Ref. Ares(2024)7484230 - 22/10/2024

D9.1 Report on Assessment and Quantification of Drilling Problem Scenarios

Deliverable No	D9.1
Work package No. and Title	WP9 Development of drilling problems detection and prediction machine learning model
Version – Status	V1.0 – draft
Date of Issue	14/06/2022
Dissemination Level	PUBLIC
Filename	OptiDrill-D9.1

This project has received funding from the European Union's Horizon 2020 research and innovation action under grant agreement No 101006964

DOCUMENT INFO

Authors

Author	Organization	e-mail
Shahin Jamali	IEG	Shahin.jamali@ieg.fraunhofer.de
Henning Knauer	IEG	Henning.knauer@ieg.fraunhofer.de

Document History

Date	Version	Editor	Change	Status
14.06.2022	V0.1	Shahin Jamali, Henning Knauer	Table of contents	Draft
22.06.2022	V0.2	Henning Knauer	Executive summary, introduction	Draft
28.06.2022	V0.3	Henning Knauer	Chapters 2,3 and 4	Draft
17.10.2022	V1.0	Shahin Jamali	Revision and acknowledgment addition	Final

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Work package number 9 has the objective of developing and validating a drilling problems detection and prediction machine learning model that is capable of predicting specific, problematic drilling events from the surface readings available in real time. This deliverable builds on the results from the deliverables 6.1 and 6.2 dealing with the drilling problematic scenarios.

The aim of the task that is being reported on was to develop methods for the further interpretation of the data gathered and for the enabling of the usage of this data for machine learning modeling application. Interpreting the data describing the identified problematic drilling scenarios and being able to correlate the data at hand in text form with the numerical data available are challenging but crucial steps for the creation of labeled drilling datasets that can be used for the supervised artificial intelligence methods. An approach for theses steps was developed and is described in this report.

1. INTRODUCTION

This report describes the interpretation methods chosen and developed and the workflow elaborated for the analysis of the data gathered describing the problematic drilling scenarios. The objective of the whole process is to enable the creation of labelled problematic drilling scenario datasets that can later be used for the development of the artificial intelligence-based module from work package number 9 for the detection and prediction of these events by analysing drilling process data available in real time at the surface of the drilling site.

Since artificial intelligence-based methods are data driven and therefore the performance of these models highly depends on the quality of the data used for the development, the creation of the datasets is a very important and fundamental process. The algorithms that will be implemented at later stages within work package 9 for the detection and prediction of problematic drilling scenarios will be supervised learning algorithms, meaning that labelled datasets are necessary. Unlike for other regression or classification tasks within this project, such as the prediction of the rate of penetration or the prediction of the lithology being drilled, where these labels are already present in the datasets gathered and can be directly accessed, for this particular task these labels have to be generated. The datasets that will be created with the methods and workflow presented within this report will be based on the surface readings extracted from the log files available and additional information describing the events that occurred during the drilling process, which will be the target of the prediction tasks. This additional information on any problematic drilling events has to be manually extracted from the documentation available for each well and requires expert knowledge, which is provided by the OptiDrill's project partners.

The whole datasets generation process can basically be divided into three subprocesses, the analysis of the drilling projects documentation available by an expert, the analysis of the drilling process parameter datasets using an anomaly detection algorithm and the correlation of the results from both processes resulting in a new drilling dataset.

The analysis of the available drilling documentation by an expert is described in section 2. This step requires expert domain knowledge and if possible also specific knowledge on the respective project. The more information is available and the better the results from this particular step will be. The points described within section two are the data at hand for the analysis, the expert analysis, interpretation and annotation process and finally the results to be achieved.

Section 3 deals with the analysis of the numerical drilling process parameter datasets. The analysis is based on the so-called isolation forest algorithm, which is used for anomaly detection purposes. The contents of section 3 are the data used for the analysis, the anomaly detection algorithm implemented and the results that will be obtained from the whole process.

The last section of this report describes the final and most challenging process of the whole task which is the drilling problem scenarios quantification. Within this process the results obtained from the drilling documentation expert analysis and the automatic anomaly detection are compared and correlated. The points described within this last section are the manual comparison and confirmation of the numerically detected anomalies and the creation of the quantified and annotated drilling problem dataset.

2. DRILLING DOCUMENTATION EXPERT ANALYSIS

Supervised learning based artificial intelligence algorithms required labelled datasets. This means that each set of inputs that will be fed into the model during the training or testing phase is provided with an associated label. In the case of the regression task for the rate of penetration prediction this label would be a continuous positive numerical value quantifying the magnitude of the speed at which the bit travels through the rock. For the task addressed within work package number 9, which is a classification task, this label will be a string describing the class of event that occurred at a certain depth. Since this kind of information is not included in the datasets containing the surface readings logged during the drilling process, it has to be gathered from the drilling documentation available.

2.1 Data

As described within previous reports dealing with data availability and the data gathering and extraction process certain parts of the documentation are not always available. The most vital part of the documentation necessary for this step in the dataset creation workflow are the daily drilling reports. Unfortunately, these documents are not always available and are sometimes partly or even completely missing for certain wells. In this case it will not be possible to gather any information that can be used for the labelling of the respective drilling process parameter datasets.

