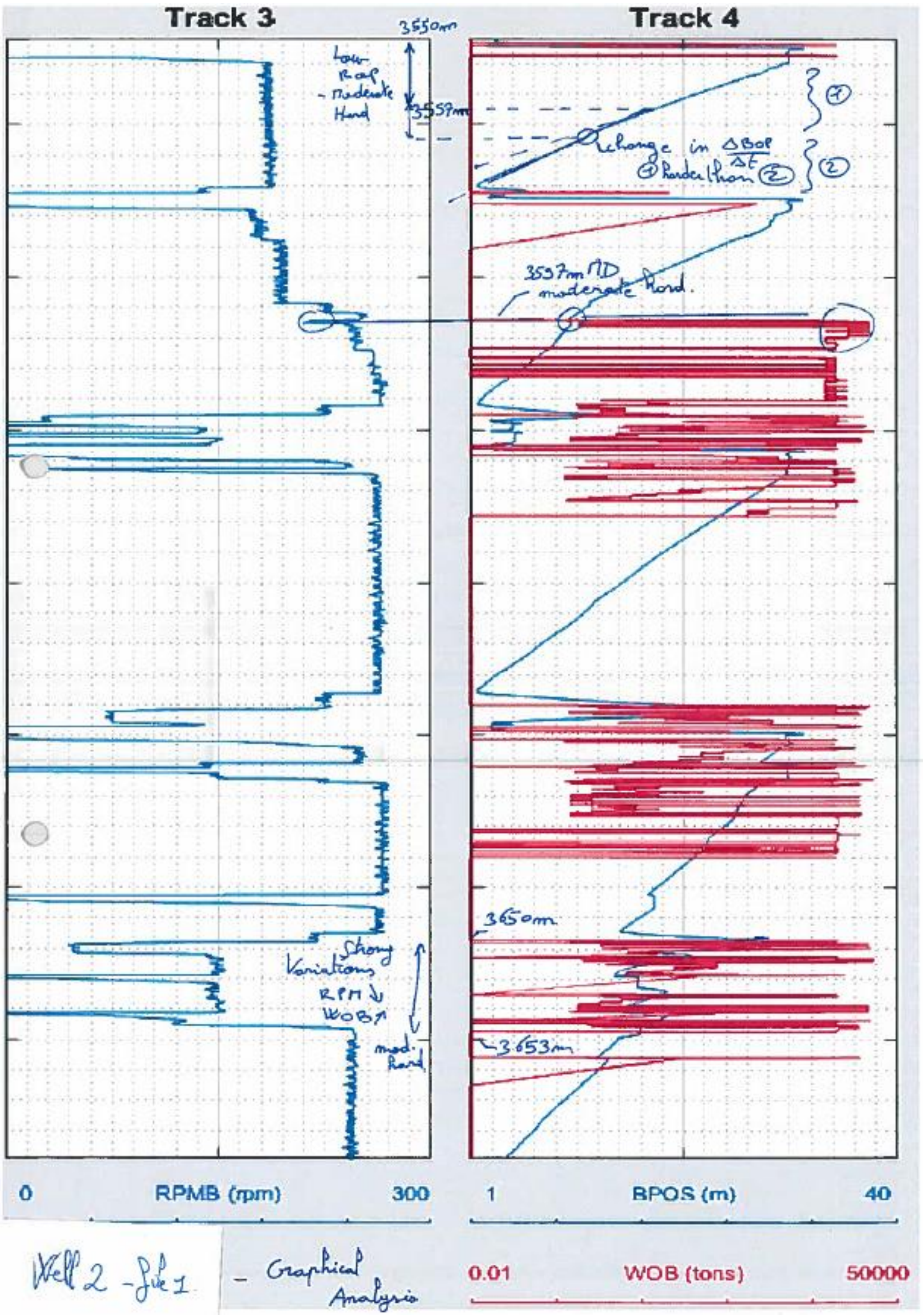
Formation	Wellbores																			
	Well 1		Well 2		Well 3		Well 4		Well 5		Well 6		Well 7		Well 8		Well 9		Well 10	
	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End	Top	End
Cromer	2526	2544	-	-	2478	2482	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Viking	2544	2546	2374	2379	2482	2490	2350	2385	2503	2514	2451	2460	2583	2610	2394	2396	2573	2594	2387	2388
Brent	2546	2611	2379	2588	2490	2625	2385	2529	2514	2585	2460	2478	2610	2617	2396	2536	2594	2750	2388	2565
Dunlin	2611	2768	2588	2615	2625	2715	-	-	2585	2081	2478	2606	2617	2693	2536	2731	-	-	2565	2600
Statfjord	2768	2941	-	-	-	-	-	-	2801	3007	2606	2829	2693	2909	-	-	-	-	-	-
Lunde	-	-	-	-	-	-	-	-	-	-	-	-	2909	2935	-	-	-	-	-	-

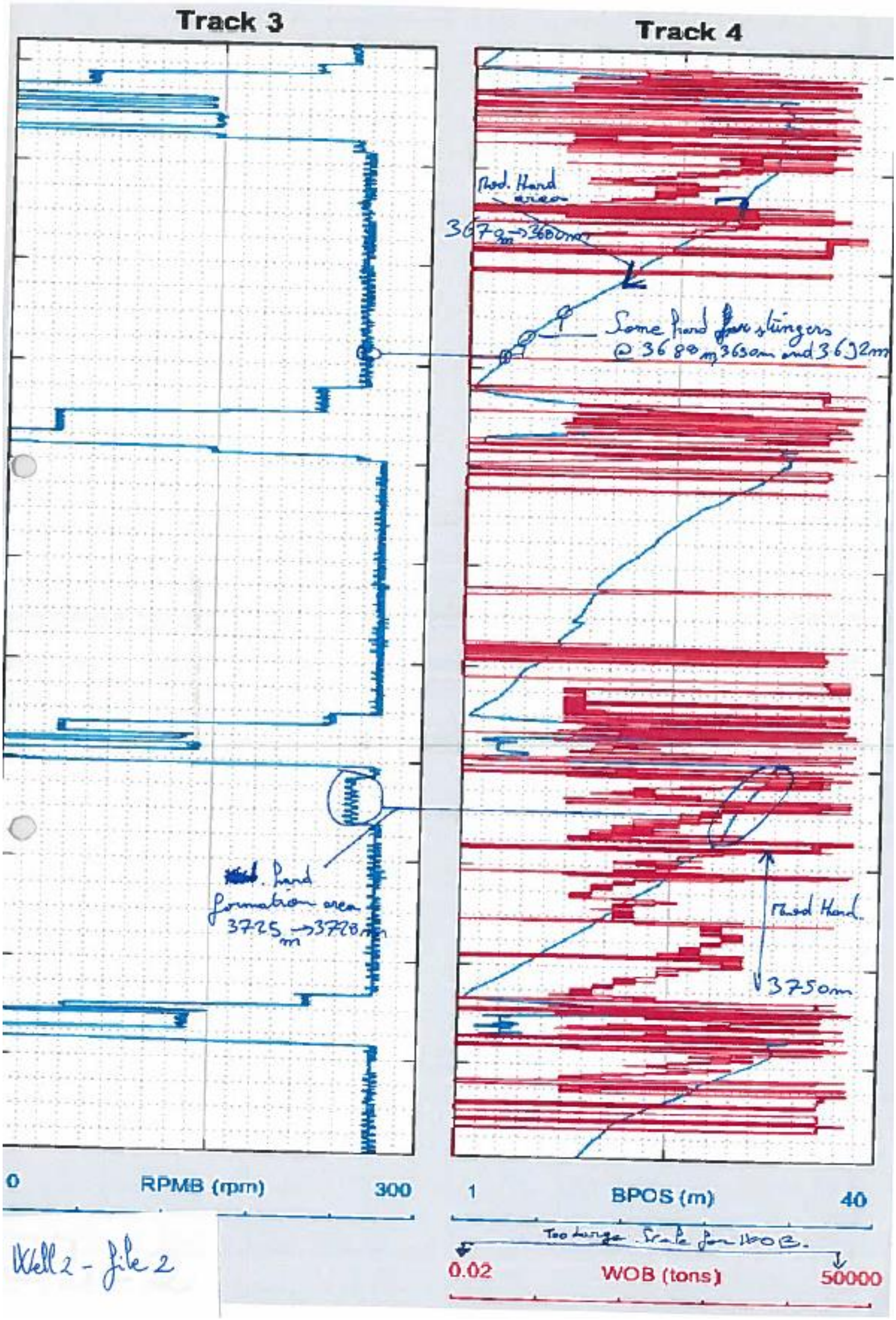
IN THIS TABLE ALL DEPTHS ARE TD DEPTH

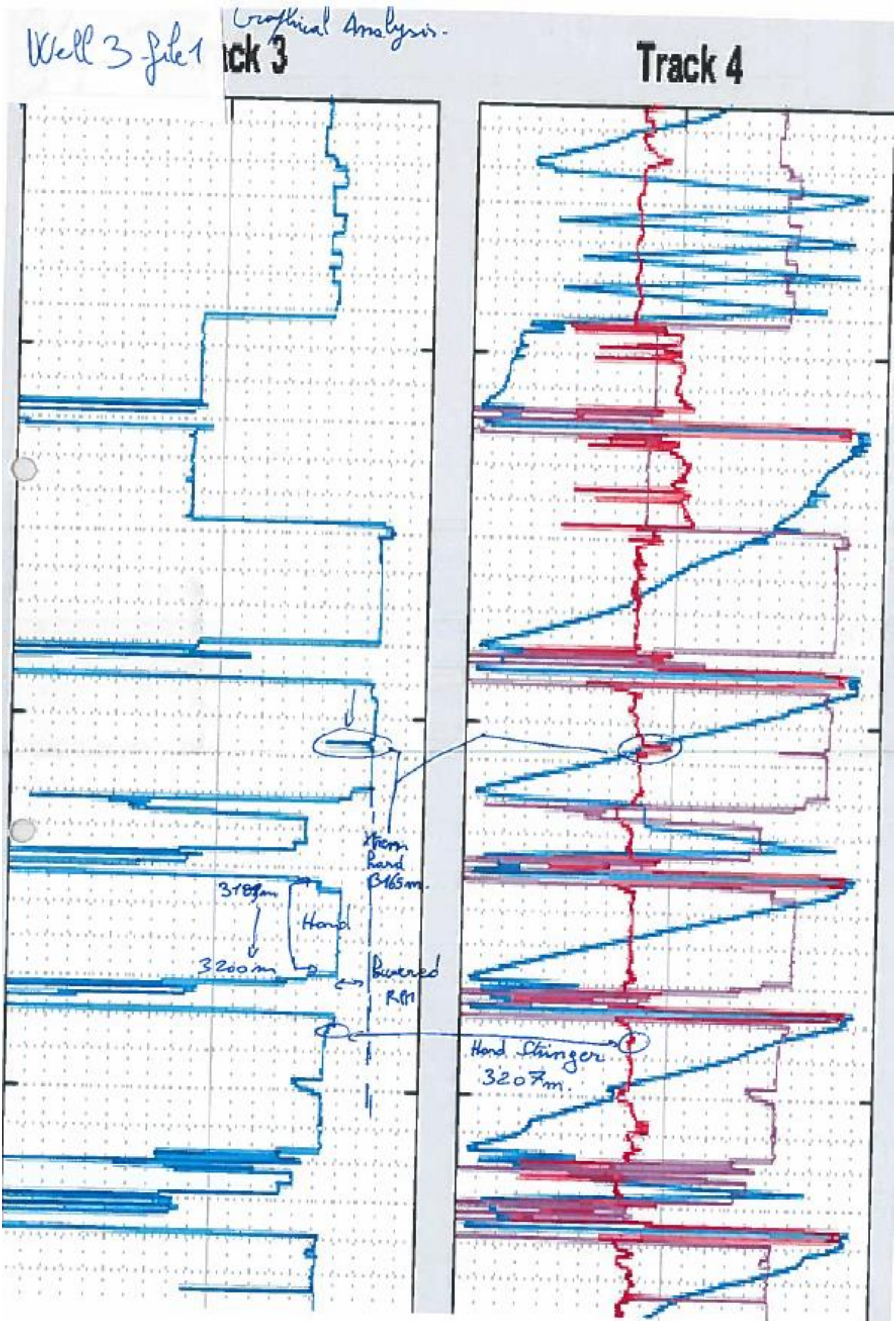
window studied

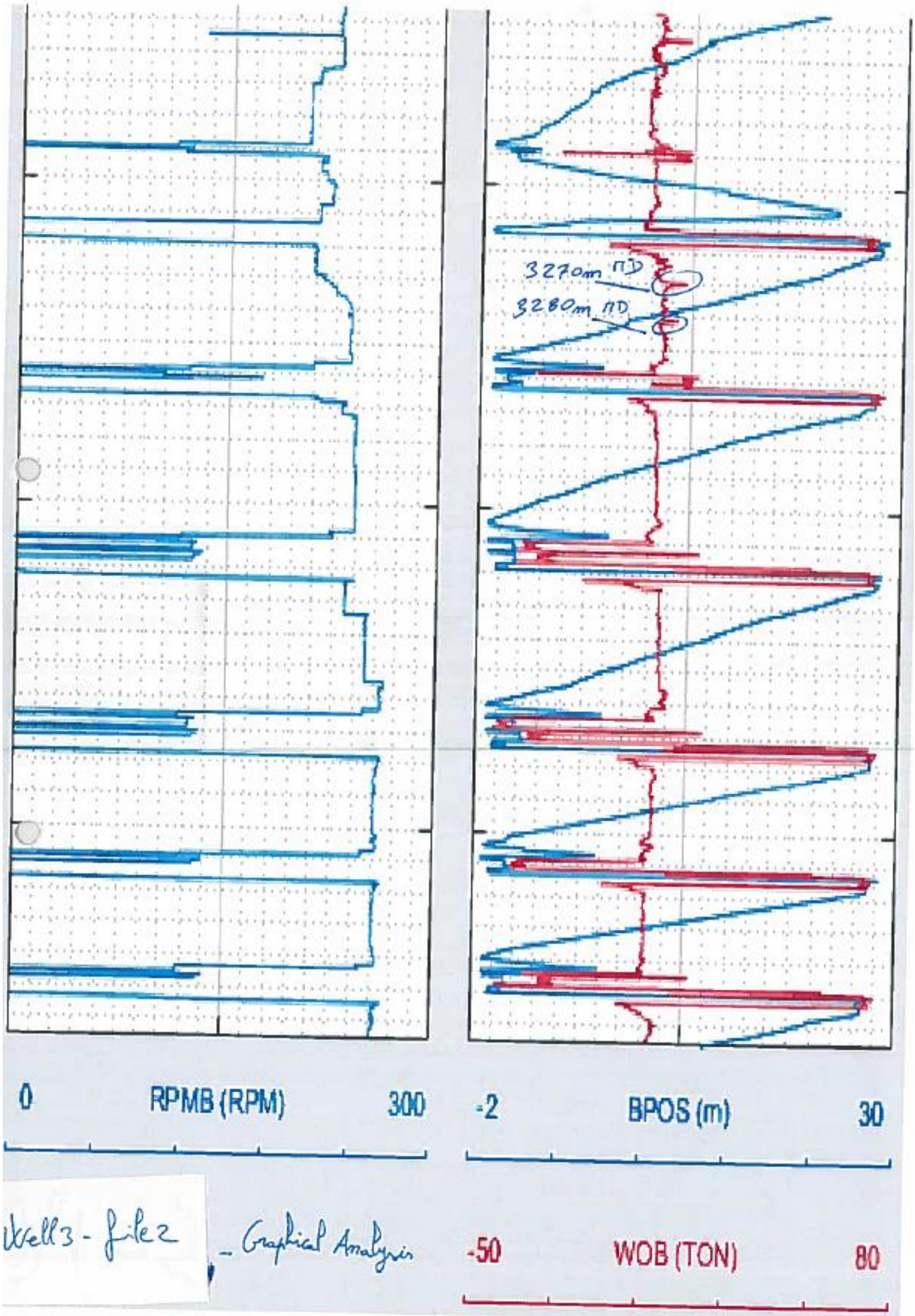
13.8 Appendix VIII – Graphical study of real time drilling data of the ten wells of the 10 wells

