Master of Science in Petroleum and Mining Engineering
A.a. 2021/2022

Master Thesis

Use of Machine Learning to Optimize Drilling Equipment Performance and Life Cycle
Focus: Predictive Maintenance

Supervisors:

Prof. Marilena CARDU
Dott. Francesco CURINA

Candidate:

Michel KATTAR

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Marilena Cardu, for her continuous support, insightful comments, helpful information, practical advice and unceasing ideas that have helped me tremendously at all times in my research and in the realization of this project.

A second appreciation would go to my company advisor Dott. Francesco Curina who always pushed me into achieving more in this never-ending learning journey. In addition to the whole team at Drillmec, especially Mr. Elia Abdo and Mr Hadi Mustapha who helped me and taught me a lot throughout my work and supported me till the end.

Furthermore, I would like to thank the reader, Dr. Jean Claude Assaf, for his precious time given for the evaluation of my project. I am also grateful for the presence of the jury members during the presentation of my project.

Moreover, I would like to thank the previous and the current head of the department, Dr. Marwan Azzi, Dr. Carole Aghnatios, for their support throughout my years in the Chemical engineering department. Additionally, a special thanks to Dr. Rany Rizk, director of the faculty of engineering in the Lebanese university branch II, as well as to the mobility office as a whole for providing us with the unique opportunity to perform a double degree program in Europe’s most prestigious Engineering schools.

Finally I would like to thank my family and friends who were really mu biggest support system throughout all my studies. I am grateful to have you by my side no matter the distances between us and always helping me in overcoming any difficulty. A special thank you goes to my father who passed away during this year, he always pushed me to give my best and to be one of the best and taught me that if I put something in my mind and work hard for it, eventually it will happen.

Table of Contents

List of Figures

Abstract

The need for a safe, reliable and cost-effective production requires early and accurate fault detection to increase performance and reach an efficient maintenance strategy. Moreover, in strategic sectors such as the oil and gas industry, being able to predict a failure before its occurrence is of high importance. By doing that, equipment lifetime can be extended and unplanned operations shutdowns can be reduced leading to a huge cut in the cost of breakdowns and maintenance. This project talks about how the creation of a predictive model was established to know the number of low-pressure alarms of the blower ahead of time. It also talks about the methodology followed to perform such a thing as well as how KPIs for the heat exchanger, the hydraulic pumps, the filter, and the lube oil were tried to be found. All that happened by first explaining different concepts used in this work such as the effect of the evolution of the industry on production, the Internet of Things, the types of maintenance, the equipment's lifetime as well as the relationship between machine learning and predictive maintenance and how a predictive maintenance program can be set. In addition to that how the data was collected using ProRig and how it was cleaned and then implemented. The model for the prediction of the number of low-pressure alarms gave a good but not so high accuracy which meant that it could still be improved. Concerning the KPIs, it wasn't possible with the collected parameters to find any, but a lot of suggestions for future work were discussed to help in getting KPIs as well as having a better model.

1 Introduction

The evolution of technology is very quick especially in the fourth industrial revolution due to the increase in the use of the Internet and the Internet of Things. Industries are trying to take advantage of that to optimize their production lines while reducing their costs. Furthermore, they are attempting to do that while keeping a safe and sustainable work environment. Focusing on making a good maintenance strategy is a very good place to meet the industries' goals. A proper maintenance strategy helps not only in improving general equipment health status but also reducing equipment failure frequency, all of that leads to a minimization in maintenance costs and maximization of the equipment's useful life [1]. In [2] they mentioned that by using digital maintenance and reliability transformation in heavy industries, asset availability can be increased by 5 to 15 % and maintenance costs can be reduced by 18 to 25 %.

Predictive maintenance is becoming one of nowadays' hot topics, taking advantage of the reduction in expenses of equipment needed for the monitoring of machines, collecting data, and storing huge amounts of it. This maintenance technique by integrating physical and digital systems can produce an optimum maintenance schedule. Its main goal is to predict when a machinery's failure could occur which helps the maintenance team in solving the issue proactively. With the use of machine learning algorithms, correlations and hidden patterns within complex and huge amounts of data can be found [3].

The objective of this work is to generate a predictive model able to predict the number of low-pressure alarms of the blower in the top drive system after having a lot of these alarms triggered in a short period. Moreover, the other goals are to try and find key performance indicators (KPIs) for the heat exchanger, the hydraulic pumps, and the filter of the hydraulic power unit, in addition to the lube oil quality of the top drive system. These KPIs have the role of helping in the monitoring of this component to assess their performance and predict their failures.

Along this project, the evolution of the industry from 1.0 to 4.0 will be discussed. Plus some concepts that are in relation to this work will be explained such as an Internet of Things system, the different maintenance types, the equipment's life, the setup of a predictive maintenance program, the relation between machine learning and predictive maintenance as well as some machine learning techniques used for the model's creation. The methodology of work, from how the data was collected, cleaned, and used for the generation of the predictive model and the KPIs will also be showcased. Finally, all the results with suggestions for improvement in terms of future works will also be explained.

2 Evolution of the Industry 1.0 to 4.0

The industry is in a continuous state of evolution and change. Historically, there are four phases of the industrial revolution: three of them have been completed and the fourth one, Industry 4.0, is currently being experienced. This paragraph will discuss each revolution alone, together with its effect on the industry, as seen in Figure 1.