Operations

Start time	End time	End Depth mMD	Main - Sub Activity	State	Remark
00:00	01:30	132	drilling -- casing	ok	Ran in with 30" conductor on dedicated landing string to 132 m MD. Filled conductor with 20 m2 seawater until returns thorugh shoe observed. Hookload after filling 108 MT.
01:30	02:00	148	drilling -- casing	ok	Stung 30" conductor into funnel on F-9 while observing with ROV.
02:00	03:00	221	drilling -- casing	ok	Ran in with 30" conductor from 148 m to 221 m MD, only minor drag in hole. Observed 112,68 m mark on landing string flush with machined ring on Centralizer deck. Meanwhile deployed hotstab line on tugger wire.
03:00	04:00	221	drilling -- casing	ok	Picked up and made up cement stand. Accidentally applied torque on TDS and put 2,5 turns righthand rotations on string. Inspected CART with ROV and backed CART in again using lefthand turns in rotary table. Performed digiquartz measurement on F-7 and F-9 indicating F-9 12 cm too high.
04:00	05:00	221	drilling -- casing	ok	Made up cement hose to side entry sub. Adjusted height on F-9 conductor 10 cm down. Repeated elevation check on conductor housing by doing consecutive measurement on F-9 and F-7. Measured F-9 elevation 3 mm higher that F-7 - ok. Bad visibility for establishing hotstab on template.
05:00	06:00	221	drilling -- casing	ok	Poor visibility on template base for establishing hotstab. Meanwhile tested cement line against lo-torque on side entry sub to 200 bar 10 min - ok. Made readings on inclinometer : pitch 1,3 deg / roll 0,3 deg.
06:00	07:00	221	drilling -- wait	ok	Waited for visibility to improve for engaging hotstab.
07:00	08:15	221	drilling -- casing	ok	Removed dummy stab and installed hotstab in bank #4 on template base. Shifted lever to "on". Energized centralizing pistons with 2000 psi and observed travel on pistons with ROV. Made readings on inclinometer : pitch 1,27 deg / roll 0,3 deg.
08:15	08:30	221	drilling -- casing	ok	Broke circulation and verified conductor full of water by observing return through fillup valve. Closed fillup valve with ROV. Hookload before circulating 87 MT.
08:30	09:30	221	drilling -- casing	ok	Stepped pumps up to 1500 lpm and circulated 50 m3 seawater. Poor visibility, assumed to be caused by returns from hole. Meanwhile performed prejob meeting prior to cementing. Hookload after circulating 87 MT.
09:30	10:45	221	drilling -- casing	ok	Performed cement job. Mixed and pumped 54 m3 1,52 sg tuned light cement slurry from cement unit. Pump rate 650-750 lpm. Pump pressure 22 bar. Displaced cement with 6,5 m3 sea water to put theoretical TOC at 217 m MD inside conductor. Bad visibility due to no current, hard to judge amount of returns from hole. Hookload after cementing 97 MT.
10:45	11:15	221	drilling -- casing	ok	Rigged down cement line and flushed through manifold.
11:15	21:00	221	drilling -- casing	ok	Waited for cement to set up. Meanwhile : -Serviced TDS -Serviced fwd and aft PRS -Changed annular element -Housekeeping, cleaning and maintenance on drillfloor
21:00	23:30	221	drilling -- casing	ok	Waited on cement to set up. Meanwhile : -Made x-y measurements on conductor relative to funnel 36 cm/36 cm -Shifted bank #4 handle to "off" and disengaged centralizer cylinders with 2000 psi through minireel -Retrieved hotstab and guidewire -Removed lo-torque on cement stand -Performed toolbox talk prior to releasing CART
23:30	00:00	221	drilling -- casing	ok	Slacked off 11 MT to obtain neutral at CART. Released CART by 5 righthand rotations. Lifted out CART from conductor housing and set slips on landing string.

Figure 1: Excerpt from a daily drilling report from well 15/9-F-9 on the Equinor Volve field

These so-called daily drilling reports, short DDRs, are brief and succinct reports that hold information on all important events that occurred on a particular day at the drill rig. They are normally in a 24-hour format and report on any events that happened with the respective time and depth. An excerpt of a daily drilling report from an offshore well in Norway in the Equinor Volve field is shown in Figure 1. In addition to the key events that occurred during the day the daily drilling reports often also contain further information on the wellbore, drilling fluid etc. Depending on the time needed to complete the well and the days of activity the number of daily drilling reports available for a certain well can vary quite significantly.

2.2 Expert Analysis, Interpretation and Annotation

The analysis of the daily drilling reports involves significant effort in terms of manually browsing through the reports and scanning them for activities conducted and any interesting drilling related events that occurred. In this particular case for the test of the methods developed and the workflow elaborated this task was carried out by Kevin Mallin and his team from the OptiDrill's project partner Geolorn Ltd.

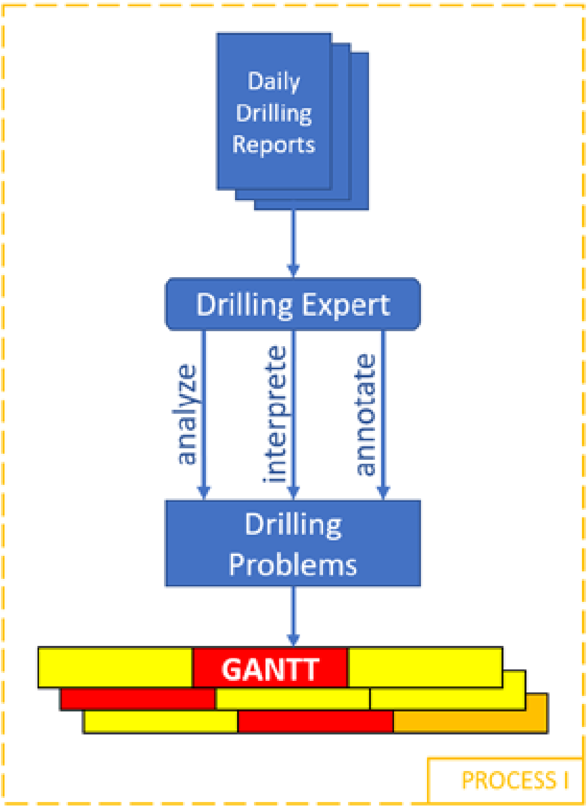

Figure 2:Expert analysis workflow (process 1)

Figure 2 gives a brief overview of the analysis process of the daily drilling reports. The daily drilling reports are normally present as separate PDF documents for each day of activity. The drilling expert checks each of the documents manually and gathers the pieces of information that are of interest for the later creation of the drilling dataset in a separate document for all the reports available.

For the first tests that have been carried out data from the Equinor Volve field[1], which is publicly available, was used since in that way there will not be any conflicts regarding confidentiality of the data and the public status of this deliverable. The data shown within the examples in this report mainly comes from the following wells:

- 15/9-F-14
- 15/9-F-9 A

The well 15/9-F-14 from the Volve field comes with a total of 134 individual daily drilling reports, which were all processed and used to create a Gantt chart representing the drilling activities conducted. An example of the result obtained from the daily drilling reports from well 15/9-F-14 is shown in the next section.

---

[1] https://www.equinor.com/energy/volve-data-sharing

2.3 Results

Figure 3 shows an excerpt from the generated Gantt chart for the well 15/9-F-14 from the Equinor Volve field. The Gantt chart was created by Kevin Mallin and his team from the OptiDrill's project partner Geolorn Ltd.