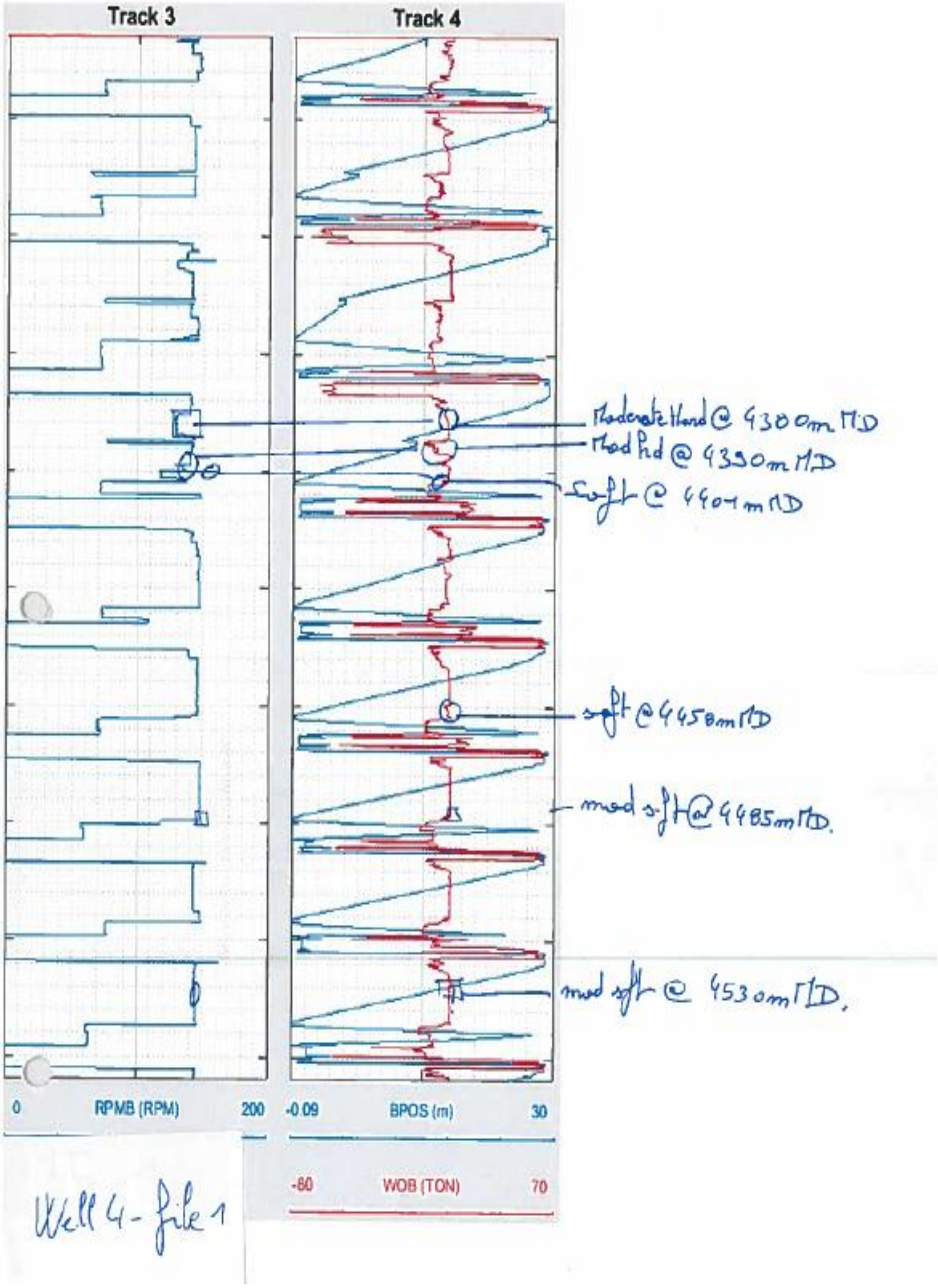


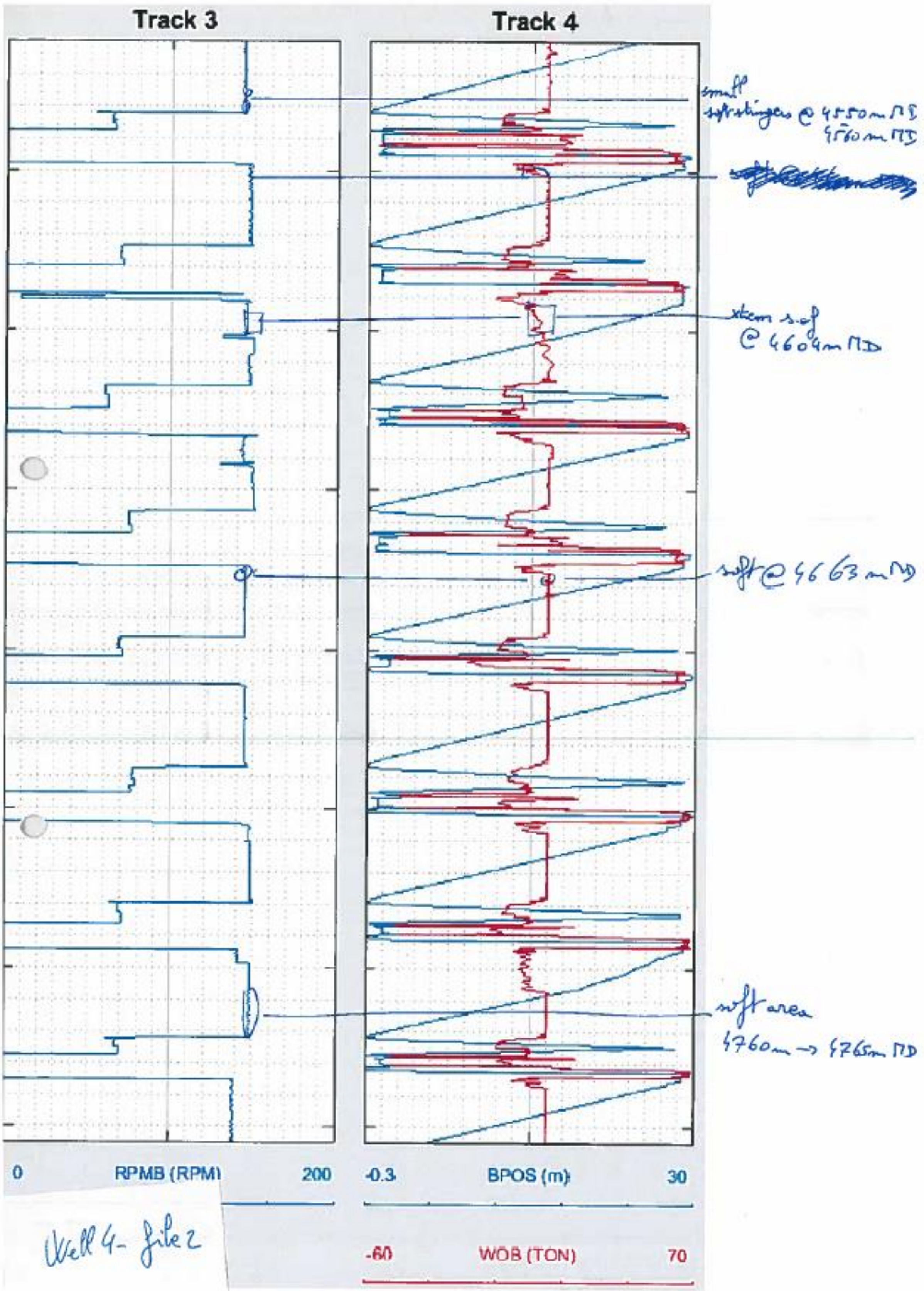

Track 3
Track 4
3550m
Low ROP
- moderate Hard
3559m
change in ΔBOP/Δt
② harder than ①
①
②
3537m MD
moderate Hard.
3650m
3653m
Strong Variations
RPM ↓
WOB ↗
mod. hard
0
RPMB (rpm)
300
1
BPOS (m)
40
0.01
WOB (tons)
50000
Well 2 - file 1
- Graphical Analysis

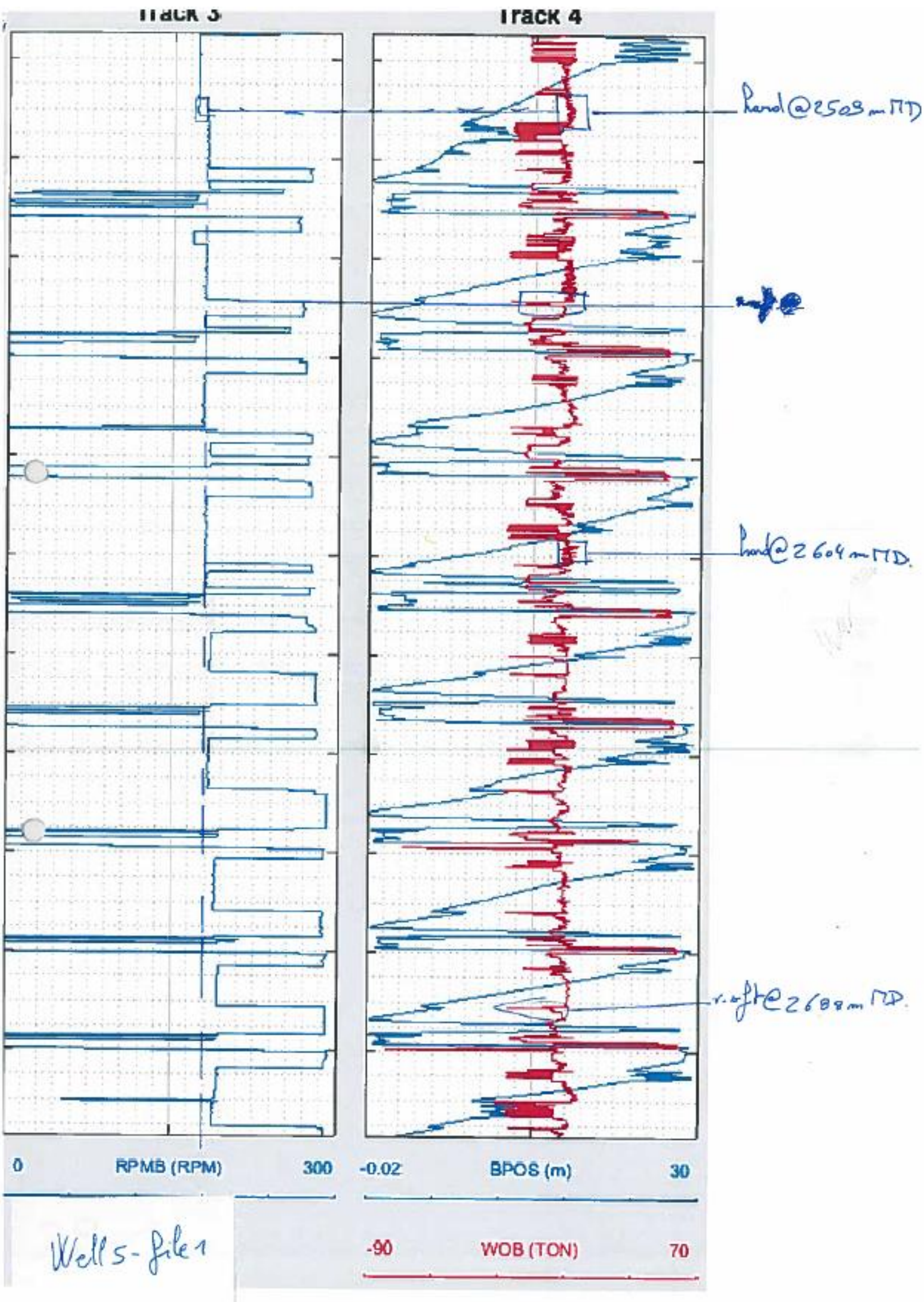

Track 3
Track 4
0
RPMB (rpm)
300
1
BPOS (m)
40
0.02
WOB (tons)
50000
3750m
Well 2 - file 2