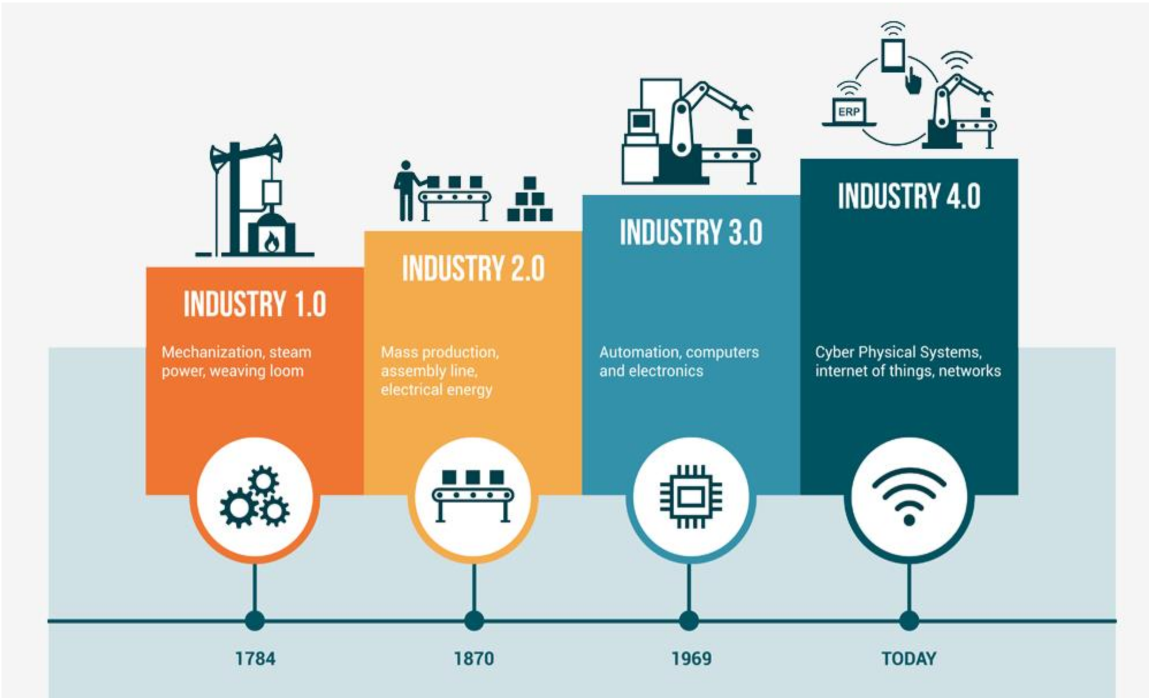

Figure 1: Evolution of the industry from 1.0 to 4.0 [4].

2.1 Industry 1.0

The first industrial revolution started around 1780 in England and then spread to other countries in the Nineteenth century. With the development of James Watt's steam engine which was powered by coal, factories were no longer tied to natural sources of energy, such as windmills or watermills. This led to the transition from individual cottage owners, agriculture, and handcrafts to the world of machines where organizations with owners, managers, and employees were needed to serve customers. This first industrial revolution increased productivity significantly and made it easier while creating a lot of job opportunities and making the transportation of goods faster and cheaper [4].

2.2 Industry 2.0

With the introduction of electricity towards the end of the Nineteenth century, the second industrial revolution began, where electrically powered machines replaced steam engines. Alongside this transition, Henry Ford's invention of the assembly line in 1913 made serial production possible in record time. In addition, the evolution of telecommunications, from letters to telephone calls and telegrams, has also played an important role in increasing production [4].

2.3 Industry 3.0

The third industrial revolution began in the last few decades of the Twentieth century. With the invention of the Z4, the first computer for commercial use, and the manufacturing of electronic devices like transistors and integrated chips, the transition from humble and completely analogue human activities to digitization and automation has become a reality. In this period, software systems were developed to capitalize on electronic hardware and two digital giants were born: Microsoft and Apple. The rise of automated machines and computers during this third revolution has made life faster and easier, but it has led to mass human unemployment, and to the fear that machines will replace humans [4].

2.4 Industry 4.0

Industry 4.0 builds on the development of the previous revolutions. It started in the early 21^{st} century and is still evolving. The introduction of the Internet and the success of the Internet of Things (IoT), which enabled information to be shared, analyzed, and used to drive intelligent actions, made digitization faster and better than Industry 3.0.

The devices are now able to perform tasks fully automatically with the help of the digital network, which offers countless possibilities. Through networking, it is possible to build an intelligent factory where employees, systems, and components communicate with each other for more efficient production.

Due to this huge and rapid growth in digitization, companies cannot produce goods in stock anymore, but must follow actual demands which changed completely the world of production. This was only possible because of the continuous expansion of information processing. This Industry, 4.0, has developed faster compared to the others' annual intervals and, thanks to it, new terms such as robotics, artificial intelligence (AI), IT, clouds, self-learning algorithms and big data have emerged [4].

3 Internet of Things

3.1 What is the Internet of Things?

The Internet of Things is a system of interconnected devices, mechanical, digital machines, computing devices, objects or even humans and animals that have unique identifiers. These devices are called things, are assigned an Internet Protocol address and can transfer data and communicate with each other over closed internet connections or even across different networking types, creating a much more connected world [5] [6].
The addition of sensors to these things and their ability to communicate real-time data without human intervention, provide these devices with digital intelligence and transform the world into a smarter and more responsive place.
Organizations around the world are increasingly using this technology in a variety of industries to enable them to perform more efficiently, learn more about their customers' needs, provide them with enhanced services, improve decision making and increase their businesses' value [5].

3.2 History of Internet of Things

The idea of connected devices started around the year 1970 under the names of embedded internet and pervasive computing, but the first to mention the term "Internet of Things" was Kevin Ashton, co-founder of the Auto-ID Center at the Massachusetts Institute of Technology (MIT): that was the name of his presentation in 1999 in which he wanted to bring radio frequency ID (RFID) to the attention of Protector and Gamble's attention by incorporating the term Internet that was the new cool trend at that time. Also in the same year, MIT professor Neil Gershenfeld in his book "When Things Start to Think" gave a clear insight into where this technology was headed.
The first application of the Internet of Things dates to the 1980s when some programmers at Carnegie Mellon University were able to check if the Coke machine contained cold drinks when they decided to go for a refreshing beverage. Apart from other small projects similar to this one, progress was very slow due to the lack of the necessary technology at the time. The adoption of RFID tags, low-power chips able to communicate wirelessly, and the addition of equipment components to track their location was one of the first applications of the Internet of Things. With that, the use of IPv6, the growth of broadband internet, cellular and wireless networks, and the huge decrease in the cost of sensors and the cost of adding the Internet to objects made the Internet of Things not only used for business and manufacturing, but also for people filling their houses and offices with smart devices that made life easier [5] [6].