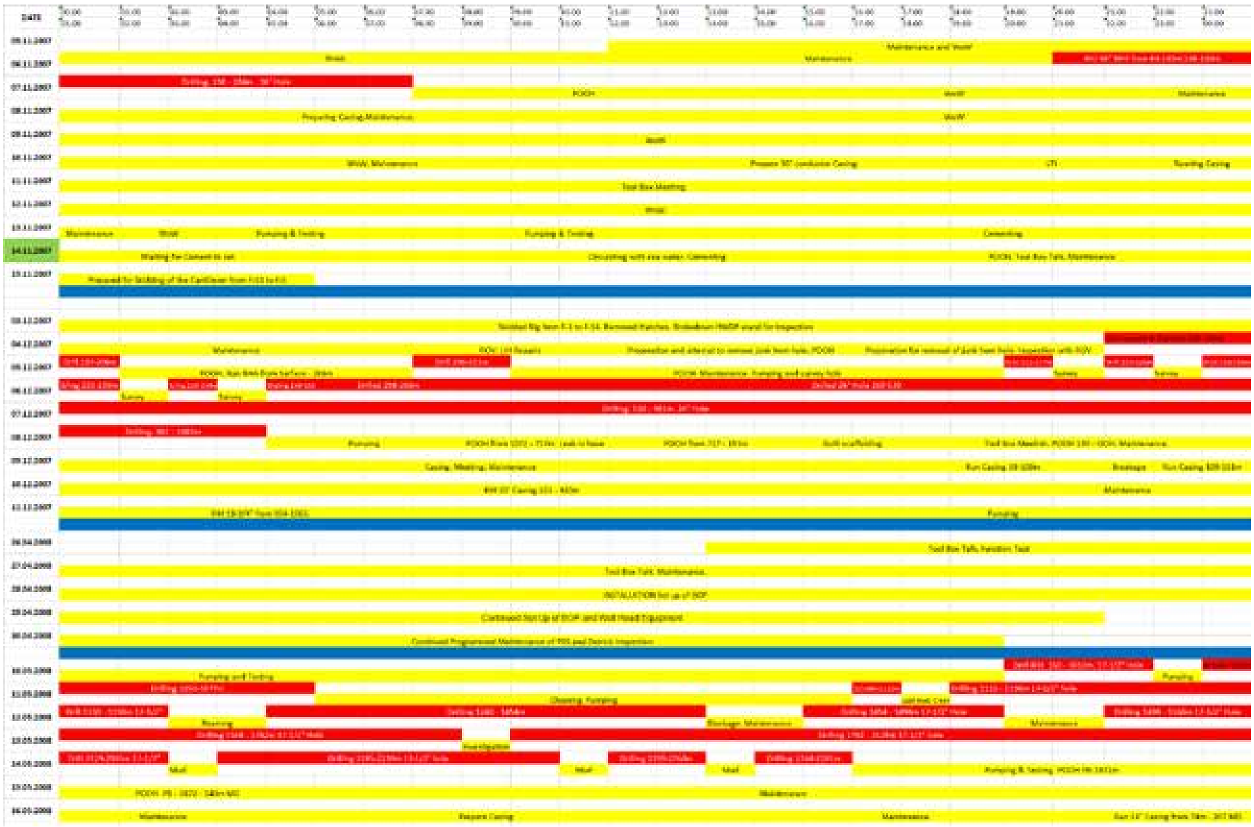

Figure 3: Excerpt from the Gantt chart for well 15/9-F-14

The Gantt chart depicted differs a bit from the typical Gantt charts known from project management purposes in regard to its structure and organisation. The first column holds the dates for which daily drilling reports are available and in the first two rows the dates are listed in one-hour increments. The activities represented by the bars filling the chart are mainly divided into four groups with different colours. Time periods in which wo work was conducted on the rig are left white without any additional label. The time periods in which actual drilling operations were conducted are coloured in red with labels describing the actual activity, such as “Drilling 530 – 981m 26” Hole”. Non-drilling hours are represented by yellow bars with the respective label describing the reason for the interruption of the drilling process, such as “Maintenance” or “Survey”. Any time periods representing activities related to well abandonment are coloured in orange. Additionally, gaps in the timeline for which no daily drilling reports are available, due to the fact that no work was conducted at the rig or that the documents are simply missing, represented by blue bars.

It can be seen that most of the time documented in the daily drilling reports accounts to non-drilling time and that the actual drilling hours have a much smaller proportion. Also, for this particular well it is very obvious that there were a number of longer interruptions of a few weeks up to over half a year during the course of the drilling programme where nothing happened at the drilling site.

The Gantt charts created within this step of the whole process will be used together with the results from the anomaly detection, which is described in the next section, for the generation of the training datasets. With the extensive and detailed overview over the whole drilling programme and the activities carried out at the drilling site, which these Gantt charts offer, the process of matching found anomalies in the data with the actual events and activities that caused these anomalies is expected to be more efficient and straight forward.

The following figures show some more detailed excerpts of the Gantt chart shown in Figure 3.