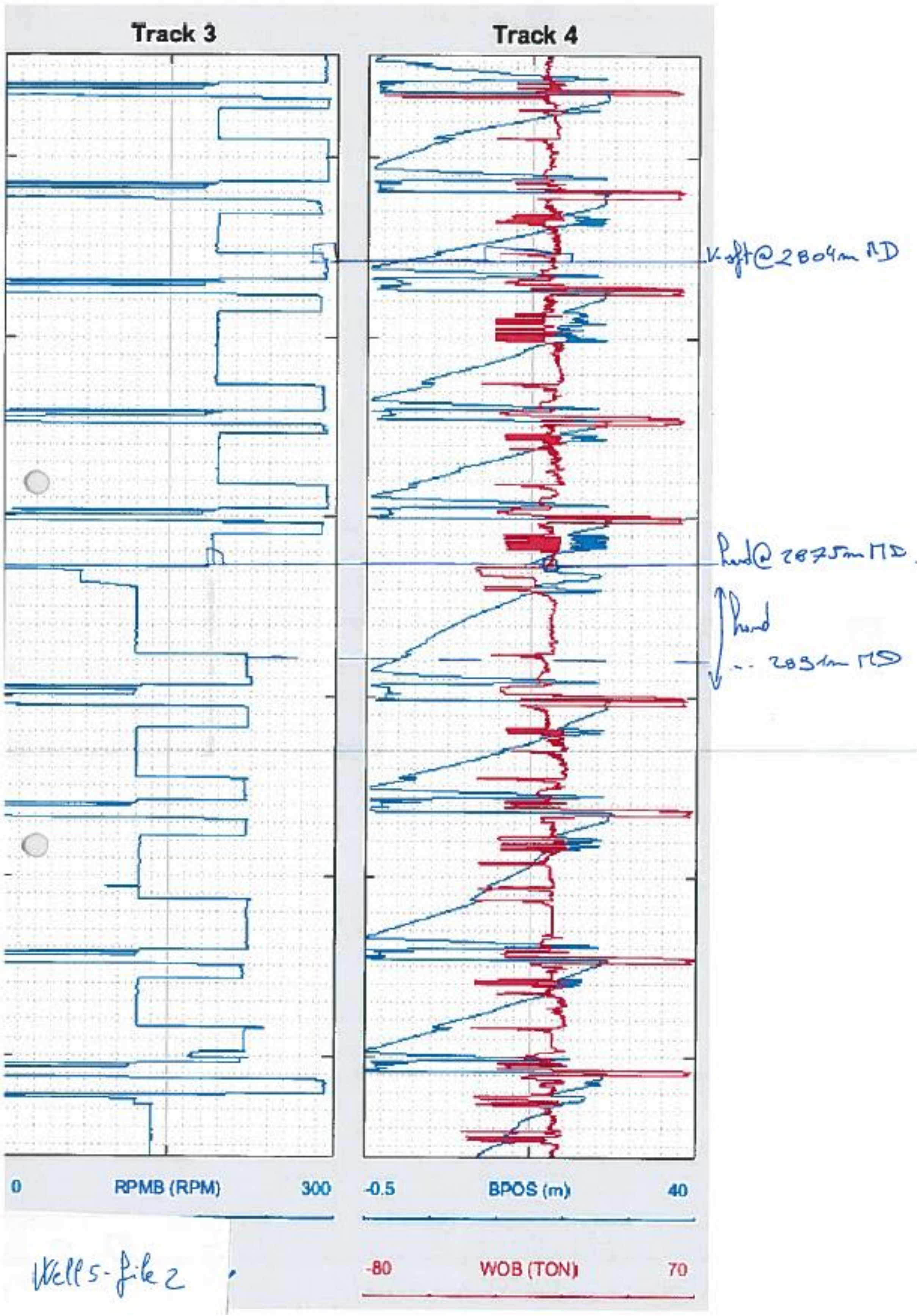

Well 3 -file 1
Graphical Analysis.
Track 4
Hard
3200 m
Hard Stinger
3207m.

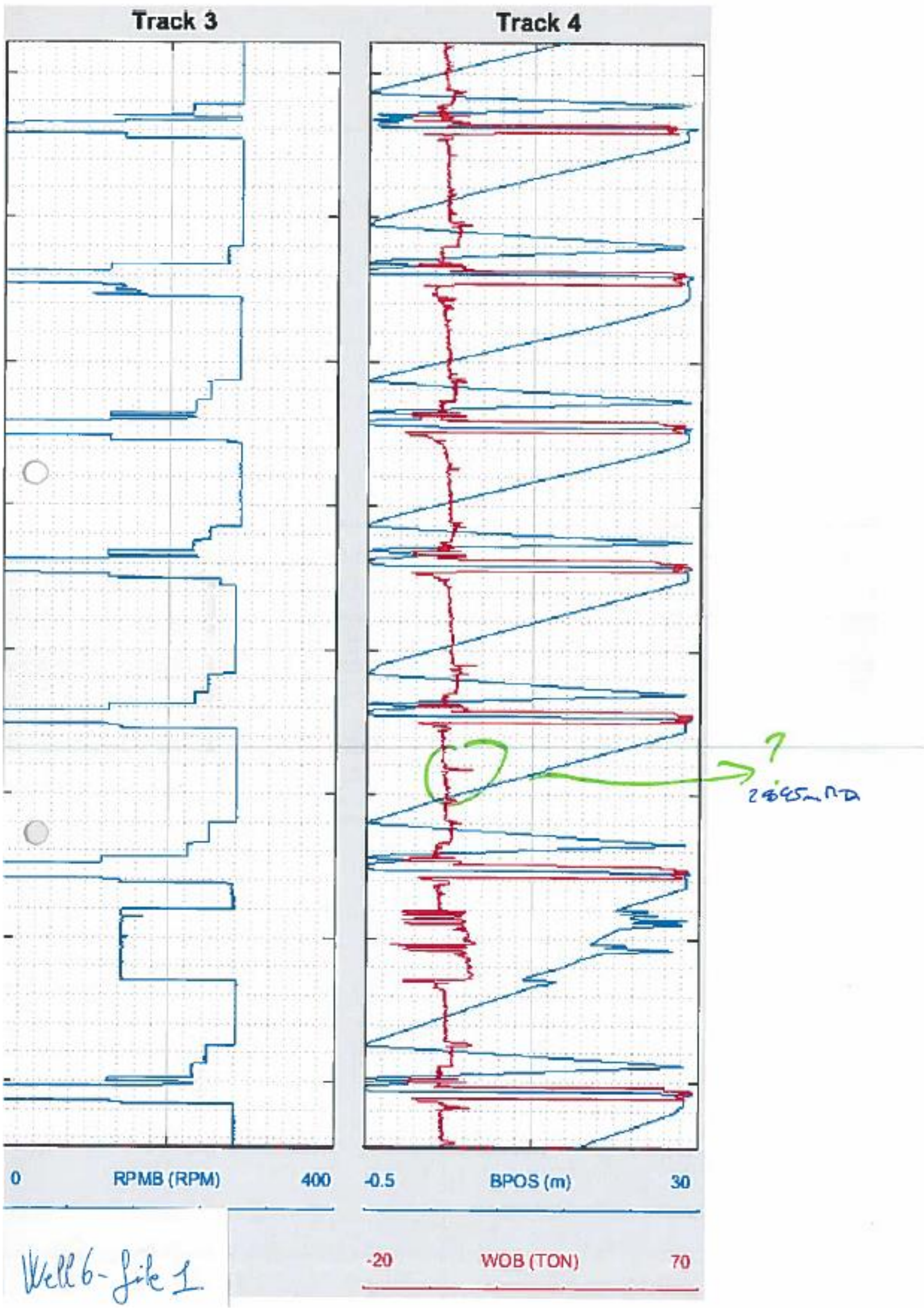

3270m MD
3280m MD
0
RPMB (RPM)
300
-2
BPOS (m)
30
-50
WOB (TON)
80
Well3 - file2
- Graphical Analysis

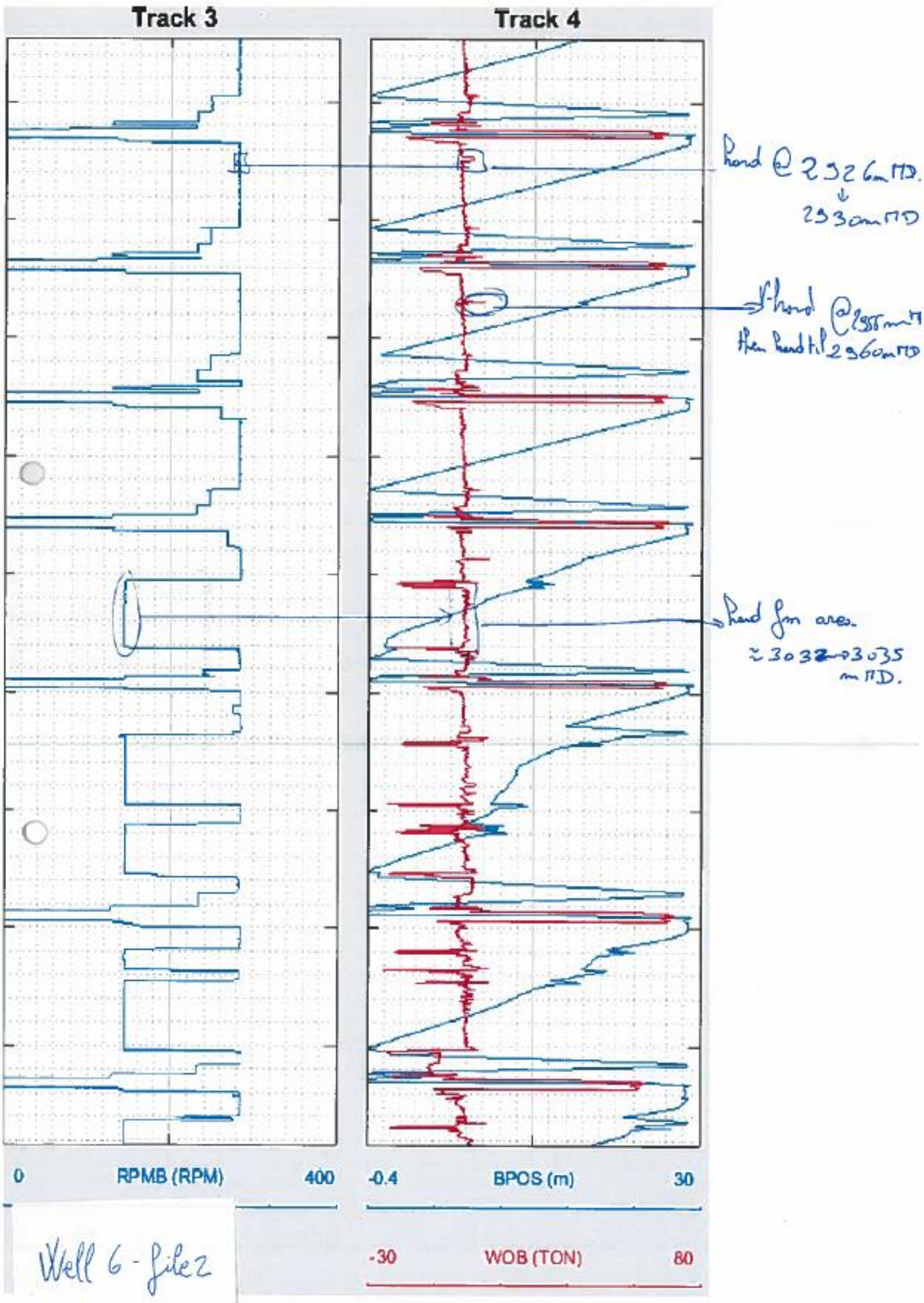

Track 3
Track 4
Moderate Hard @ 4300m MD
Mod Hd @ 4330m MD
Soft @ 4404m MD
soft @ 4458m MD
mod soft @ 4485m MD.
mod soft @ 4530m MD.
0
RPMB (RPM)
200
-0.09
BPOS (m)
30
-80
WOB (TON)
70
Well 4 - file 1

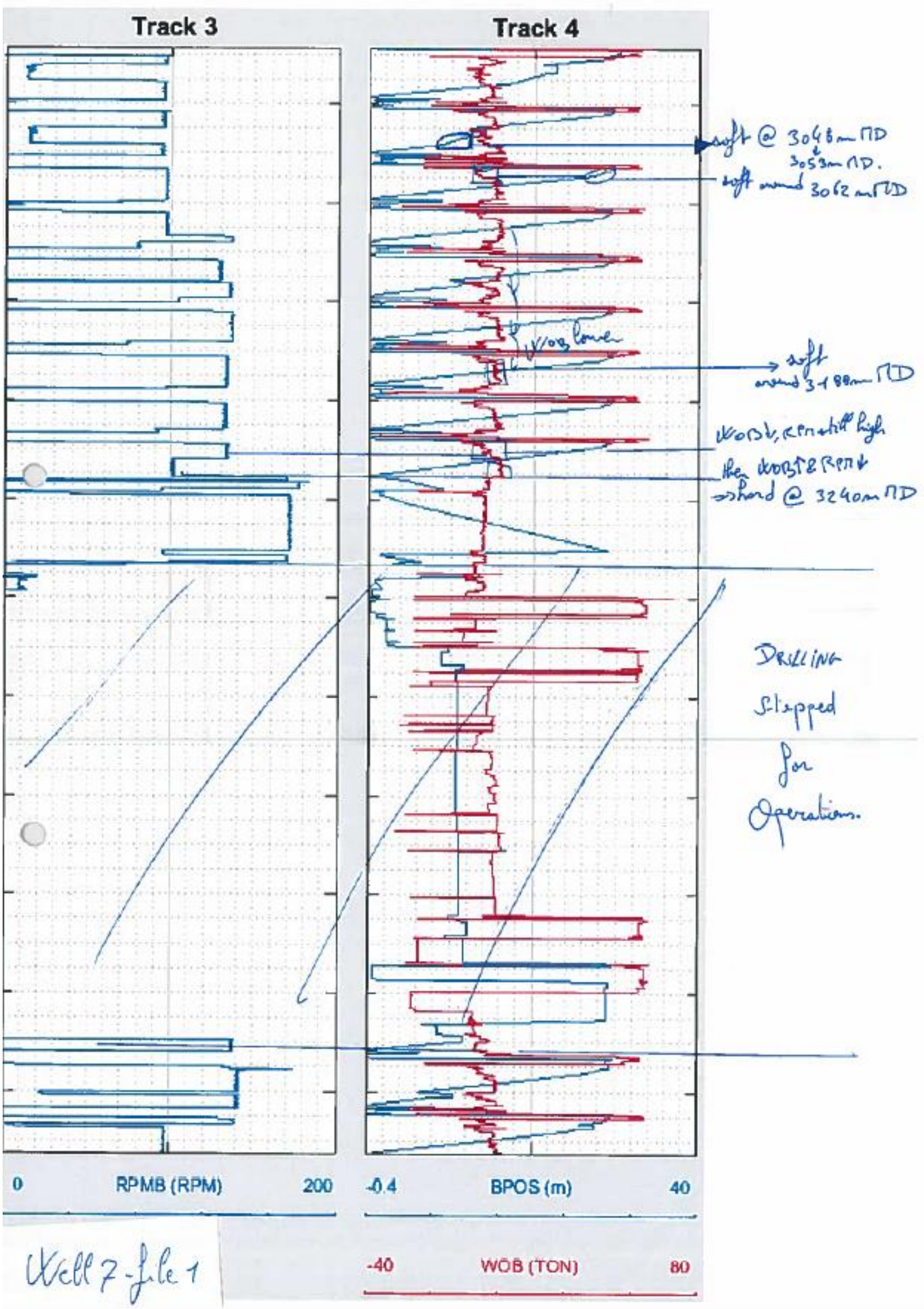

Track 3
Track 4
small
softstingers @ 4550m MD
4560m MD
stem s.of
@ 4604m MD
soft @ 4663m MD
soft area
4760m → 4765m MD
0
RPMB (RPM)
200
-0.3
BPOS (m)
30
-60
WOB (TON)
70
Well 4- file 2