3.3 How Does It Work?

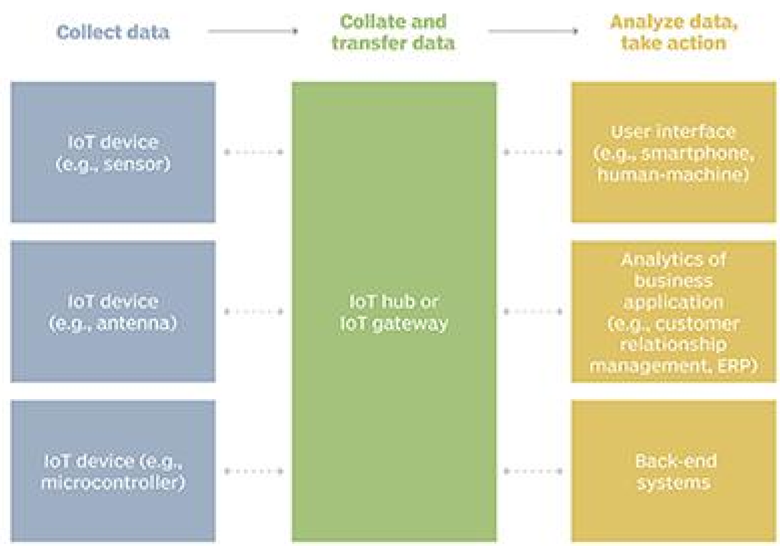

Figure 2: Internet of Things workflow [5]

Figure 2 above shows how an IoT system works and will be explained in this paragraph. An Internet of Things system is made of four main components: sensors and actuators, connectivity, data processing, and user interface. First of all, the sensors are responsible for collecting data from the object (e.g. temperature) while actuators make the system perform a certain action (e.g. switch off the heater). Second, connectivity is where the aforementioned components are connected to a gateway typically wirelessly, but a wired connection could also be used. This gateway transforms the message into an understandable format and uploads it to the cloud on the internet. Now in the cloud, which is a network of computers on the Internet, is where the data storing and analysis is done by using simple rules or more complex ones with the aid of machine learning and artificial intelligence to take smart decisions. Lastly, in the user interface the information is made useful for the end user, for example by sending a notification to his phone about the temperature in the storage room, and then, depending on the application, the end user might act setting the temperature manually, or an automatic response from the system might be triggered like turning on or off the heater to set back the temperature to the defined and acceptable range.

Then, to recap in an Internet of Things system, sensors collect and send data to the cloud via a gateway, this data is then stored, processed, and analyzed by software to make good use of it and take smart decisions and perform actions on the system by using actuators. The person can interact with the system thanks to the user interface that allows him to act and override it, so that his orders are sent in the opposite direction, from the user interface to the devices where the actuators do the work for him [7]. This concept is of high importance for this study since the data collected and analyzed to perform it was done through an IoT system, the ProRig, which will be discussed in paragraph 8.1.

4 Maintenance Types

Maintenance is the ensemble of processes and practices performed on different equipment to guarantee its efficient and continuous operation. This means that by using the best maintenance plan for the plant or industry, the longevity of assets could be increased, machines' performance optimized, unscheduled downtimes avoided, and costs minimized. There are two main types of maintenance, reactive maintenance, and preventive maintenance, very different from each other, each with its pros and cons, and a better place to be used.

4.1 Reactive Maintenance

Reactive maintenance, also referred to as run-to-failure maintenance is, as its name indicates, the repair or replacement of equipment or equipment component after its failure has occurred. The tasks are carried out with priority, from the one that affects safety and production the most to the one that affects them the least. This is usually low-cost maintenance as it could be done with fewer resources and maintenance infrastructure, but the problem is that in the long run, catastrophic events can occur, which become very expensive and the average time to repair will increase. Furthermore, reactive maintenance does not fix the cause of the problem and this means that the mean time between failures will decrease and that will cause the same problem to be repeated several times in a decreasing time interval [8].

Reactive maintenance could be broken down into four different types of responsive strategies:

1. Breakdown maintenance, takes place when a piece of equipment won't start or operate. Usually, it is an unplanned event, and the apparatus must be quickly fixed to resume production. If any delay occurs, the bottom line of the business will be affected, so it is always advisable to have backup equipment for the most critical ones. Sometimes, a planned breakdown could occur for which workers are ready to respond [9].

2. The run-to-fail maintenance, is letting the machine work until it breaks down after that the maintenance operation is performed. This type is considered a planned one since the decision of letting the component fail is deliberately taken and often a plan is put ahead to repair the item without interrupting the production. Furthermore, the equipment that is permitted to fail does not cause health or safety risks to the operators [9].

3. Corrective maintenance is to carry out maintenance operations on malfunctioning equipment before its total failure, so before causing serious problems to the system. Generally, this malfunctioning or potential failure-causing area is detected while fixing other problems [9].

4. Emergency maintenance is never planned for: it is the quick response to an unforeseen accident and it could cause health and safety problems if the situation is not addressed properly. For this type, there is usually an excavation plan, for which the workers are planned to perform if any dangerous event occurs [9].

Even though reactive maintenance is the most basic one, it still has some pros and cons. Some of its pros are that it needs no planning or initial cost, less or no manpower is needed since there are no frequent checkups and it could be outsourced and excluded from the skill set on the payroll; moreover, it is more cost effective on non-critical and low impact equipment. On the other hand, accurate budgets for it are difficult to make and unplanned downtime puts production schedules at risk and alters the supply chain. This reduces profit, due to the absence of operators and spare tools and more delays will occur to bring them to the production site. That will cause disorganized work and distracts the team; in addition to all these, the most important problem is that safety risk is high especially that maintenance workers need to solve the problem quickly to minimize the downtime [10].

4.2 Preventive Maintenance

Preventive maintenance has the goal of taking the required actions and precautions to avert equipment failures from happening and minimize accidents risk [8]. Here there are two maintenance activities, nonintervention activities, and intervention activities. In nonintervention activities, the equipment is monitored by maintenance teams following maintenance rounds which consist of predetermined routes within a specific area of the plant with a well-known number of equipment and systems to be inspected to assess their performance and to see whether any signs of degradation and variance from normal operating conditions are present. Nonintervention tasks could also comprise scheduled legislative inspections set by the government. Intervention activities, they consist of repair, replacement, and intrusive servicing of the equipment. These types of activities often need the system to be shut down so that additional inspections can be made to detect other possible failures and address them. Sometimes, if many maintenance activities are needed, the stopping of production is necessary, so that they can be assembled and performed in a preplanned facility shutdown not to affect the supply chain program or even the production plan [8].
Under the umbrella of preventive maintenance, four types of maintenance could be found, which are:

1. Time-based maintenance: it consists of periodic cleaning and checking of the equipment to monitor its condition, which could be every month or annual basis depending on the manufacturer's recommendations. Even if there were no recommendations, any business should always try to inspect the most critical machines regularly, to avoid any major incidents. This means that when a maintenance schedule is being prepared, the focus should be on the major utilities, equipment, tools, and technology that the company relies on to succeed. Furthermore, to better plan time-based maintenance, detailed notes about past failures or problems should be kept so that systems and equipment that need additional care are detected. An example of periodic maintenance is air conditioner servicing after a couple of months after summer [11].