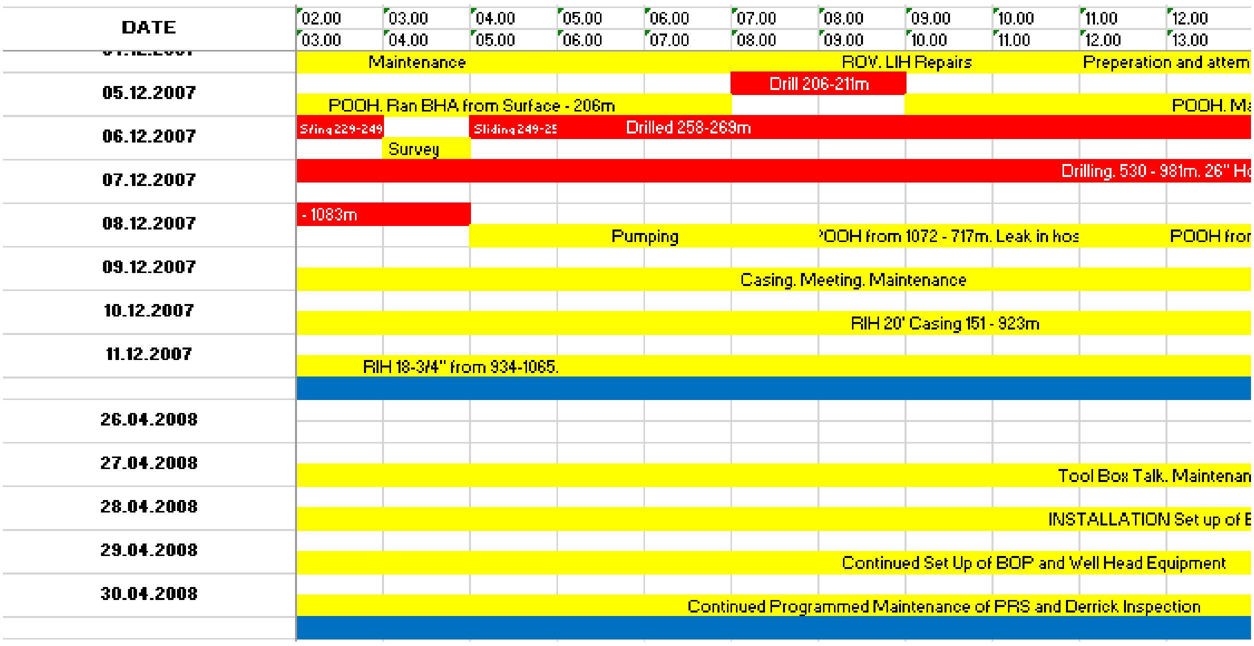

Figure 4: Gantt chart excerpt 1

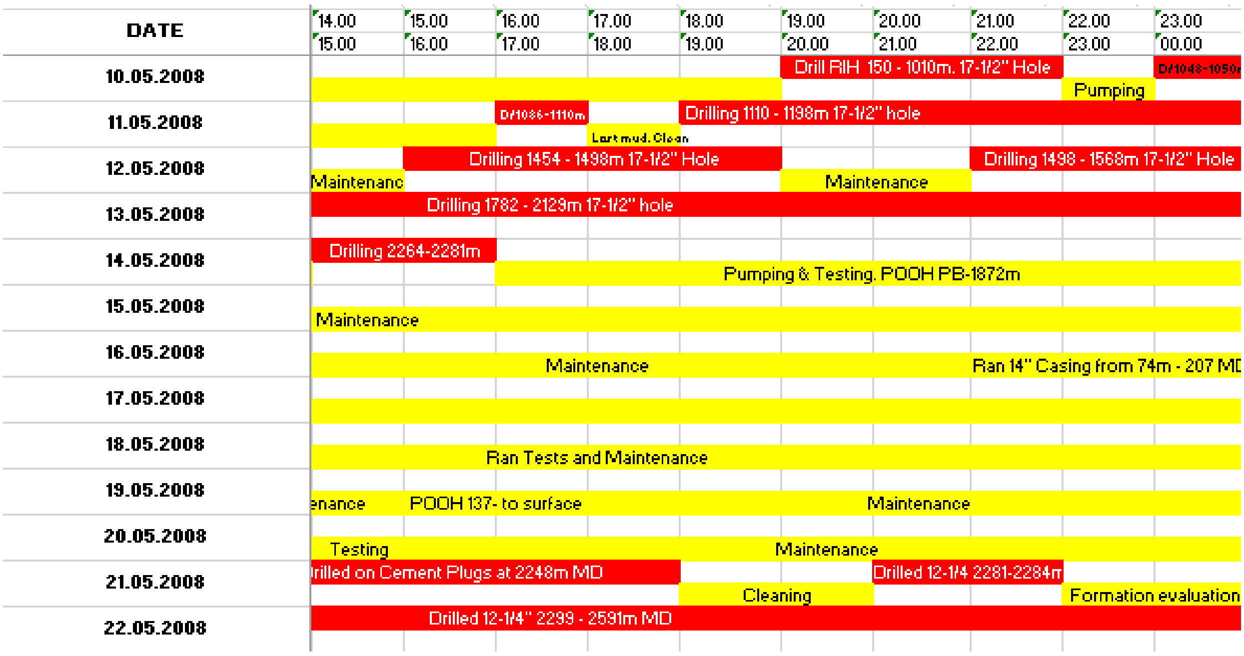

Figure 5: Gantt chart excerpt 2

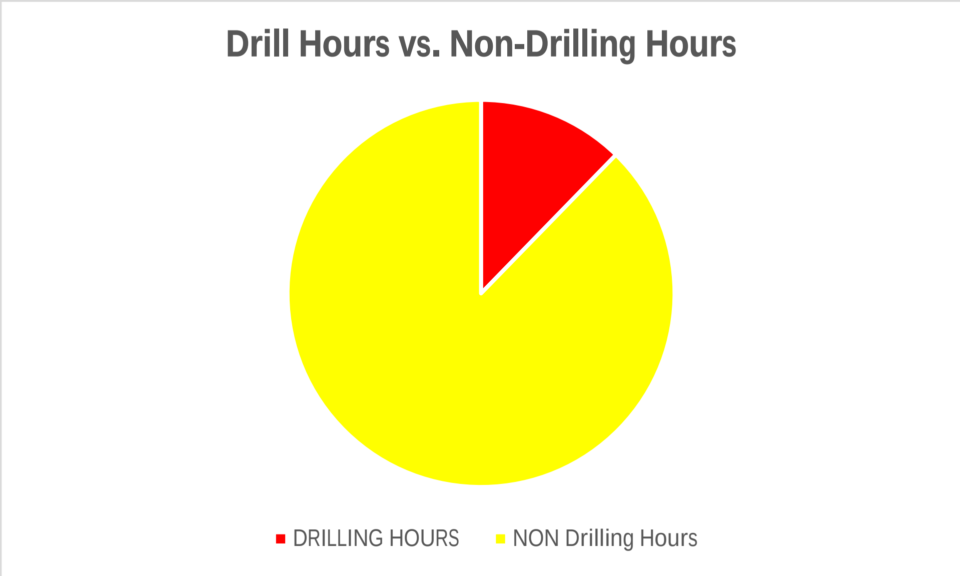

Figure 6: Pie chart showing the proportions of drilling and non-drilling hours

Another very striking insight into the drilling programme that you get from looking at the Gantt chart is that there is a significant imbalance in the distribution of time spent for actual drilling and time spent for non-drilling activities. The whole amount of time spent working at the drill rig accounts to 2173 working hours. Only 267 hours were spent on drilling activities, accounting to roughly 12% of the whole working hours. The remaining 1906 hours were spent on other activities, such as maintenance, cleaning, installation of equipment etc., not resulting in actual depth wise progress of the borehole.

3. DRILLING DATASET AUTOMATIC ANOMALY DETECTION

The next important step in the whole process of creating datasets that can be used as a basis for the development of drilling problem detection and prediction models is the automatic anomaly detection. In this process the framework developed and reported on within deliverable 5.4 under section 4 called "ANOMALY DETECTION AND ANOMALY FEATURE EXTRACTION FOR DRILLING PROBLEM DETECTION" is used to find anomalous data point within the datasets at hand. These data points differ from the majority of the data points present in the dataset and are most likely to point to events and problematic drilling scenarios that are of interest for the dataset creation.

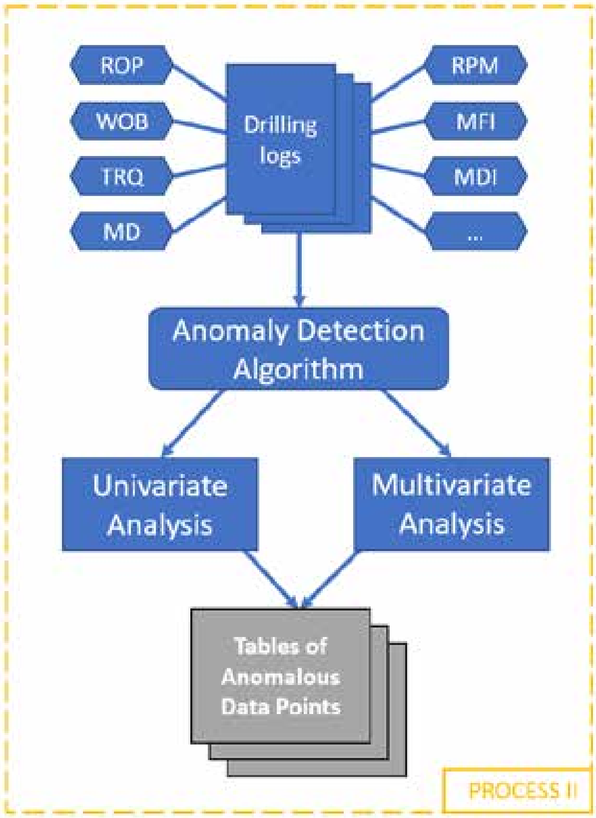

Figure 7: Automatic anomaly detection workflow (process 2)

Figure 7 shows the basic workflow followed within this step of the process.