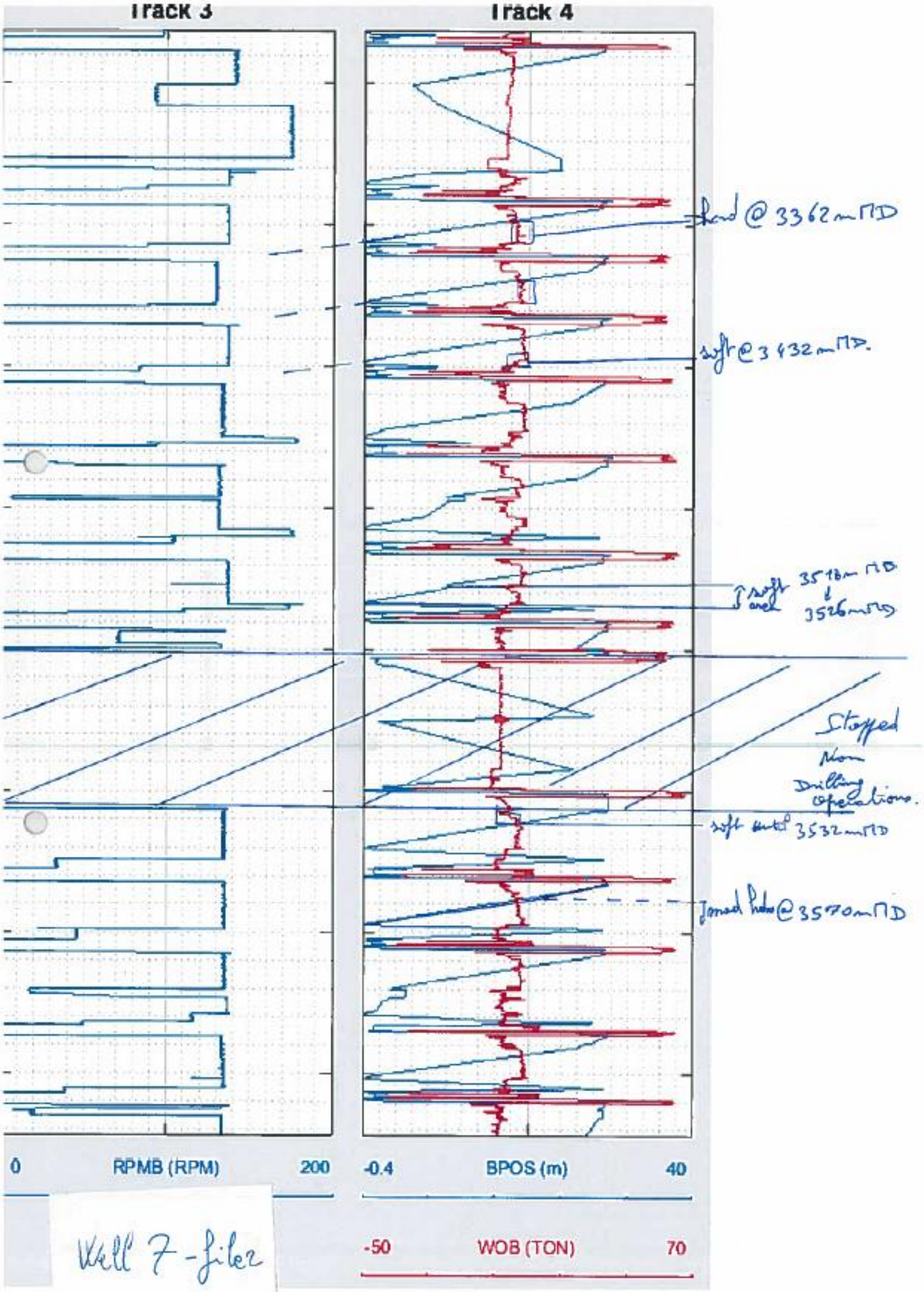

Track 3
Track 4
Rand@2503 m MD
Rand@2604 m MD.
r.aft@2688 m MD.
0
RPMB (RPM)
300
-0.02
BPOS (m)
30
Well 5 - file 1
-90
WOB (TON)
70


Track 3
Track 4
0
RPMB (RPM)
300
-0.5
BPOS (m)
40
-80
WOB (TON)
70
Well 5 - file 2

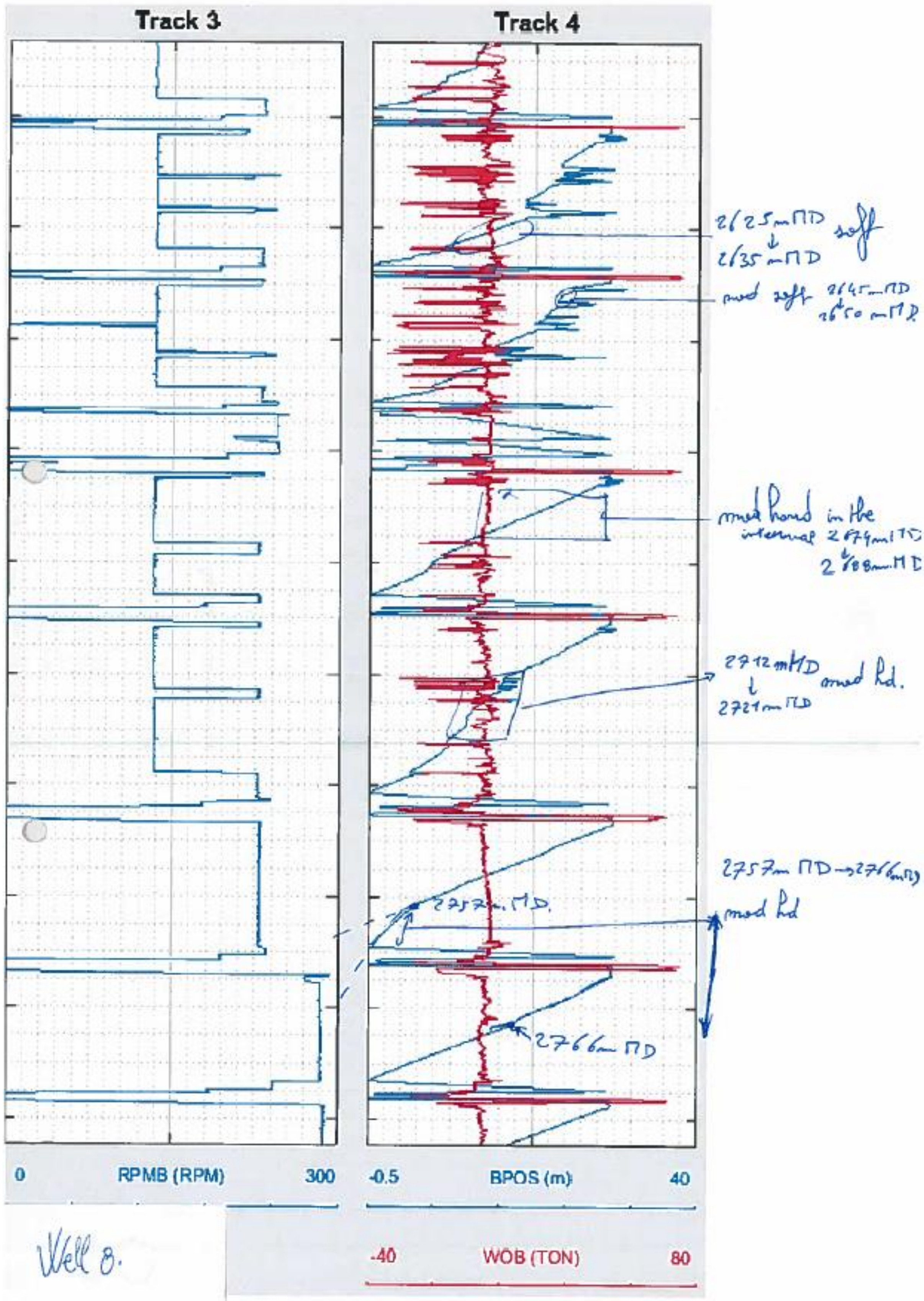

Track 3
Track 4
0
RPMB (RPM)
400
-0.5
BPOS (m)
30
-20
WOB (TON)
70
Well 6- file 1

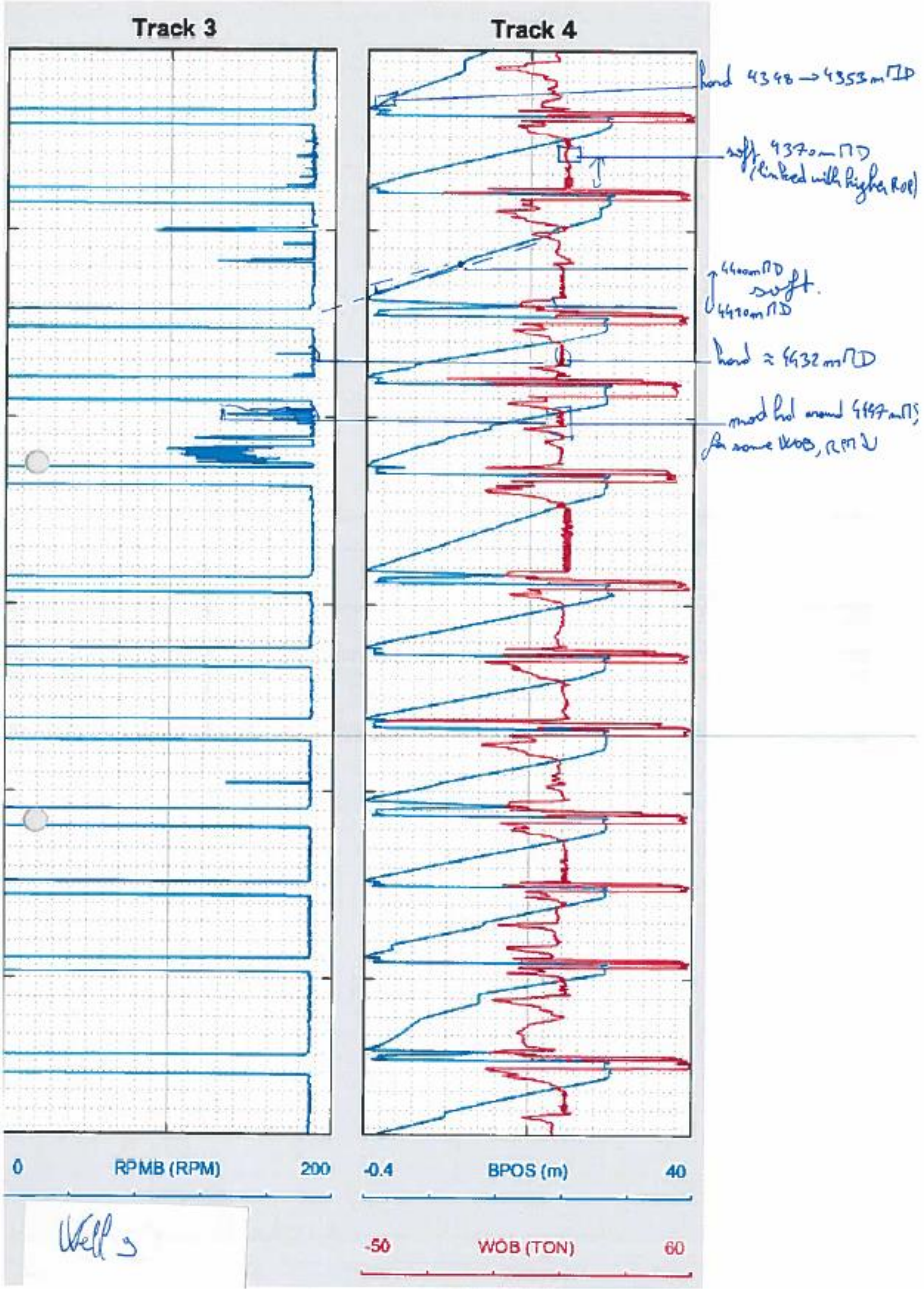

Track 3
Track 4
0
RPMB (RPM)
400
-0.4
BPOS (m)
30
-30
WOB (TON)
80
Well 6 - file2
Hard @ 2926m MD.
2930m MD
Hard fm area.
≈3032→3035
m MD.


Track 3
Track 4
0
RPMB (RPM)
200
-0.4
BPOS (m)
40
-40
WOB (TON)
80
soft @ 3048m MD
3053m MD.
soft around 3062m MD
WOB lower
soft around 3188m MD
WOB, RPM still high
then WOB & RPM
hard @ 3240m MD
Drilling stopped for Operations.
Well 7-file 1

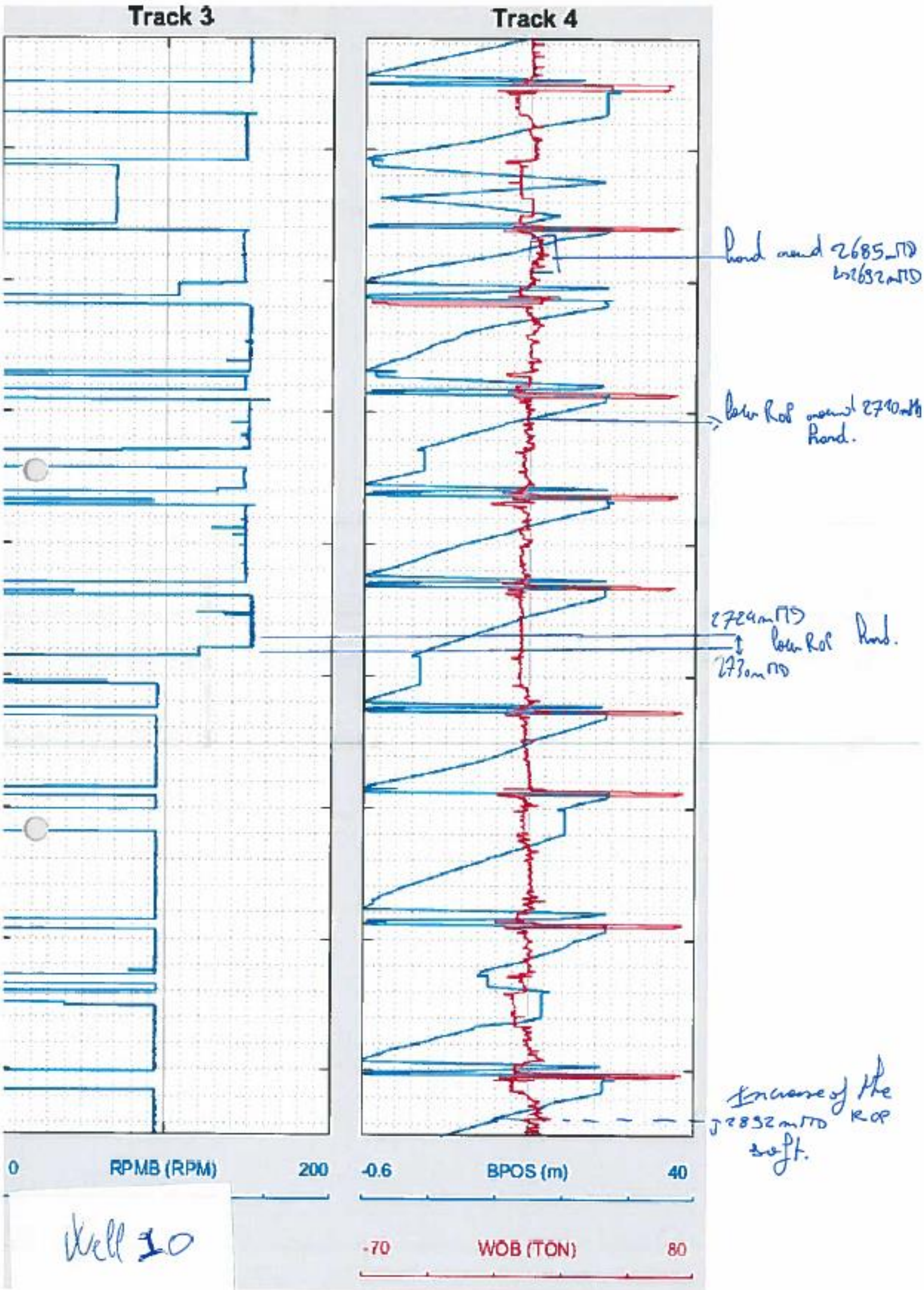


Track 3
Track 4
hard @ 3362 m MD
soft @ 3432 m MD.
soft and 3518 m MD & 3526 m MD
Stopped Non Drilling Operations.
soft until 3532 m MD
0
RPMB (RPM)
200
-0.4
BPOS (m)
40
-50
WOB (TON)
70
Well 7 - filer


Track 3
Track 4
0
RPMB (RPM)
300
-0.5
BPOS (m)
40
-40
WOB (TON)
80
Well 8.