2. Usage-based or meter-based maintenance is done for machinery that works daily, so tracking its usage, like operating hours and production cycles, particularly on systems that don't give notifications on the number of operating hours is critical because, by doing that, operators will stay on top of proper care and will guarantee a prolonged use of critical equipment [11]. The most common example of this type of preventive maintenance is the oil change in a car after a certain mileage has been reached [12].

3. Predictive maintenance's goal is to reduce the number of necessary planned repairs and it consists of an advanced form of preventive maintenance [12]. To do that, predictive maintenance captures equipment conditions through sensors that gather data, by doing that it can track the machine's condition and detect any abnormal patterns so that it would be able to predict what failure and when could occur in the future and a maintenance schedule could be set [11]. This means that predictive maintenance is based on monitoring the condition of equipment and having any equipment undergo a certain cycle of failure. Hence, the faster a failure is perceived, the more time the maintenance team has to choose the best intervention plan to avoid the problem without interrupting the production. This was the goal of this work, predicting the failure of machines to try and perform a predictive maintenance program in the future to avoid failures from reoccurring. A good example of predictive maintenance is that if a certain car owner goes off-roading in the summer, the oil change should occur every 3000 miles instead of the normal 5000 miles in the winter [12].

4. Prescriptive maintenance is very similar to predictive maintenance, but it uses advanced analytics, machine learning, and artificial intelligence to predict and suggest the best maintenance strategy [11]. What happens is that the software gathers data about equipment's status and analyzes it to give specific recommendations to avoid any risk. For instance, in the previous example of predictive maintenance, by using a prescriptive maintenance approach to the same model, the software would advise oil to be changed

every 4000 miles in the summer instead of the 3000 of the predictive maintenance model, to avoid dusty filters [12].

Similarly to reactive maintenance, preventive maintenance also has many advantages and disadvantages. First the advantages:

- Increase in productivity, as poor maintenance strategies reduce productivity by 20%, operational managers through preventive maintenance digitize the details of essential equipment, assign recurring work orders, review asset history from their smart devices and reduce business downtime caused by unexpected failures [11].

- Longer equipment lifespan, as badly performing parts are fixed regularly through schedules that ensure a steady performance of the assets according to manufacturers and consumers guidelines, the frequency of capital expenditures to buy new equipment is reduced [12].

- Reduced costs, unplanned downtimes, and running a machine until it fails may cost ten times more than periodically checking, monitoring, and fixing it. Thus, companies that use preventive maintenance have a higher output than the ones that adopt reactive maintenance as they experience fewer breakdowns [12]. In addition to that, due to the planning done and the reduction of unforeseen downtimes, the amount of money paid on overtime for workers to fix problems is reduced [11].

- Enhanced safety: due to regular machinery checks, the occurrence of hazardous problems is reduced, which leads to the reduction of unpredicted downtimes, health h,azards and liability lawsuits [12].

- Less energy consumption and reduction of utility bills, because weakly maintained equipment consumes more energy than those normally operating and that means that energy-robbing parts will be rapidly fixed [12].

After having examined the five very beneficial advantages of preventive maintenance, here are some of its disadvantages:

- Budgetary constraints, with the high prices of the software used in this field and the needed man-hours to fulfill the required job, smaller businesses can't afford this digitally advanced maintenance solution, even with providers who try to make it more accessible [12].

- More required resources: with the increase of the demanded tasks throughout the year due to preventive maintenance, there will be an increase in the entailed personnel, parts, and monthly capital [12]. Furthermore, some employees would have to work overtime or be removed from their daily tasks to perform preventive maintenance assignments [11].

- Time consuming: since more work needs to be done to set up a schedule for the periodic maintenance duties and assets, inspection becomes more complicated with a huge number of parts taking more time [12].

- Difficult to organize, with the high number of equipment to be checked in a company: organizing the work to be done may be very difficult to prepare for and, if not done correctly, may lead to a disaster [12]. This problem will be even greater if no past planning has been done or if there have been no previous records of maintenance activities [11].

- Excess of preventive maintenance leads to spending huge quantities of money on unnecessary maintenance tasks [11].

5 Equipment Life

An equipment's life can be described as physical life, profit life, and economic life, where the last two should be defined and calculated to make an equipment replacement decision since they give two important means to address the replacement analysis. In addition, obsolescence, inflation, maintenance and repairs, depreciation, downtime and investment are all fundamental to replacement analysis. Thus, by combining all these concepts, an equipment manager can make a good equipment replacement decision [13].

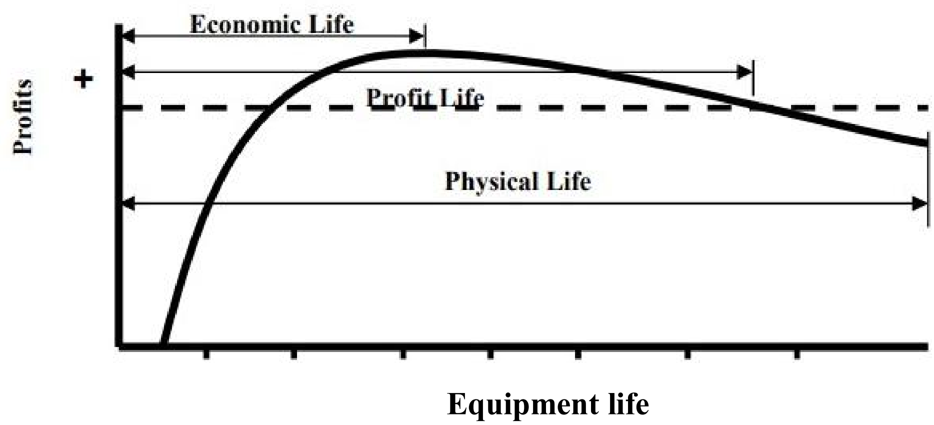

Figure 3: Profits vs. Equipment life [13]

Figure 3 shows the three stages of an equipment's life cycle where it can be seen that in its initial years, the equipment needs some time to earn enough to cover the capital cost of its procurement. After that, it moves to the second phase where it earns more than the cost of operation, ownership, and maintenance. Finally, the end of a machine's life is when keeping it working and the production time lost due to maintenance cost more than what it earns while normally operating. Thus, an equipment fleet manager requires the necessary tools to capture that time when the asset is no more profitable and needs replacement [13].