3.1 Data

The data used for this process can be made available to the anomaly detection script in different file formats. The script which is in this case a Jupyter notebook programmed in python that accesses a python library, both developed by the project partner BGS and already mention in the deliverable 5.4, can handle .LAS files as inputs files and with some minor modifications also .csv files. Since for the majority of the datasets available at this stage of the project extensive csv files have been generated they can be directly used as inputs to the Jupyter notebook. Also, for some of the datasets no .LAS files and only csv or excel files are available, so being able to process csv files is absolutely necessary and will also speed up the whole process since not huge amounts of log files have to be browsed and processed.

The drilling process parameters we are looking for in the input data are basically all of those that are available in real time at the surface while drilling. The parameters of most interest used for an anomaly detection, among a number of other drilling parameters, could for example be the following:

- Measured depth (M_D)
- Rate of penetration (ROP)
- Revolutions per minute (RPM)
- Torque (TRQ)
- Weight on bit (WOB)

Two approaches for the automatic anomaly detection using the same algorithm were implemented a single-variate and a multi-variate version. The single variate version is using only the rate of penetration as input, while the multi-variate version can take a selection of different drilling process parameters as input. Both versions are further described in the next section.

3.2 Anomaly detection algorithm

The anomaly detection algorithm chosen for this particular task of finding anomalous data points within drilling process parameter datasets is the so-called isolation forest algorithm. The isolation forest algorithm dates back to the year 2008 and was first presented in the paper by Liu et al. for the detection of data-anomalies using binary trees. This resembles the well-known random forest algorithm used for classification and regression tasks in supervised learning applications. The isolation forest however is an unsupervised learning method, meaning that it can work with unlabelled data, which is the case in our application. We are planning to use this algorithm for the purpose of labelling our data for further development steps.

The isolation forest algorithm assumes that the data points that are anomalies are a minority in the whole dataset and they differ from the majority of the normal data points regarding their attribute values. Similar to the random forest algorithm the isolation forest algorithm is based on an ensemble of binary trees that are build on the dataset used. Thanks to the nature of the algorithm it normally converges very quickly with a small number of trees.

The isolation forest algorithm can be implemented in python using the scikit-learn library. It can be found in the ‘ensemble’ package. The algorithm once trained on the dataset return an anomaly score for each data sample that defines whether the respective sample is anomalous or not. In case of an anomaly the algorithm returns a score of -1 and in case no anomaly was detected a 1 is returned.

Since the algorithm offers only a very small number of hyperparameters their tuning does not play an important role in the overall implementation of the algorithm. The most important hyperparameter for the implementation of the isolation forest for our use case is the ‘contamination’. It can be set using a float value between 0 and 0.5 and determines the proportion of outliers present or expected in the dataset. Setting the contamination values to 0.1 means that the algorithm will categorise the most anomalous 10% of whole dataset as anomalies.

As mentioned before two versions of the algorithm were implemented using different subsets of input parameters for the anomaly detection. The two versions are briefly described in the next two sections and the result plots are depicted.

3.2.1 Uni-Variate Analysis

The uni-variate implementation of the isolation forest only takes one input feature into account for the determination of anomalous data points. In this case the chosen input feature is the rate of penetration. The plotted results from the algorithm are shown in Figure 8. The x-axis displays the measure depth of the well in meters and on the y-axis the log values of the rate of penetration are plotted. The different background colours represent the different hole sections of the wellbore. Data points that are classified as anomalies are highlighted in red.

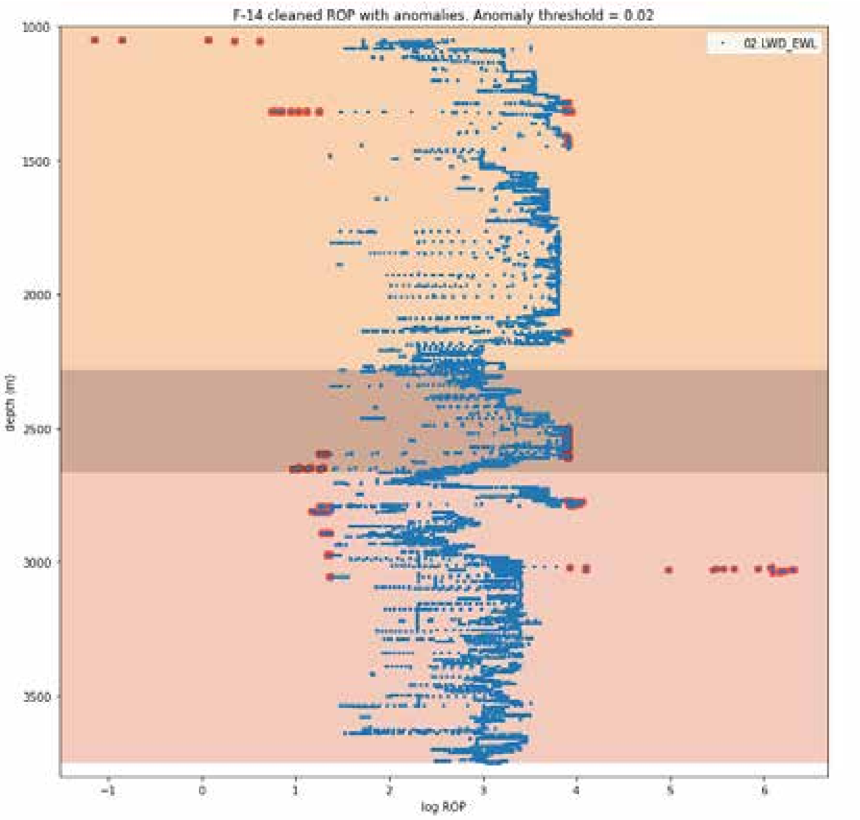

Figure 8: Anomalies detected using the uni-variate version of the isolation forest in well 15/9-F-14

For this run the available log data for the well 15/9-F-14 from the Equinor Volve field was used. The algorithm used a contamination value of 0.02 meaning that 2% of all data points are classified as anomalies.

3.2.2 Multi-Variate Analysis

The multi-variate implementation of the isolation forest can take multiple input features into account for the determination of anomalous data points. In this case the chosen input features were the following:

- Measured depth
- Rate of penetration
- Total flow
- Weight on bit
- Revolutions per minute