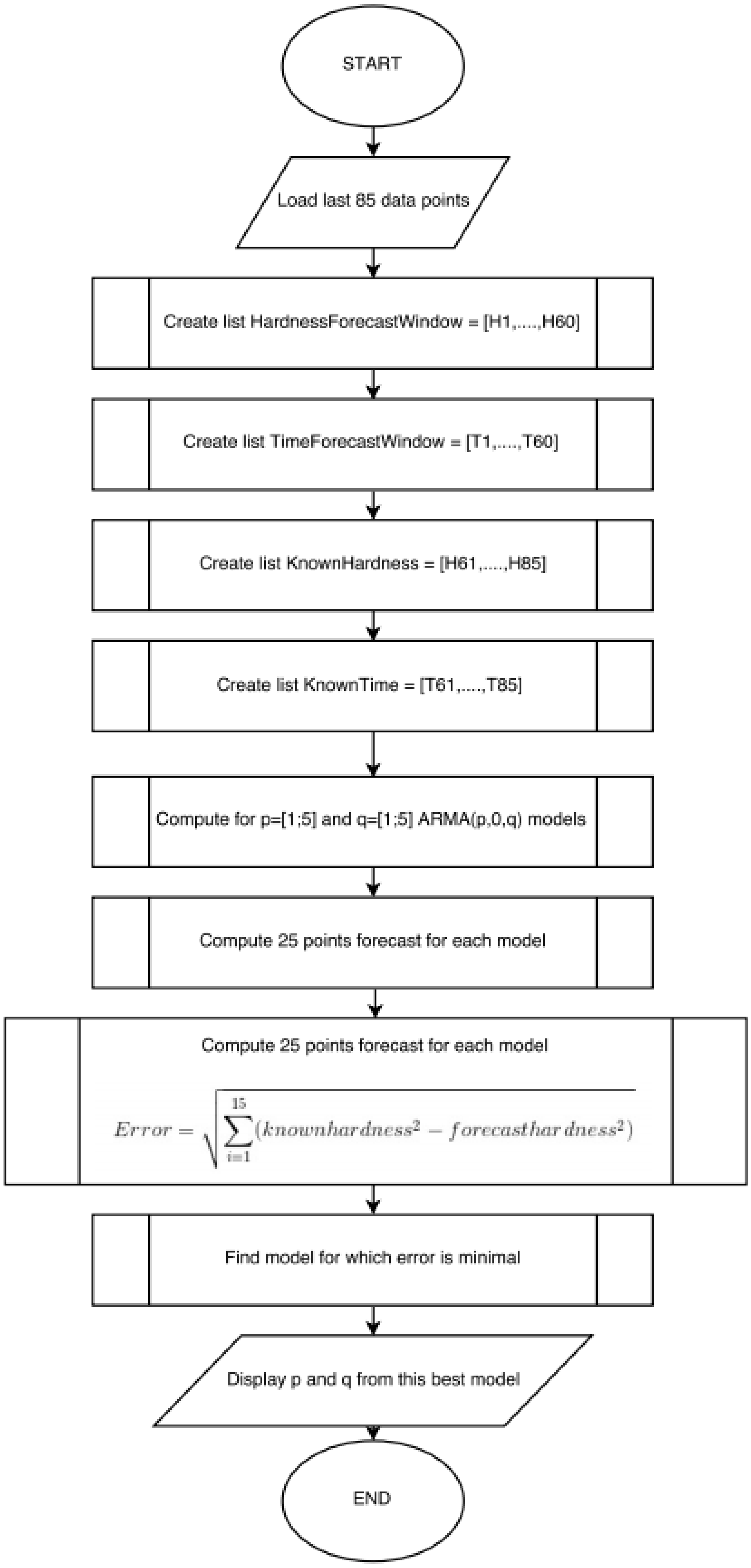


Track 3
Track 4
hard 4348 → 4353 m MD
soft 4370 m MD
(linked with higher ROP)
4400 m MD
soft.
4410 m MD
hard ≈ 4432 m MD
0
RPMB (RPM)
200
-0.4
BPOS (m)
40
-50
WOB (TON)
60
Well 3

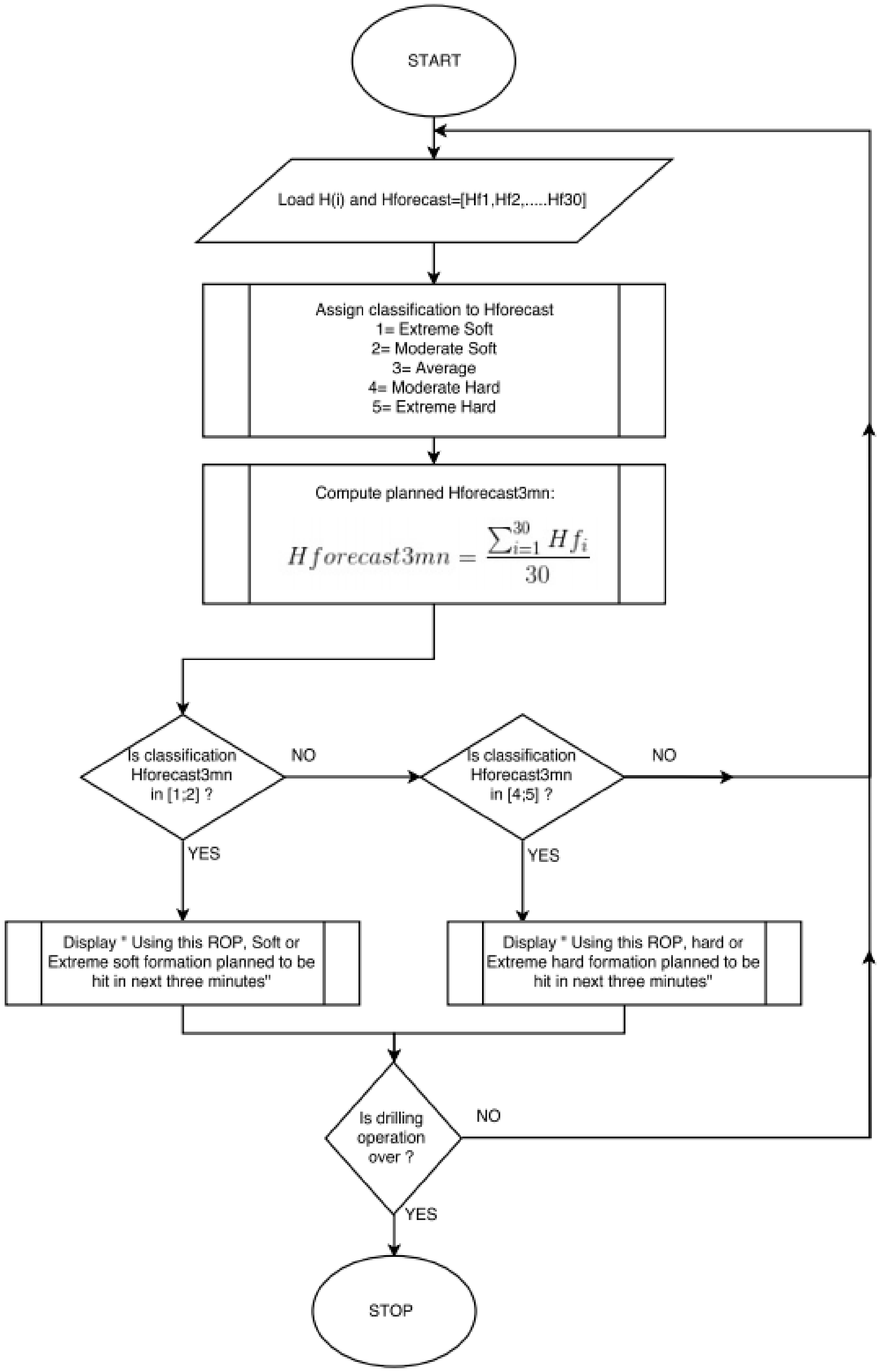

Track 3
Track 4
Hard around 2685 mTD
to 2692 mTD
Lower ROP around 2710 mMD
Hard.
2724 m MD
Lower ROP Hard.
2730 m MD
Increase of the ROP
2892 m MD
soft.
0
RPMB (RPM)
200
-0.6
BPOS (m)
40
-70
WOB (TON)
80
Well 10

13.9 Appendix XIX – Flow charts of agent sub functions

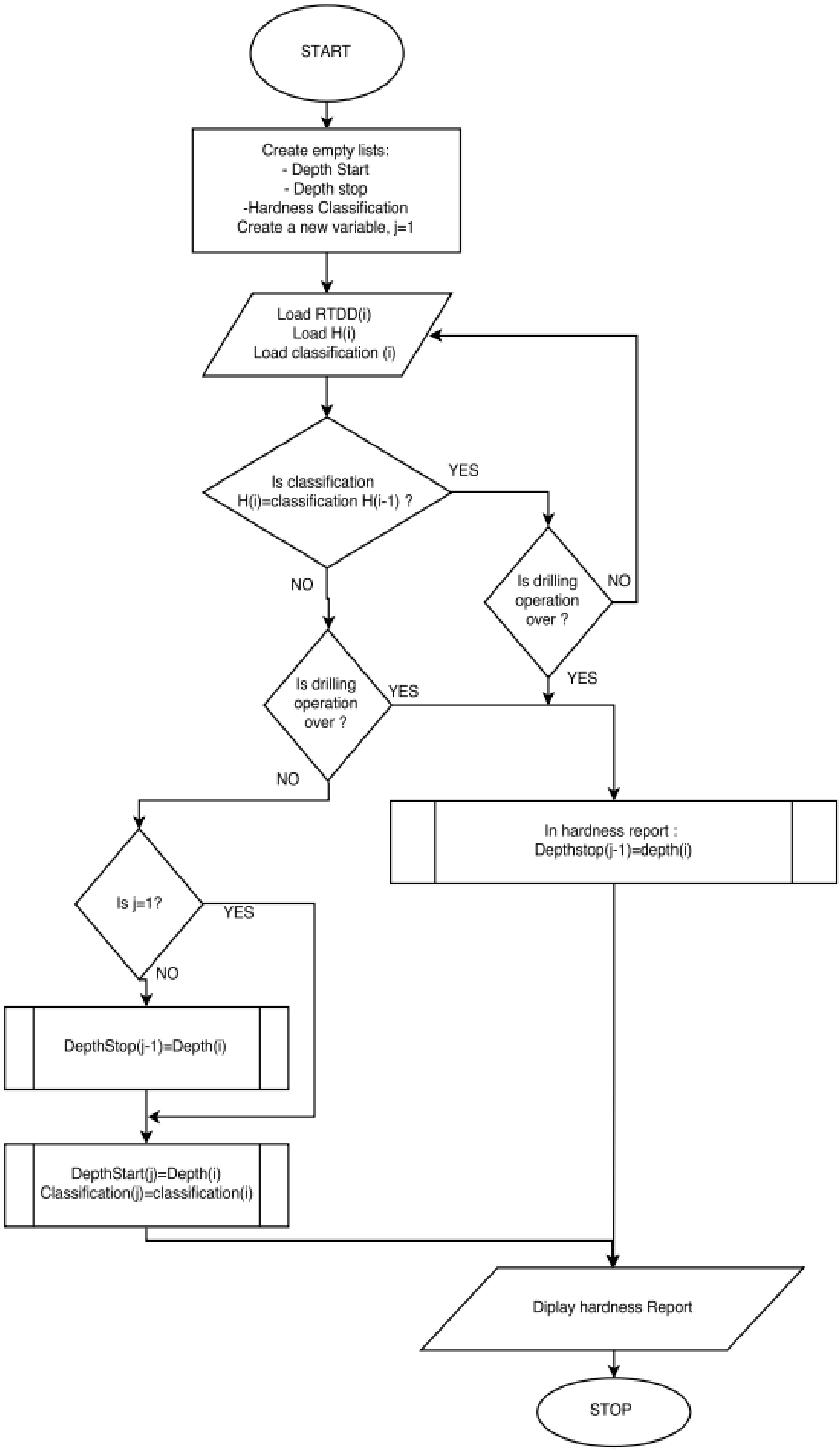
13.9.1 Flow chart of the "launch ARMA test" function