5.1 Physical Life

The physical life of equipment is equivalent to its service life, and it ends when it's unable to operate anymore [14]. It is mainly affected by the maintenance and repairs done on this machine during its lifespan. This means that an apparatus that was given less attention during its operation period will break down earlier than one that was under preventive maintenance. Thus, a piece of equipment's physical life will be impacted by two things, itself and if it was maintained properly [13].

5.2 Profit Life

An equipment's profit life is the period in which the equipment generates profit. That is the most craved life since after it the machine will start operating with a loss [14]. Increasingly costly repairs exacerbate this as major components wear out and need to be replaced. Consequently, this stage needs to be maximized to get the most profit and efficiency out of the asset, and the manager has to know the period where a replacement plan should be implemented for a new machine while the components are still useful [13].

5.3 Economic Life

Economic life is the period where ownership costs decrease while operating costs increase and it continues until the operating costs exceed the ownership costs, which means that the component is costing more to operate than to own [14]. Thus, to have the maximum profit, an equipment should be replaced before the end of its economic life. The proper timing of equipment replacement prevents erosion of profitability by the increased cost of maintenance and operation as the equipment ages beyond its economic life [13].

5.4 Life Cycle Assessment and Costing

The depletion of natural resources became along with climate change global concerns. The industrial sector is currently the largest consumer of energy and resources, and a major contributor to the climate change problems. Thus, a noticeable transition of the industrial sector towards sustainability is needed.

Life cycle assessment (LCA) is a popular tool applied across many industrial sectors to determine the environmental impact linked to a product's lifecycle. Its application to support decision-making in different sectors like supply-chain management, marketing, process optimizations, and eco-design or strategic decisions has been already performed. It is favorable to use it in specific cases to assess a product's lifecycle. This assessment is not done on all the products but only the ones with high criticality and where the manager feels the need to perform a such thing. However, life cycle assessment is rarely applied on industrial manufacturing systems and even less on high energy-consuming equipment to assess their environmental impact. A brief literature review tells that the progress made in life cycle assessment methodologies, is essentially focused on the improvement of individual processes such as impact assessment and data collection. On the other hand, this method should satisfy stakeholders' requirements while balancing cost, time, effectiveness, and expertise for the assessment.

Life cycle costing (LCC) focuses on assessing all costs that occurred during the lifecycle of a product, i.e., from conceptualization, manufacturing and operation to the end of its useful life. It can be modified to give guidelines to manufacturers and clients from their perspectives. In addition to financial representation, life cycle costing also includes integrated analysis comprising economic, social, and environmental aspects.

A combination of environmental and economic aspects gives better support to decision-making procedures in developing products or services. Nevertheless, life cycle assessment and life cycle costing is performed separately and then their results are combined but this leads to a series of problems including inconsistencies, tedious data, and loss of significance. Procedure-oriented models for integrated LCC and LCA by dividing lifecycle into system and subsystem levels have been developed, which enable a consistent and structured handling of partial problems. However, the only problem with these models is their high evaluation procedure complexity [15].

Even though the concepts of equipment's physical, profit, and economic life as well as life cycle assessment and costing where not used in this work, it would be of a great plus if in future works they were taken into consideration since they attack the subject from a more economical and environmental point of view.

6 Predictive Maintenance Program

Suggestions in production and maintenance are that there is a need to change maintenance policies from short-term problem solving to considering long-term goals. Predictive maintenance is a maintenance policy where specific physical variables are selected and associated with sensors on operating machines that measure their values continuously to record the data and use, them later on, to analyze, visualize and obtain important and useful information [16]. This policy uses these data along with other predictive tools like historical data, integrity factors, statistical interference methods, and engineering approaches to detect machine failures from an early time [1]. Predictive maintenance progressed from simple visual inspection to automated methods by taking advantage of the developed signal processing techniques based on pattern recognition and machine learning. These methods by collecting and detecting sensitive information from machines, help by finding the causes of the failures and avoiding unneeded downtime to improve equipment's efficiency. What differentiates this maintenance technique from others that also provide a schedule is that this one makes the schedule based on real-time data [17]. In addition to that predictive maintenance also increase the remaining useful life of machines [1].
To set up a predictive maintenance program it is of high importance to understand the technical peculiarities of the procedures and instrumentations that make the diagnosis possible. Hence the success or failure of that will rely on the knowledge of the program planner about this subject. The activities carried out in a predictive maintenance program are:

1. Data acquisition: it is performed by using routes so that the time dedicated to this activity is minimized. Data must be collected by the same operator and in the same location to minimize the effect of conditions' variables on the measurements [16]. For this project data acquisition was carried out using an IoT system that will be discussed in paragraph 8.1.

2. Information processing: after acquiring the predictive data, they must be sent to a PC where processing is carried out depending on each company's policies; the transfer is done automatically. Normally a graph is created for the visualization of the different parameters and only more advanced analysis will be done if deviations in the behavior appear [16]. These graphs are similar to the ones shown in paragraph 8.2. for the raw data and paragraph 8.3. for the processed data.

3. Drawing up of diagnosis: it is a critical step in the predictive maintenance program and only professionals do it, to avoid the accumulation of unanalyzed data. The analyzers use graphic representation and software to better perform the task. The technician issues the state of the equipment and in case of any problem he provides a

diagnosis. He also has a system that by using some preliminary data knows if there is any failure and how to resolve it [16].

4. Transfer of information to personnel responsible for the maintenance activities: issuing work orders. Generally, the people involved in corrective maintenance are different from those responsible for predictive maintenance, thus the predictive maintenance program needs to work so that this information is sent quickly to the correct people, to avoid further development of the fault [16].

5. Diagnosis verification: after completing the maintenance job, predictive data is acquired again to check that everything is back to normal; in case of further troubles, information has to be collected and analyzed to avoid their repetition in the future [16].

6. Transfer of the information to the computerized maintenance management system and enterprise resource planning software, which is not a common step for all companies. Linking the computerized maintenance management system to the predictive diagnosis software helps in improving the predictive maintenance program. Its integration into the company's maintenance policies facilitates the setup of the maintenance schedule. On the other hand, the enterprise resource planning software helps in taking decisions concerning costs, human and technical resources production scheduling [16].