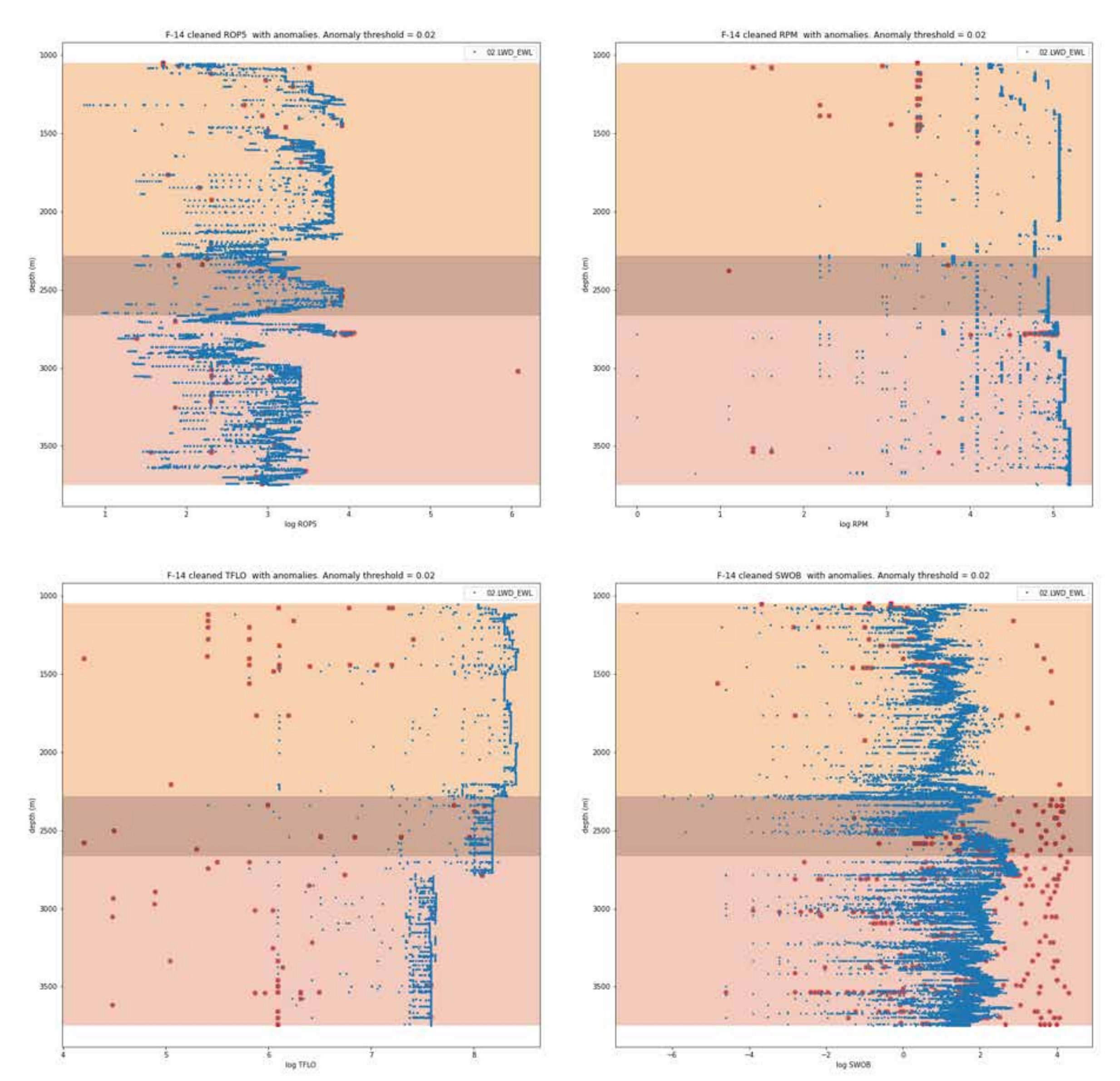

Figure 9: Anomalies detected using the multi-variate version of the isolation forest in well 15/9-F-14

The plotted results from the algorithm are shown in Figure 9. Just as before the different background colours represent the different hole sections of the wellbore. Data points that are classified as anomalies are highlighted in red.

3.3 Results

The plotted results from the uni- and multi-variate anomaly detection for well 15/9-F-14 are depicted in Figure 8 and Figure 9. Figure 10 compares the plotted rate of penetration values with the detected anomalies for both approaches.

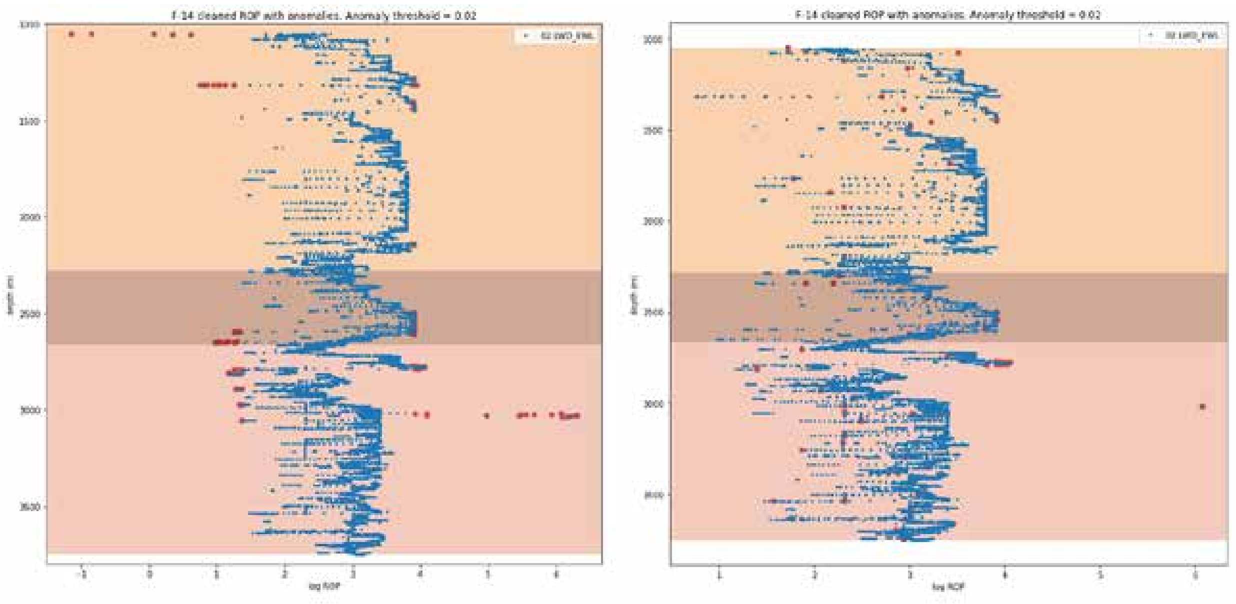

Figure 10: Plotted ROP curves with anomalies for the uni-variate and multi-variate isolation forest

Looking at the plotted rate of penetration with the highlighted anomalies of the uni- and multivariate anomaly detection analysis shows that the anomalies detected are not exactly the same for both plots. This is due to the fact that for both runs different sets of parameters are used as inputs, as described in the previous two sections. Another factor contributing to the differing results are the actual data samples used as inputs for the models. Because it is most likely that not all the process parameters used as inputs for the multi-variate anomaly detection analysis are available at each depth of the well the data used will be more incomplete compared to the data used for the uni-variate version of the algorithm.

In addition to the plot the anomaly detection framework also outputs a csv file with all the anomaly scores for all data points in the dataset. Table 1 shows an excerpt of the output csv file listing the depths with the assigned anomaly scores for a small section of the well 15/9-F-14.