13.9.2 Flow chart of the "launch warning message" function



13.9.3 Flow chart of the “launch hardness review” function



13.10 Appendix X – Matlab code of the agent

13.10.1 Main code

```
% Main Matlab code of agent
%Task 1: -forecast and detection of change in formation hardness agent
%Task 2: -Provide a hardness classification along the well path
%for further wells in the area
%Coded by C.Donne, 2017, Master student, NTNU

%delete previous operation made in Matlab
clear
clc
%load the workspace in this example the file name is well1.mat
load well1.mat
%WARNING : The program only works if the first column of RTDD curves
%are TIME,DBTM,DMEA,ROP,WOB, RPM in the same order!
%Be sure to have the appropriate file .mat before launch
%Test if the file have the proper shape
test1=strcmp(RTDD.curve_info(1,1),'Time');
test2=strcmp(RTDD.curve_info(2,1),'DBTM');
test3=strcmp(RTDD.curve_info(3,1),'DMEA');
test4=strcmp(RTDD.curve_info(4,1),'ROP');
test5=strcmp(RTDD.curve_info(5,1),'WOB');
test6=strcmp(RTDD.curve_info(6,1),'RPM');
Test= [test1 test2 test3 test4 test5 test6];
if Test==[1 1 1 1 1 1]
    disp(['the file uploaded is on the good format']);
else disp(['the file uploaded is not on the good format']);
end
clear test1 test2 test3 test4 test5 test6 Test
%Assign ROP,RPM,WOB,DMEA,DBTM,TIME into separated lists
LengthRTDD = structfun(@(field) length(field),RTDD);
NumberData=LengthRTDD(length(LengthRTDD));
Time1=[RTDD.curves(1:NumberData)];
DBTM1=[RTDD.curves(NumberData+1:2*NumberData)];
DMEA1=[RTDD.curves(2*NumberData+1:3*NumberData)];
ROP1=[RTDD.curves(3*NumberData+1:4*NumberData)];
WOB1=[RTDD.curves(4*NumberData+1:5*NumberData)];
RPM1=[RTDD.curves(5*NumberData+1:6*NumberData)];
clear RTDD LengthRTDD
%Remove data from other operations than drilling creating clean lists
TimeFullPath=[];DBTMFullPath=[];DMEAFullPath=[];ROPFullPath=[];
WOBFullPath=[];RPMFullPath=[];
for i = 1:NumberData
    if DBTM1(i)==DMEA1(i)
        TimeFullPath=[TimeFullPath Time1(i)];
        DBTMFullPath=[DBTMFullPath DBTM1(i)];
        DMEAFullPath=[DMEAFullPath DMEA1(i)];
        ROPFullPath=[ROPFullPath ROP1(i)];
        WOBFullPath=[WOBFullPath WOB1(i)];
        RPMFullPath=[RPMFullPath RPM1(i)];
    end
end
clear Time1 DBTM1 DMEA1 ROP1 WOB1 RPM1
%We want to simulate a use of the agent during drilling, so we have to
%rewrite the workspace at every new point, just as we were receiving new
%data every 5 seconds
%Create Bourgoyne and Young coefficients
a5_xtremsoft=0.5;
a6_xtremsoft=0.4;
```

```
a5_midsoft=0.8;
a6_midsoft=0.6;
a5_average=1.1;
a6_average=0.65;
a5_midhard=1.5;
a6_midhard=0.75;
a5_xtremhard=1.9;
a6_xtremhard=0.9;
%we start providing a list of first 90 points for the agent
Time=[];DBTM=[];DMEA=[];ROP=[];WOB=[];RPM=[];
for i = 1:90
        Time=[Time TimeFullPath(i)];
        DBTM=[DBTM DBTMFullPath(i)];
        DMEA=[DMEA DMEAFullPath(i)];
        ROP=[ROP ROPFullPath(i)];
        WOB=[WOB WOBFullPath(i)];
        RPM=[RPM RPMFullPath(i)];
end
%Compute number of loops needed to have results for all the path
Nloops=length(TimeFullPath)-90;
%and we create the algorithm of the agent presented Figure 6-1 in Nloops
%iteration to simulate the full drilling operation
%variable needed for video caption
kk=1;
%now simulation of RTDD begin with a new input at every loop
for k=90:(Nloops+85)
%we start computing hardness along all data available a5 and a6 average
%Compute hardness along the path
HardnessFromData=[];
DepthHardness=[];
TimeHardness=[];
for i=1:length(Time)
    hardness=abs(((WOB(i)^a5_average)*(RPM(i)^a6_average))/(ROP(i)));
    if hardness < Inf
    HardnessFromData=[HardnessFromData hardness];
    DepthHardness=[DepthHardness DMEA(i)];
    TimeHardness=[TimeHardness Time(i)];
    end
end
%Creation of the empty classification list
ClassificationHardness=zeros(length(HardnessFromData),1);
for i=1:length(HardnessFromData)
    ClassificationHardness(i)=3;
end
%we find median hardness
Hmed=median(HardnessFromData);
%we recompute hardness according classification presented table 3-6
NN=length(HardnessFromData);
for i=1:NN
    if HardnessFromData(i) < Hmed/100
        HardnessFromData(i)=...
            abs(((WOB(i)^a5_xtremsoft)*(RPM(i)^a6_xtremsoft))/(ROP(i)));
        ClassificationHardness(i)=1;
    elseif HardnessFromData(i) <= Hmed/10 && HardnessFromData(i )> Hmed/100
        HardnessFromData(i)=...
            abs(((WOB(i)^a5_midsoft)*(RPM(i)^a6_midsoft))/(ROP(i)));
        ClassificationHardness(i)=2;
    elseif HardnessFromData(i) > Hmed*10 && HardnessFromData(i )<= Hmed*100
        HardnessFromData(i)=...
            abs(((WOB(i)^a5_midhard)*(RPM(i)^a6_midhard))/(ROP(i)));
        ClassificationHardness(i)=4;
```

```
    elseif HardnessFromData(i) > Hmed*100
        HardnessFromData(i)=...
            abs(((WOB(i)^a5_xtremhard)*(RPM(i)^a6_xtremhard))/(ROP(i)));
        ClassificationHardness(i)=5;
    end
end
%Then we launch ARMA(p,q) test to determine coefficients p,q needed for
%forecast - see code chapter 13.10.2
ARMATest;
%Then we compute the prediction on the next 3mn, 30 next data points
prevision=30;
N=length(HardnessFromData);
%We call datawindow the last 60 data used for the forecast
datawindow=60;
HardnessforForecast=zeros(datawindow,1);
TimeforForecast=zeros(datawindow,1);
for i=1:datawindow
    HardnessforForecast(i)=HardnessFromData(N-datawindow+i);
    TimeforForecast(i)=TimeHardness(N-datawindow+i);
end
y=HardnessforForecast;
x=TimeforForecast;
n=length(x);
%p,q come from TEST ARIMA done before
try
Model=arima(p,0,q);
Fit=estimate(Model,y);
[HardnessForecast,ymse] = forecast(Fit,prevision,'Y0',y);
end
TimeForecast=zeros(prevision,1);
for j=1:prevision
    TimeForecast(j)=x(n)+j*5;
end
%plot hardness known and then the prediction
fig=figure(1)
semilogy(x,y,'Color',[.7,.7,.7]);
hold on
semilogy(TimeForecast,HardnessForecast,'k','LineWidth',2);
legend('Hardness from data','Hardness forecasted using ARMA','Location',...
    'southoutside')
title('Hardness Forecast using ARMA methodology - Well1')
hold off
%we capture the plot to get a video at the end to simulate the evolution of
%the forecast while receiving a new RTT at every loop
F(kk)=getframe(fig);
kk=kk+1;
%Launch of the warning message procedure according prediction - see the
%code chapter 13.10.3
WarningMessage;
%Acquisition of the new data as the driling operation is not over
Time=[Time TimeFullPath(k)];
DBTM=[DBTM DBTMFullPath(k)];
DMEA=[DMEA DMEAFullPath(k)];
ROP=[ROP ROPFullPath(k)];
WOB=[WOB WOBFullPath(k)];
RPM=[RPM RPMFullPath(k)];
end
%Implementation of the report providing hardness along path at the end of
%drilling operation - see code chapter 13.10.4
HardnessReport;
%Video forecasting evolution versus time, while receiving new RTDD at each
```

```
%iteration
v = VideoWriter('Well1.avi');
v.FrameRate=2;
open(v)
writeVideo(v,F)
close(v)
%Finally give back the report providing hardness classification versus
%depth along all the well path.
Report
```

13.10.2 Code of the sub-function "ARMATest"

```
%ARMA test needed to determine p,q for next prediction
%We use data available so far : HardnessFromData
N=length(HardnessFromData);
%Prevision length
prevision=30;
datawindow=60;
Hardnessforecastwindow=zeros(datawindow,1);
Timeforecastwindow=zeros(datawindow,1);
TrueHardness=zeros(prevision,1);
TrueTime=zeros(prevision,1);
%We use the last 90 points we know.
%from 0 to 60 it is points we use to compute the forecast
%60 to 90 is our forecast, but we also now them
%So we are going to compare forecast and real to find model suits best
%it is the one that minimize forecast and real data
for i=1:datawindow
    Hardnessforecastwindow(i)=HardnessFromData(N-prevision-datawindow+i);
    Timeforecastwindow(i)=Time(N-prevision-datawindow+i);
end
for i=1:prevision
    TrueHardness(i)=HardnessFromData(N-prevision+i);
    TrueTime(i)=Time(N-prevision+i);
end
%We test from p=1 to p=5 and q=1 to q=5
MatrixDeterminepq=zeros(5,5);
for p=1:5
    for q=1:5
        try
y=Hardnessforecastwindow;
x=Timeforecastwindow;
n=length(x);
Model=arima(p,0,q);
Fit=estimate(Model,y);
[yy,ymse] = forecast(Fit,prevision,'Y0',y);
%we compute the error between the prediciton and real case
Error=zeros(prevision,1);
for i=1:prevision
    Error(i)=sqrt(abs((yy(i)).^2-(TrueHardness(i)).^2));
end
MatrixDeterminepq(p,q)=mean(Error);
        catch
            MatrixDeterminepq(p,q)=Inf;
        end
    end
end
for i=1:5
    for j=1:5
        if MatrixDeterminepq(i,j)== Inf
```

```
            MatrixDeterminepq(i,j)=999999999999999;
        end
    end
end
%we find the best model with the smallest error
[p,q] = find(MatrixDeterminepq == min(abs(MatrixDeterminepq(:))));
p=min(p);
q=min(q);
```

13.10.3 Code of the sub function “WarningMessage”

```
%Warning message program
%The Main program provide the classification of the known hardness
%Here we are going to assign hardness classification for the prediction
%aswell
ClassificationHardnessForecast=zeros(length(HardnessForecast),1);
for i=1:length(HardnessForecast)
    ClassificationHardness(i)=3;
end
NNN=length(HardnessForecast);
for i=1:NNN
    if HardnessForecast(i) < Hmed/100
        HardnessForecast(i)=1;
    elseif HardnessForecast(i) <= Hmed/10 && HardnessFromData(i )> Hmed/100
        HardnessForecast(i)=2;
    elseif HardnessForecast(i) > Hmed*10 && HardnessFromData(i )<= Hmed*100
        HardnessForecast(i)=4;
    elseif HardnessForecast(i) > Hmed*100
        HardnessForecast(i)=5;
    end
end
%Then we provide an estimation of the coming hardness in next 3minutes
Hforecast3mn=sum(HardnessForecast)/NNN;
if Hforecast3mn >=1 && Hforecast3mn <= 2
    display 'At this ROP, Soft or Extreme soft formation planned to be hit
in next three minutes'
elseif Hforecast3mn >=4 && Hforecast3mn <= 5
    display 'At this ROP, hard or Extreme hard formation planned to be hit
in next three minutes'
end
```

13.10.4 Code of the sub program “HardnessReport”

```
%hardness report program
%%We start creating lists needed for subprogram HardnessReport:
DepthStart=[];
ClassifHardnessWell=[];
TotalLength=length(ClassificationHardness);
for i=2:TotalLength
    if i==1
        DepthStart=[DepthStart DepthHardness(i)];
        ClassifHardnessWell=[ClassifHardnessWell,
ClassificationHardness(i)]
    end
    if ClassificationHardness(i)== ClassificationHardness(i-1)
    else
        DepthStart=[DepthStart, DepthHardness(i)];
        ClassifHardnessWell=[ClassifHardnessWell,
ClassificationHardness(i)];
    end
end
DepthStop=zeros(length(DepthStart),1);
for i=1:length(DepthStart)-1
    DepthStop(i)=DepthStart(i+1);
end
DepthStop(length(DepthStart))=DepthHardness(TotalLength);
%Then we create a plot of hardness classification versus depth
ntable=2*length(DepthStart);
Depthh=zeros(ntable,1);
Classif=zeros(ntable,1);
j=1;
for i=1:2:ntable
   Classif(i)=ClassifHardnessWell(j);
   Classif(i+1)=ClassifHardnessWell(j);
   Depthh(i)=DepthStart(j);
   Depthh(i+1)=DepthStop(j);
   j=j+1;
end
figure (3)
plot(Classif,Depthh)
title('hardness classification versus Depth - Well1')
set(gca,'YDir','reverse')
xlabel('hardness classification')
xlim([0 6])
ylabel('Depth [m]')
legend('Hardness classification along the well path: 1 stand for extreme
soft,2 stand for moderate soft, 3 stand for average, 4 stand for moderate
hard, 5 stand for extreme hard','Location',...
    'southoutside')
Report=[transpose(DepthStart) DepthStop...
    transpose(ClassifHardnessWell)];
```