7. Combination of predictive information with the production process: that is important because it draws up the predictive information to be communicated to the management so that available resources for the predictive maintenance program are increased and maintained. This report should include an evaluation of the overall costs of a predictive maintenance program about its benefits [16].

7 Machine Learning for Predictive Maintenance

7.1 How Are They Linked?

Machine learning is a subcategory of artificial intelligence, it is defined as an algorithm or program that can learn with small or no support at all. Additionally, machine learning techniques solve many different problems and extract knowledge out of existing data. The required technologies and processes to perform predictive maintenance are provided in Figure 4.

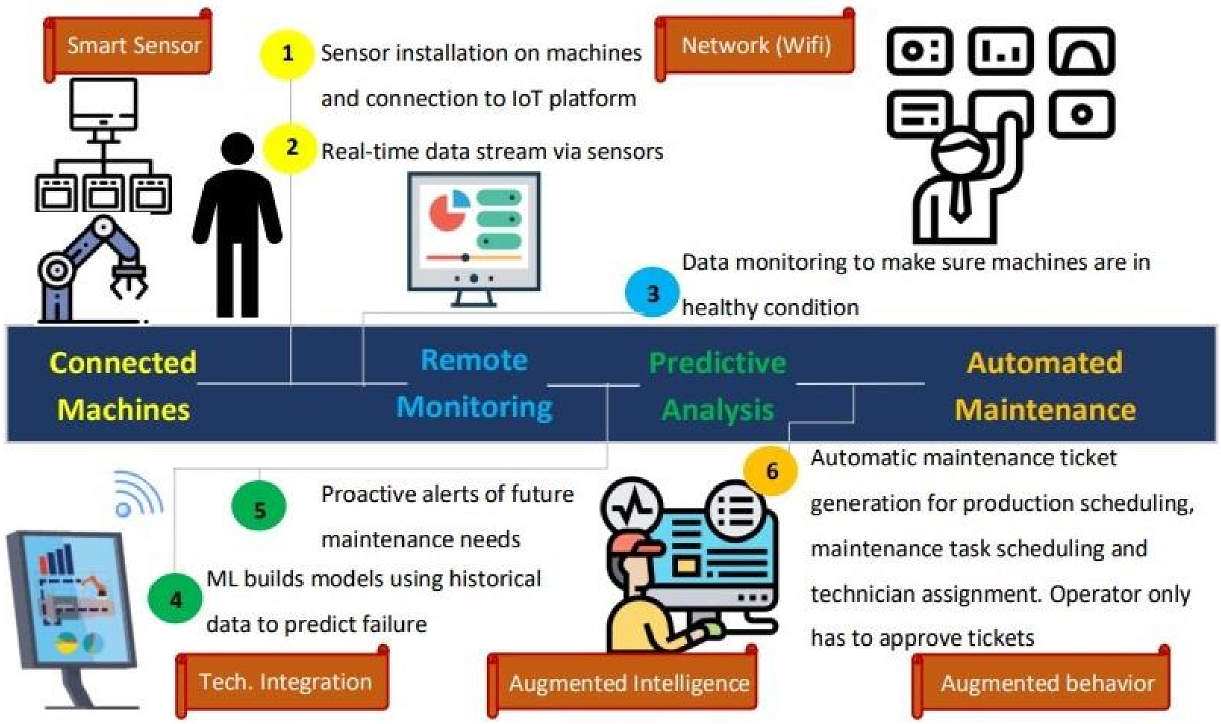

Figure 4: Process and technologies to drive predictive maintenance. [18]

The five different categories that drive predictive maintenance are smart sensors, network, integration, augmented intelligence, and augmented behavior. Smart sensors collect the needed data and information from the equipment and these are built-in sensors or environmental data with the use of external sensors. The network is responsible for storing this data and transferring it via Wi-Fi or Bluetooth. After that, the technology integration helps in the management and accumulation of data via Internet of Things. Augmented intelligence assists in data processing and analytics, while augmented behavior grants the visualization, computing and service platform through apps and tickets to help the operators. These technologies were used in this study to build a model that predicts failures. In case of high accuracy, it will be integrated to perform predictive maintenance after some

testing.
Machine Learning is used in computer science and other areas like predictive maintenance of machines, tools and equipment because it can solve problems using the huge data generated by industries. Multiple techniques can be combined to have the most efficient model [18]. The steps required to build a machine learning model are the historical data selection step, data pre-processing step, model selection, model training and model validation step, and finally model maintenance step, as shown in Figure 5.

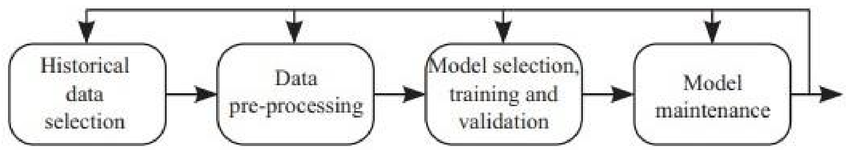

Figure 5: Main steps for the design of a machine learning model [1].

The historical data selection step has the role of deciding how and which data should be collected, so it is valuable for the model design, this step in predictive maintenance is the data acquisition step as will be seen in paragraphs 8.1 and 8.2. The next step is data preprocessing and, as the name itself says, is the one in which all the transformations and efficient data processing by the model are performed; it includes data cleanup, such as handling missing data and removing outliers as done in paragraph 8.3. In predictive maintenance, the data is analyzed in this step, to understand and have a better interpretation. After data processing, there is the phase of model selection, training, and validation. During this phase, the appropriate model or models are selected, then trained and validated to verify their accuracy towards the desired goal. The equivalent of this step in predictive maintenance is the maintenance decision-making step, where the best algorithm is chosen to be applied as done in paragraph 9. The machine learning techniques used in this study will be discussed in paragraph 7.2. Finally, the last step is model maintenance, to make sure that the model is still efficient over time because industrial activities may change and could lead to a decrease in the model's performance [1].

7.2 Machine Learning Techniques Used

To train the predictive model in this study, five machine learning techniques were used thanks to the powerful ThingWorx software that helps in running multiple tests using different combinations of techniques in a short amount of time and without the need of writing complex algorithms. The following techniques will be discussed: artificial neural networks, gradient boost, linear regression, decision tree, and ensemble.