Table 1: Excerpt from the csv output file for the multi-variate anomaly detection for well 15/9-F-14

DEPT	ROP5	TFLO	SWOB	RPM	anomaly
2339.49	6.73	3296.44	54.29	41.92	-1
2772.46	50.53	3227.49	17.50	153.33	-1
2772.61	50.96	3227.49	18.35	152.00	-1
2772.77	50.66	3227.49	18.35	154.00	-1
2772.92	50.83	3227.49	17.74	134.21	-1
2773.07	51.01	3227.49	17.74	141.86	-1
2773.22	51.04	3227.49	17.42	143.09	-1
2773.38	50.94	3227.49	17.51	140.55	-1

4. DRILLING PROBLEM SCENARIOS QUANTIFICATION

The last step in the drilling problem scenarios quantification and the dataset creation process is depicted in Figure 11. The inputs for this last subprocess are the results from the processes 1 and 2 described in the previous two sections. The Gantt charts created in process 1 give an extensive and detailed overview of the activities that were carried out at the drilling site and offer valuable information for the manual processing of the data and documentation. The anomaly plots and the tables with all the data points detected during the automatic anomaly detection using the isolation forest algorithm will provide more insight into the actual numerical data and will play the most important role in the labelling process of the original data and the final dataset creation process.

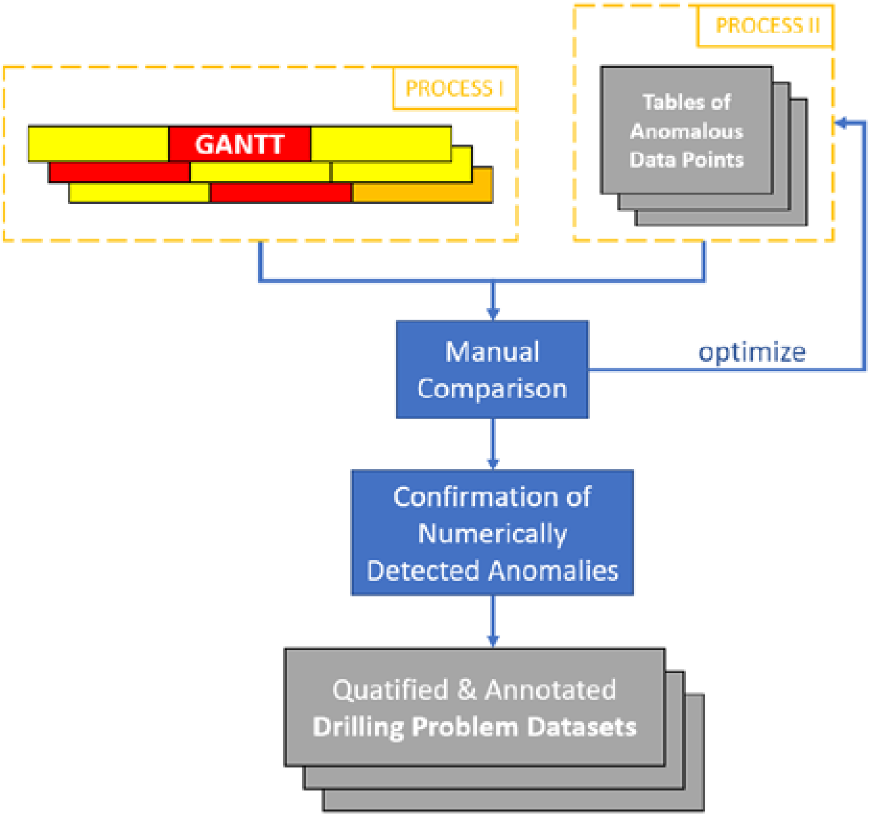

Figure 11: Manual comparison and drilling problem datasets creation workflow (process 3)

This labelling process will have to remain a manual process for the generation of the training datasets since this very part relying of expert human knowledge cannot be automated. Figure 11 shows the workflow of the process with the results of the previous processes at the top followed by to main tasks which are represented by the blue boxes. These tasks, the manual comparison and the confirmation of the numerically detected anomalies are described in the next section. The arrow pointing back at process 2, which is the automatic anomaly detection, emphasizes that this process will most likely have to undergo some further development and optimization depending on the results achieved. Some ideas for possible approaches for this optimization will be discussed in section 4.2.

4.1 Manual Comparison and Confirmation of Numerically Detected Anomalies

For the manual comparison of the results from the processes 1 and 2 the OptiDrill project partner BGS has developed a Jupyter notebook that generates and plots the illustrations shown in Figure 12 and Figure 13. On the left we can see the results from the automatic anomaly detection algorithm described in chapter 2. The illustration on the right is generated using the information found in the respective Gantt chart described in chapter 3. It shows a time versus depth plot with the added annotated activities found in the daily drilling reports. For the Jupyter notebook to work, the information about the activities carried

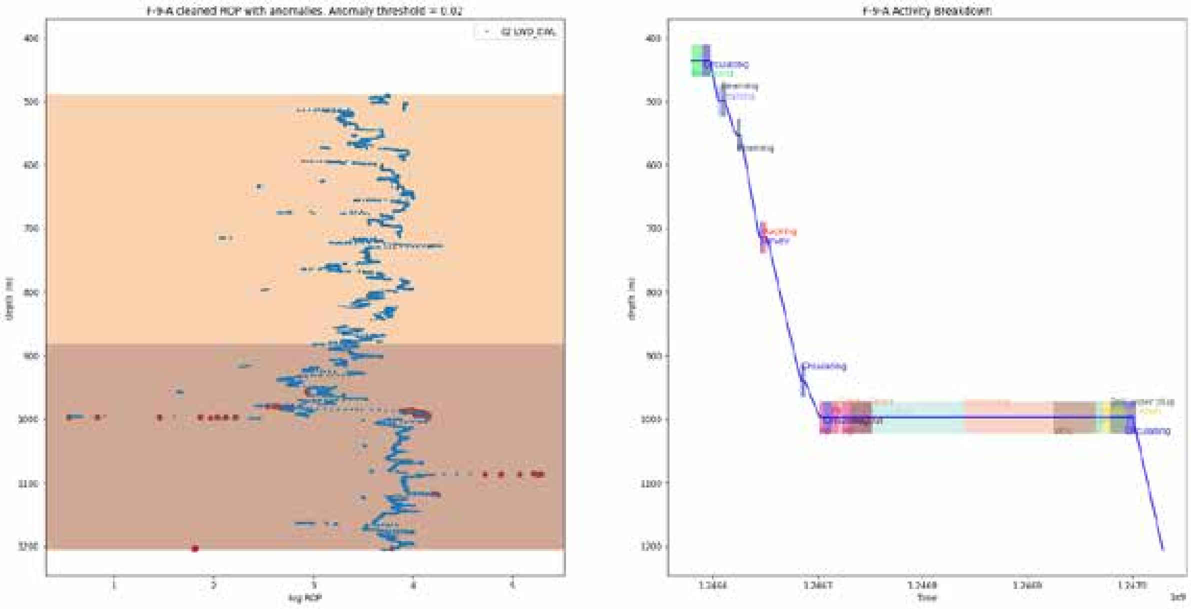

Figure 13: Comparison of anomaly detection results with the activities shown in the Gantt chart for well 15/9- F-9 A