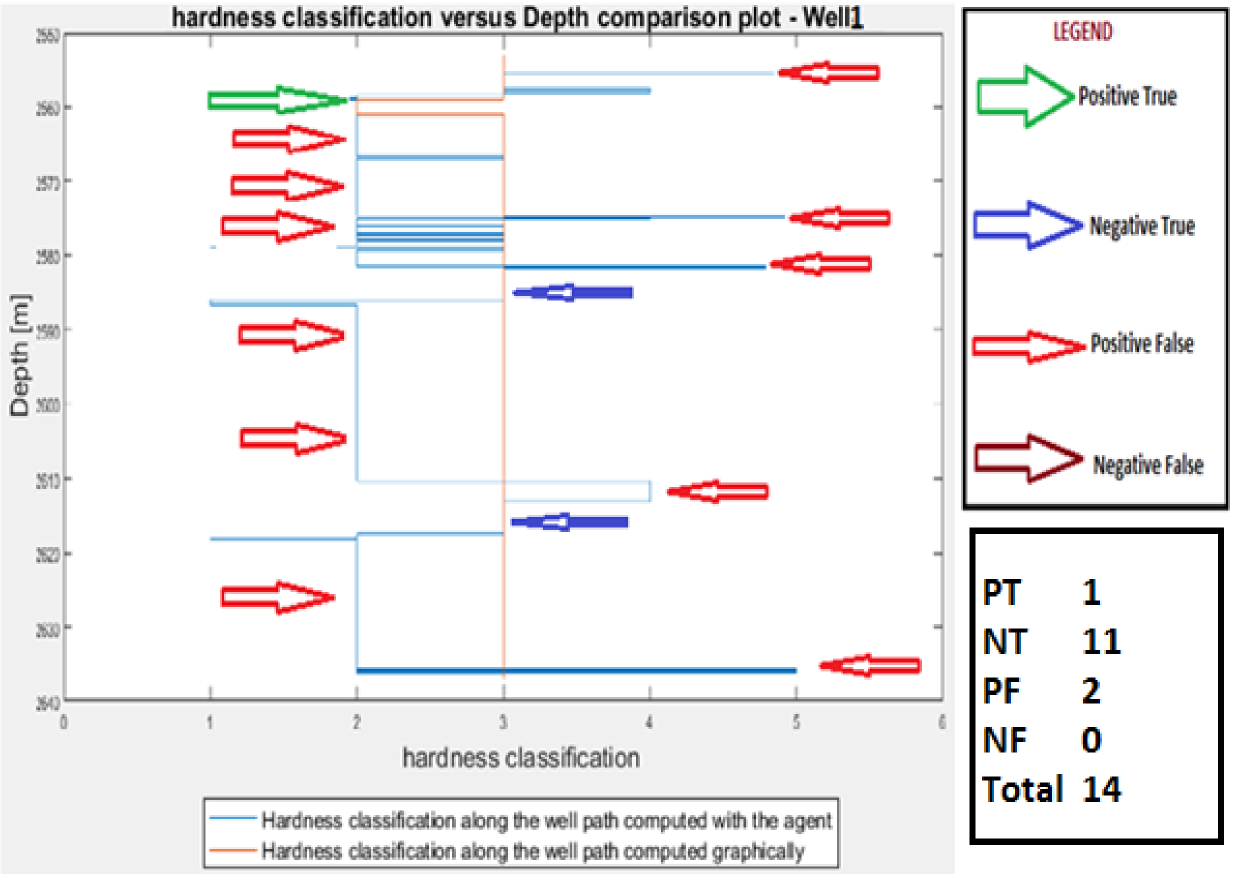
13.11 Appendix XI – New “HardnessReport” sub program code to make comparison plots

```
%hardness report program modified for comparison plot
%load Graphical1.mat
%%We start creating lists needed for subprogram HardnessReport:
%All Graphical files are in the .zip folder attached to the thesis
load Graphical1.mat
MainAgent
DepthStart=[];
ClassifHardnessWell=[];
TotalLength=length(ClassificationHardness);
for i=2:TotalLength
    if i==1
        DepthStart=[DepthStart DepthHardness(i)];
        ClassifHardnessWell=[ClassifHardnessWell,
ClassificationHardness(i)]
    end
    if ClassificationHardness(i)== ClassificationHardness(i-1)
    else
        DepthStart=[DepthStart, DepthHardness(i)];
        ClassifHardnessWell=[ClassifHardnessWell,
ClassificationHardness(i)];
    end
end
DepthStop=zeros(length(DepthStart),1);
for i=1:length(DepthStart)-1
    DepthStop(i)=DepthStart(i+1);
end
DepthStop(length(DepthStart))=DepthHardness(TotalLength);
%Then we create a plot of hardness classification versus depth
ntable=2*length(DepthStart);
Depthh=zeros(ntable,1);
Classif=zeros(ntable,1);
j=1;
for i=1:2:ntable
   Classif(i)=ClassifHardnessWell(j);
   Classif(i+1)=ClassifHardnessWell(j);
   Depthh(i)=DepthStart(j);
   Depthh(i+1)=DepthStop(j);
   j=j+1;
end
%Then we create a plot of hardness classification versus depth following
%data we found in the grapfycal analysis
ntableGraphical=2*length(Graphical);
DepthGraphical=zeros(ntableGraphical,1);
ClassifGraphical=zeros(ntableGraphical,1);
j=1;
for i=1:2:ntableGraphical
   ClassifGraphical(i)=Graphical(j,3);
   ClassifGraphical(i+1)=Graphical(j,3);
   DepthGraphical(i)=Graphical(j,1);
   DepthGraphical(i+1)=Graphical(j,2);
   j=j+1;
end
figure (3)
plot(Classif,Depthh,ClassifGraphical,DepthGraphical)
title('hardness classification versus Depth comparison plot - Well1')
set(gca,'YDir','reverse')
xlabel('hardness classification')
```

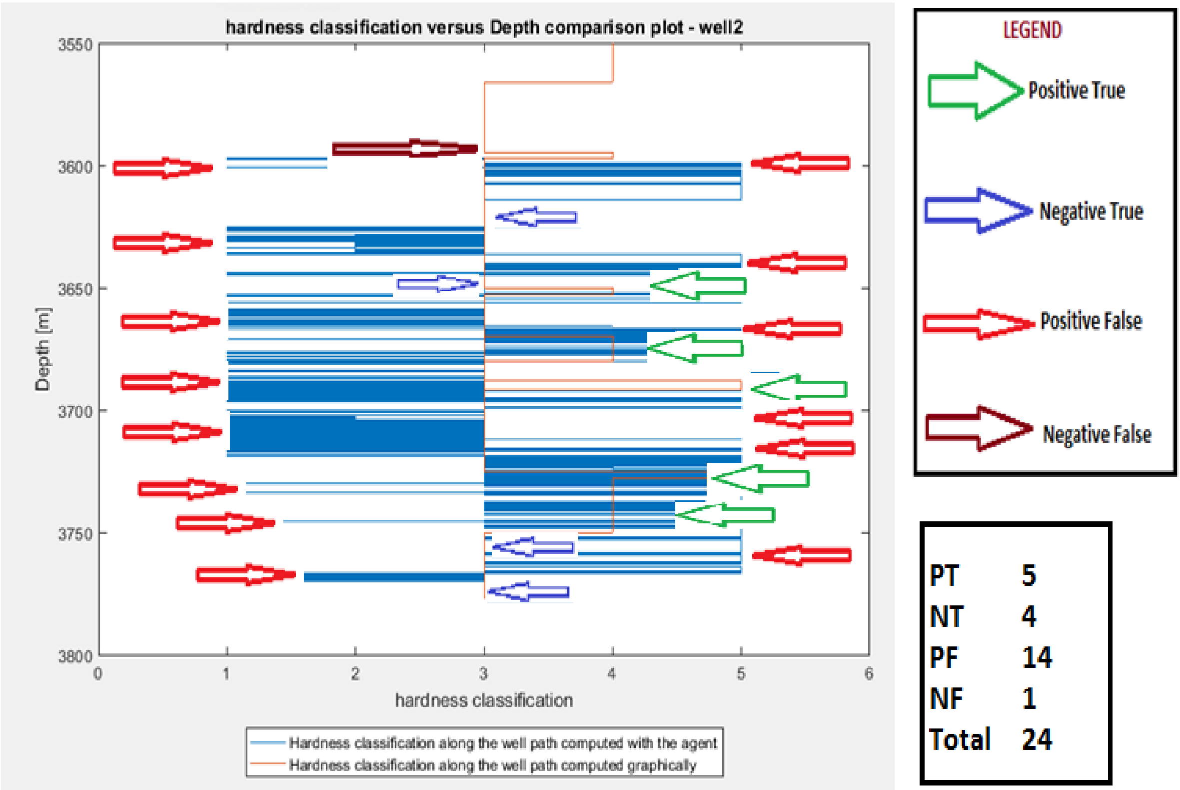
```
xlim([0 6])
ylabel('Depth [m]')
legend('Hardness classification along the well path computed with the
agent: 1 stand for extreme soft,2 stand for moderate soft, 3 stand for
average, 4 stand for moderate hard, 5 stand for extreme hard','Hardness
classification along the well path computed graphically'...
    ,'Location','southoutside')
Report=[transpose(DepthStart) DepthStop...
    transpose(ClassifHardnessWell)];
```

13.12 Appendix XII - Hits determination figures for the ten wells

13.12.1 Hits determination figure – Well 1

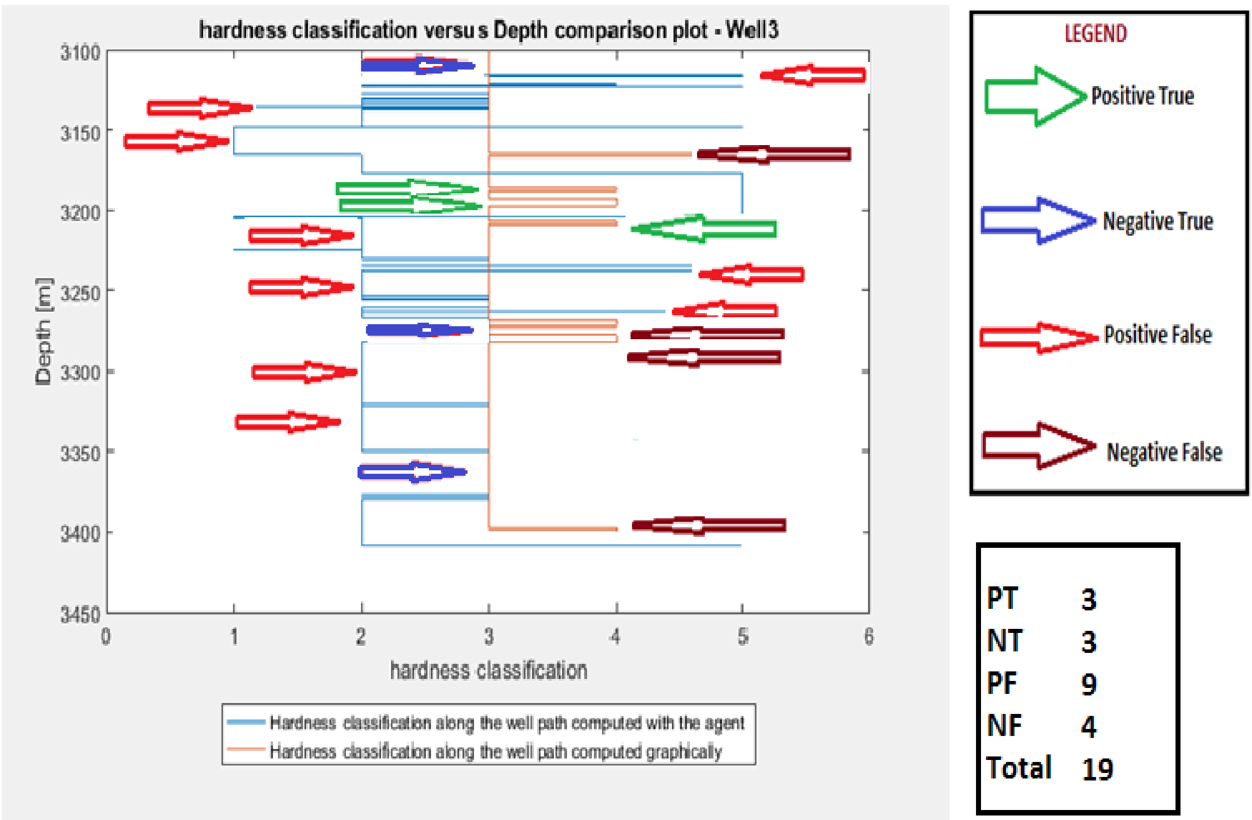


13.12.2 Hits determination figure – Well 2



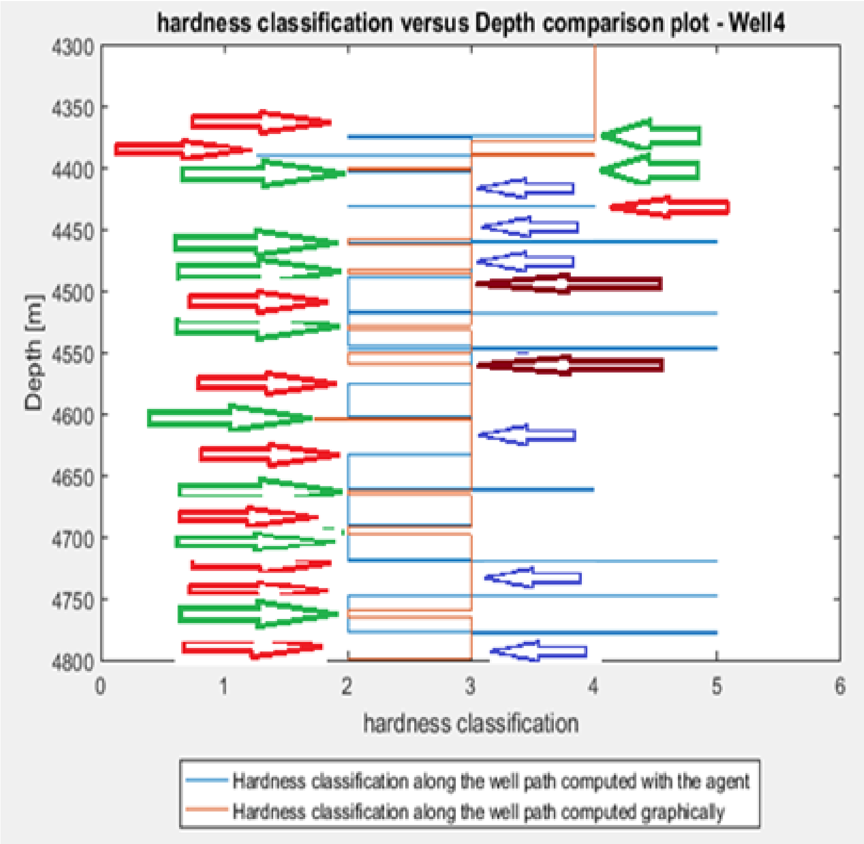
PT	5
NT	4
PF	14
NF	1
Total	24

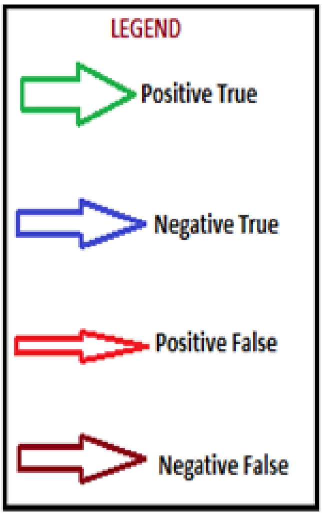
13.12.3 Hits determination figure – Well 3