7.2.1 Artificial Neural Network

The artificial neural network model was developed using a biological approach based on the structure and function of the human brain. It is a huge parallel computing system formed by a massive number of interconnected simple processors which can learn the basic laws from the set of given symbolic situations in examples instead of following the set of laws specified by humans [18].

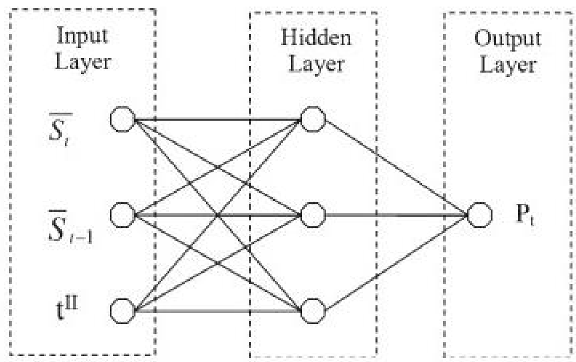

Figure 6: Artificial Neural Network Structure [19].

As seen in Figure 6 above, the structure of an artificial neural network is made up of neurons, which consist of interconnected processing nodes. These neurons are arranged in a sequence of several layers: the input layer, one or more intermediate layers, and the output layer. The input layer acquires the data and is not responsible for any computation. The intermediate layers or the hidden layers are connected to the input and output layers: every neuron in it and the output layer receives signals from all the neurons in the previous layer and performs a weighted summation and transfer function of the input. Finally, the output layer provides the network's answer for the corresponding input [19]. This model is used in diverse fields of study, such as dynamic system modeling, pattern recognition, data classification, prediction, multi-dimension mapping, and novelty detection due to its capability of learning from examples and its ability to find solutions from random, fuzzy, and nonlinear data [18] [20]. An artificial neural network can learn from historical data, experience, and examples to perform modeling, classification prediction, and clustering [20].

7.2.2 Decision Tree

A decision tree is one of the most used machine learning techniques. Its structure resembles a tree with roots branches and leaves. It starts with a root node and along the way the data are grouped into smaller groups and divided into branches. The inner joints of the tree are called nodes and represent the properties, while the branches represent the decision rule and the leaves are the outcomes [21]; an example is provided in Figure 7.

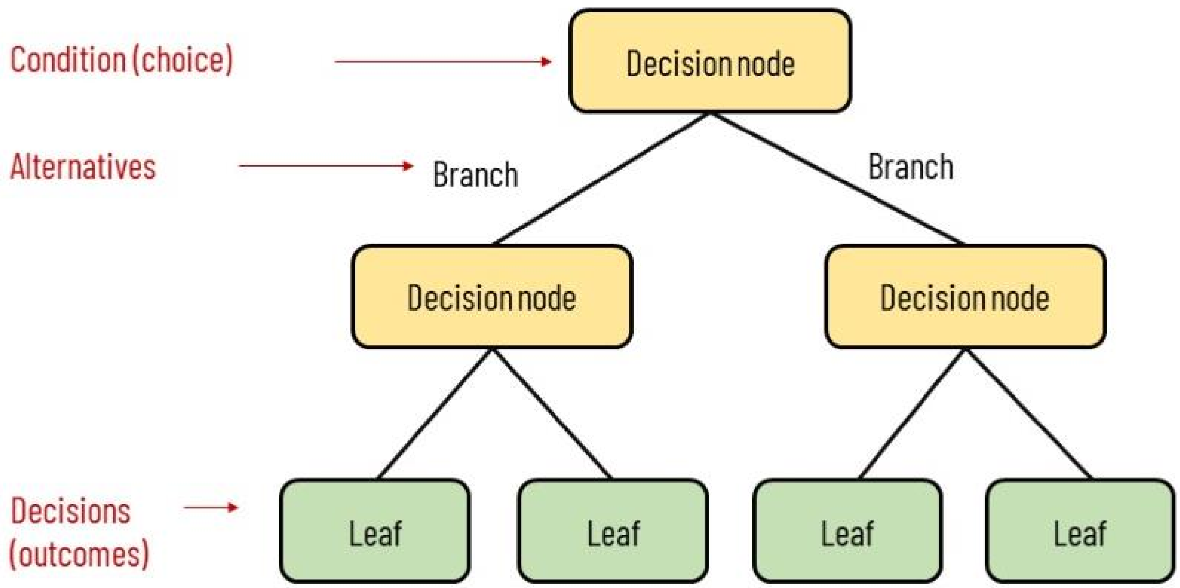

Figure 7: Elements of a Decision Tree [22]

One advantage a decision tree has over other machine learning methods is the interpretability of its model, which helps support the design of future data analysis work by providing information relating on identifying important characteristics and interclass relationships [23]. Moreover, its construction is quick and can be converted into SQL statements: in this way, efficient access to databases is possible [24]. The most optimized algorithm for decision trees is the one that gives the split with the highest information gain [25].

7.2.3 Gradient Boost

The idea behind the gradient boost technique is to develop the weak learner towards the best result by using the gradient direction, as shown in Figure 8.