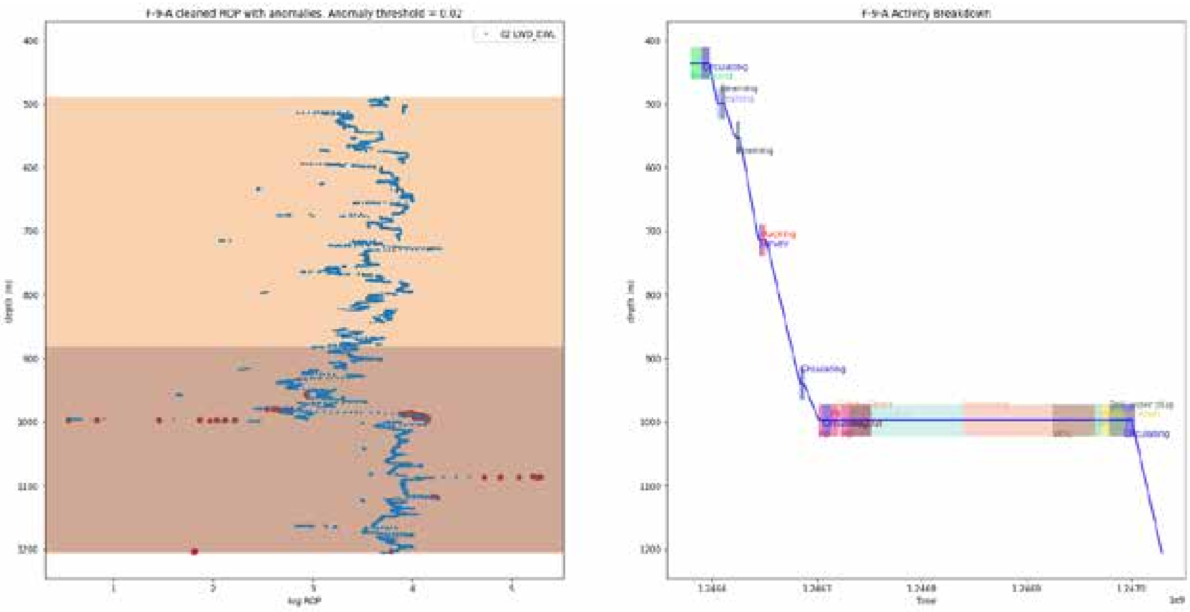

Figure 12: Comparison of anomaly detection results with the activities shown in the Gantt chart for well 15/9-F-14

out has to be extracted from the Gantt chart and presented in an excel file, which is not further described at this point.

The results shown in Figure 12 are obtained by analysing the data from well 15/9-F-9 A and the results in Figure 13 are derived from well 15/9-F-14.

As mentioned before the final step of correlating the data from the daily drilling reports and the additional what if scenario data, if provided for the respective datasets, with the numerical data and the marked anomalies will be a manual step that will require a lot of time and effort. At this point the anomalies found in the numerical data have to be checked using the depth and if available also the timestamp of the respective data point and compared with the data available in the well's documentation. The Gantt chart and the activity breakdown generated will serve as some kind of look-up table to help with the whole correlation and comparison process. At the end the anomalies found by the isolation forest algorithm have to be classified into correctly or incorrectly detected anomalies through this correlation. Correctly detected anomalies would be any anomalies detected that can actually be correlated to any problematic drilling event that happened at the depth of the anomaly. Incorrectly detected anomalies would be data points that have been classified by the algorithm as anomalous data points but can not be correlated to any events from the documentation at hand.

At the end of this step we would ideally have a table with the data points remaining that were classified as correctly detected anomalies. This table would include an extra column describing the type of anomaly, the event or problematic scenario that caused it to happen and any additional information that could be interesting. Having this information describing the causes and the type of the event that led to the anomaly will enable us to create the according labels for all the data points available.

4.2 Quantified and Annotated Drilling Problem Dataset

The final product of the processes described in the previous chapters will be a drilling problematic scenario dataset consisting of all of the drilling process parameters available for the respective well and labels describing the class of each data point. These classes will most likely be string values naming the type of event that occurred at the given data point, such as 'drilling' for normal drilling operation without any problems or 'lost circulation', 'pipe sticking' or 'kick' for problematic events just to name a few.

In order to be able to quantify the drilling datasets and generate new datasets in numbers that can be used for the further development steps to reach the objective of work package number 9, the processes 1 and 2 described in the chapters 2 and 3 have to be optimized. Process number 3 will most likely never be automated due to the human supervision and expert knowledge required.

Process 1 which is the drilling documentation expert analysis, offer huge potential for automation. The task manually of browsing through numerous daily drilling reports, which can easily sum up to far over one hundred per well, in order to extract interesting information about the drilling process and activities from it is very time consuming. Doing this manually for a few wells is fine but thinking about doing this for a couple of dozen or even a few hundred wells will pretty quickly get too expensive and soon use up all the time resources available. Automating this task using scripts for the processing of the daily drilling reports in order to quickly and efficiently extract the most important information from them will be necessary and one of the next steps in terms of optimization. First test using available python libraries for the extraction of text from PDFs file have been carried out and will be further elaborated. The biggest challenge in automatically processing the daily drilling reports is their inconsistent format. Depending on the company that did the reports and the time they were written they can have varying formats and quality. In addition, automating the subsequent step of generating the Gantt charts, after being able to automatically process the daily drilling reports would be a huge improvement.

Process 2 which is the drilling dataset automatic anomaly detection also offers some opportunities for improvement. Since the uni-variate version of the algorithm will not be sufficient to detect all events of interest within the data, the multi-variate version will most likely be used for the anomaly detection. An import decision here will be the choice of process parameters that are used as inputs for the anomaly detection algorithm. Finding the most suitable and important parameters for the reliable detection of anomalies will be another challenge. At this point using some feature importance measuring functionalities could help to solve this issue. Running an anomaly detection with a larger set of parameters available to get some preliminary results and then trying to rate the importance of all the parameters used using the random forest algorithm could be an option to try out. Apart from that instead of feeding all data available to the anomaly detection algorithm it could be a better option to feed the data into the algorithm in subgroups divided by bit diameters or formation. This could also lead to better and more reliable results in the detection of anomalies and will be tested in the future.

5. Conclusion

The processes and workflows elaborated and described in this deliverable aim at enabling the creation of development datasets that can be used for the training of a drilling problematic scenario detection and prediction model based on artificial intelligence methods. The subprocesses presented have been tested on a small scale on two wells from the Equinor Volve field. The results show that the approach followed is reasonable and could be optimized and scaled up to be efficiently used even on larger amounts of data.

The main developments that have to be made to achieve this are discussed in section 4.2 and aim at both the expert analysis of the drilling documentation and the automatic anomaly detection. Automation will play an important role in the processing of drilling documentation for the extraction of the information needed for the dataset generation. Apart from that tuning the algorithm and its implementation for the automatic anomaly detection will also be crucial for the success of the dataset generation.

6. REFERENCES

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth ieee international conference on data mining* (pp. 413-422). IEEE.