PT	3
NT	3
PF	9
NF	4
Total	19

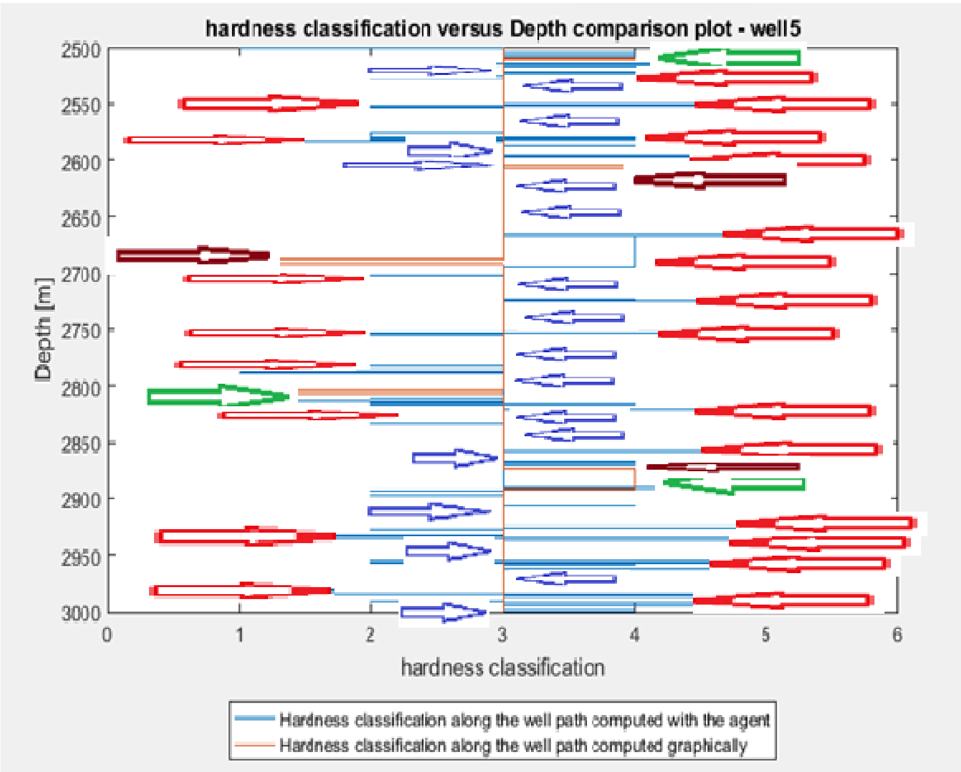
13.12.4 Hits determination figure – Well 4

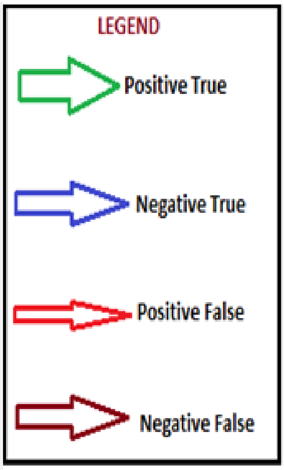




PT	10
NT	6
PF	10
NF	2
Total	28

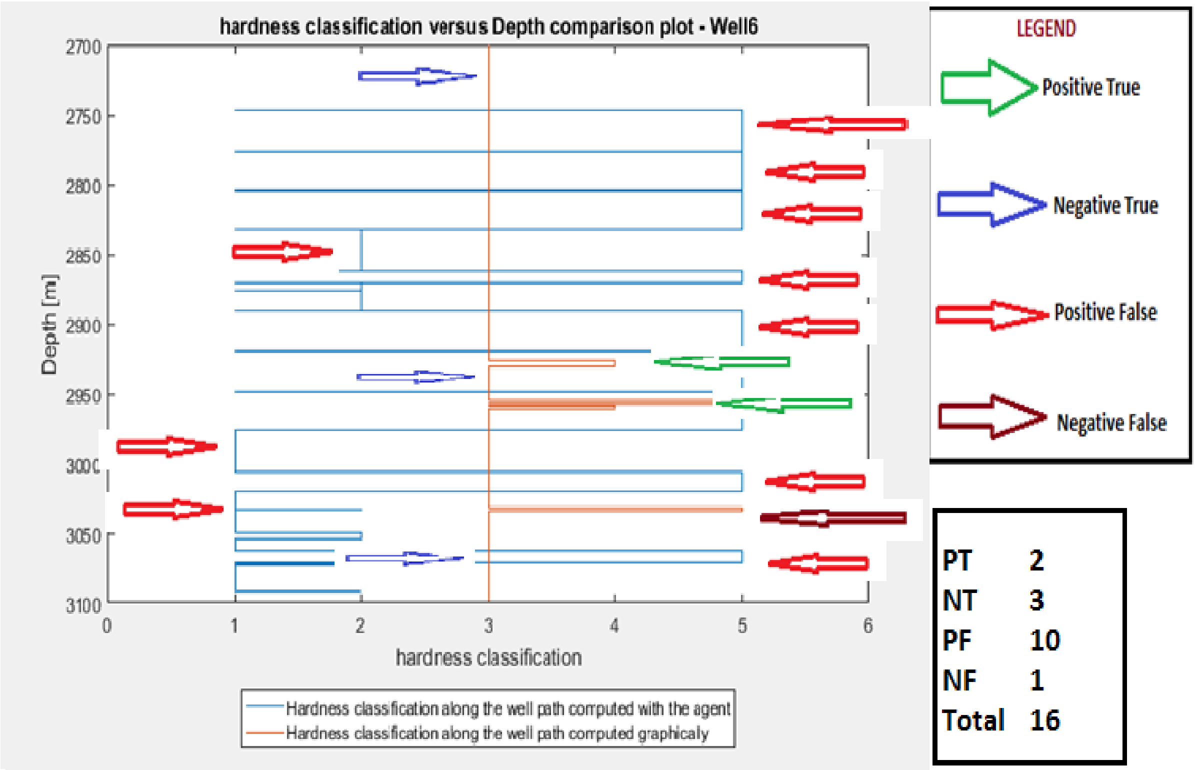
13.12.5 Hits determination figure – Well 5



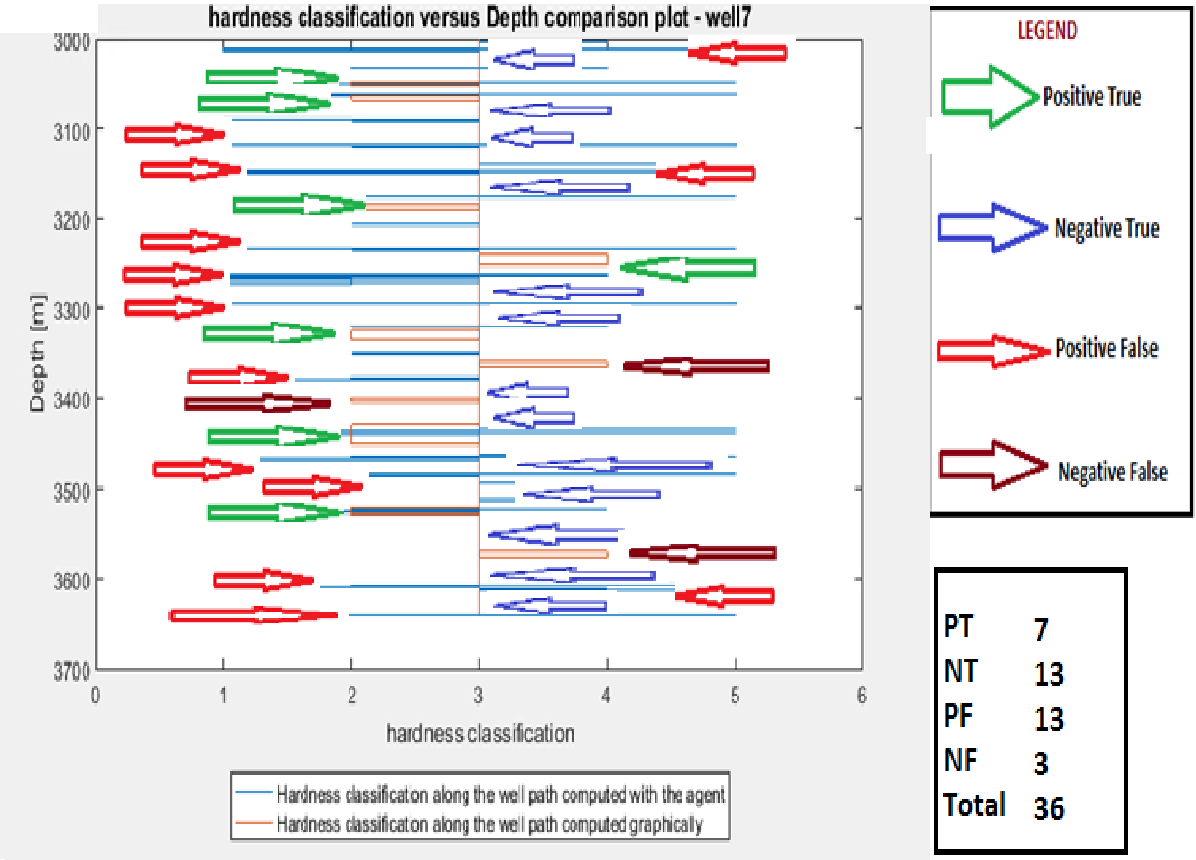


PT	3
NT	18
PF	22
NF	3
Total	46

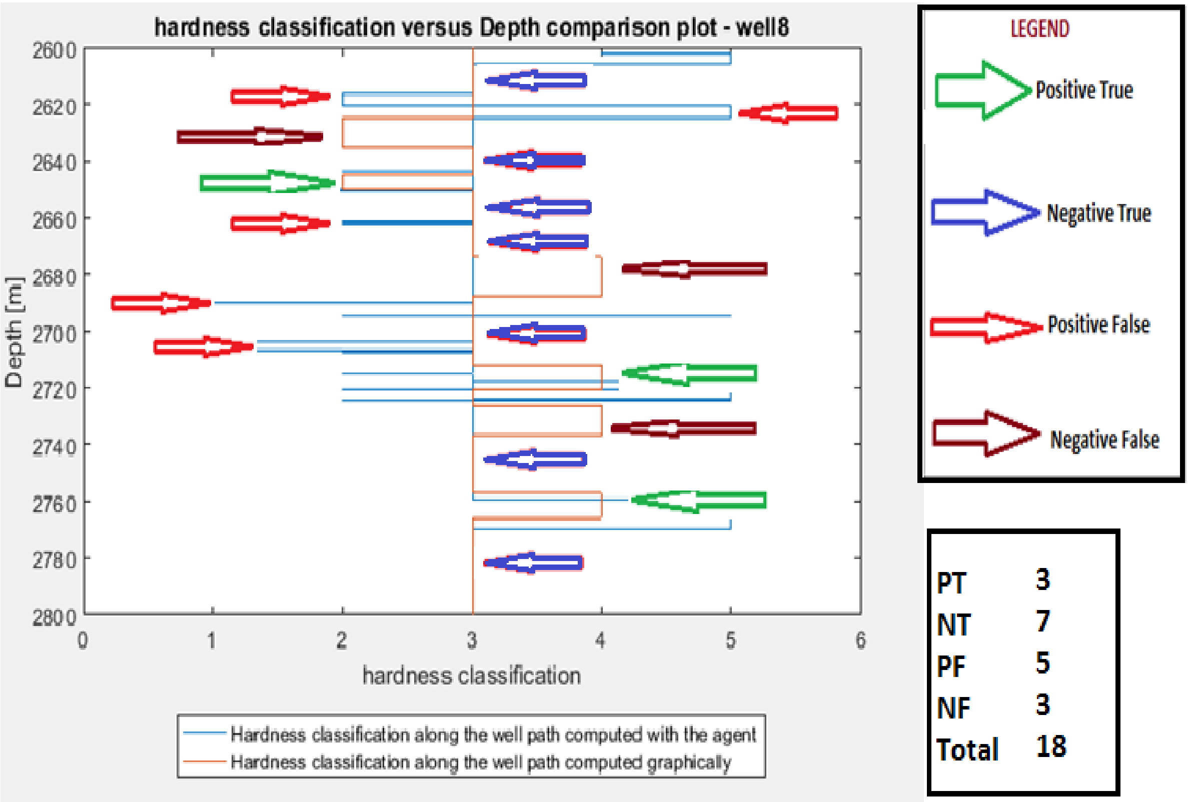
13.12.6 Hits determination – Well 6



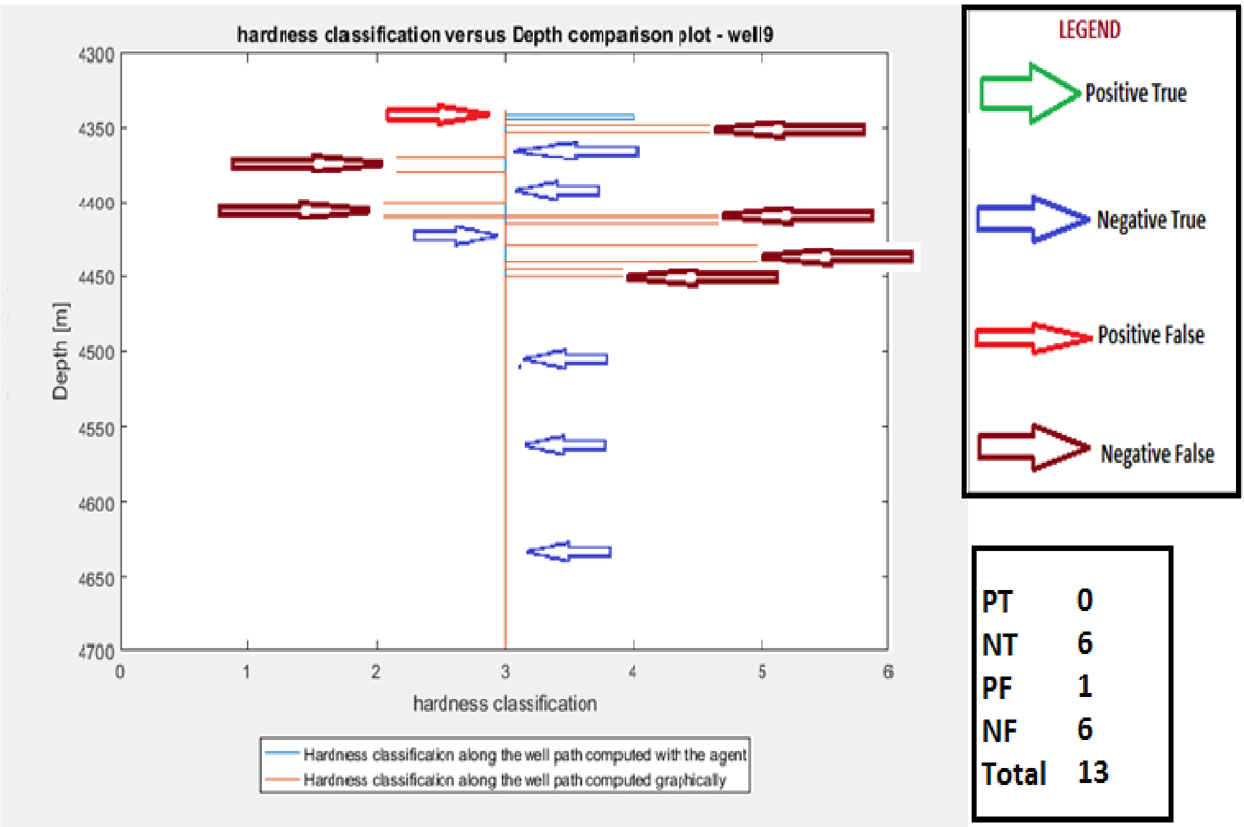
13.12.7 Hits determination – Well 7



13.12.8 Hits determination – Well 8



13.12.9 Hits determination – Well 9



13.12.10 Hits determination – Well 10