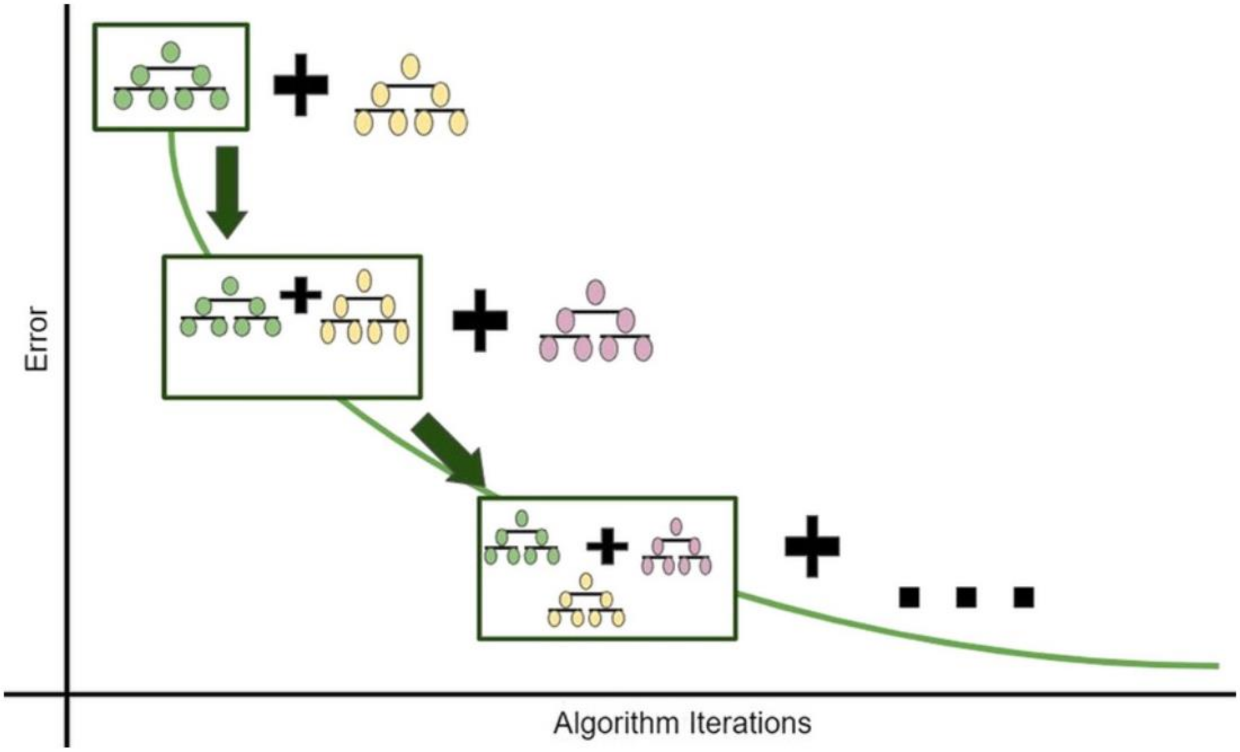

Figure 8: Gradient Boost [26]

The sample is checked at every iteration and the boosting algorithm gives the highest weight to the misclassified samples for it to be prioritized. This process is repeated until the boosting algorithm satisfies the weighted sum of all the ensembles prediction [27] [28]. Gradient boosting is useful in classification problems as well as regression problems [29]. Furthermore, this machine learning technique was successful in many applications and showed the ability to update prediction results on new consecutive models [30].

7.2.4 Linear Regression

Regression analysis is a term used to mean two things: the first is to forecast or predict, and in this area regression models' application overlap with machine learning, the second is to find relations between independent and dependent variables [31]. An example is provided in Figure 9.

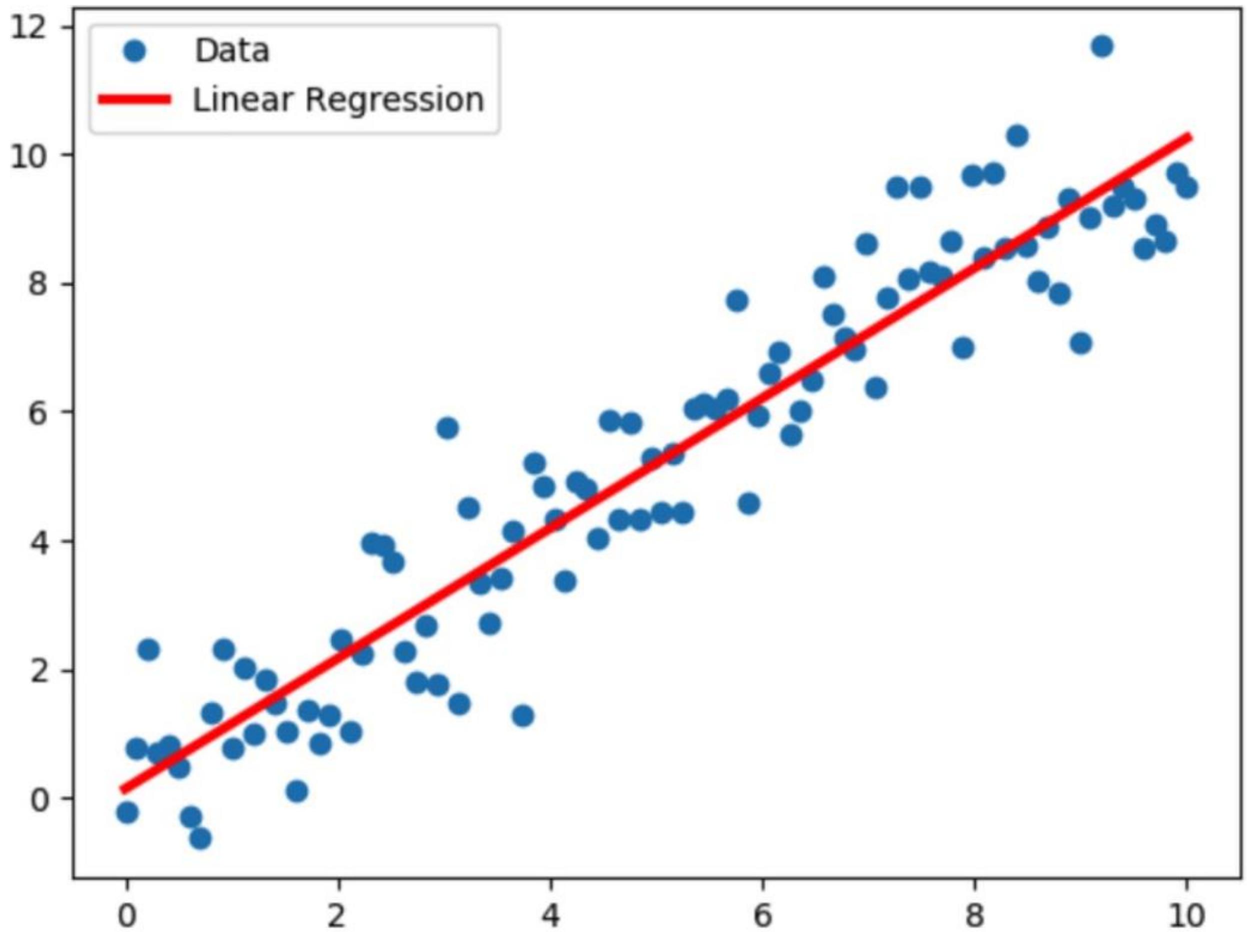

Figure 9: Linear Regression Model [32]

Linear regression refers to a multivariate linear combination of regression coefficients (i. e. constants and weights of input variables). The coefficients are estimated by the generalized least square technique [33]. Linear regression can predict the dependent variable from its correlation with one or more independent variables and to quantify this relationship. Linear regression is a widely used technique since it is simple, and easy to interpret; it could also overcome more sophisticated and complex nonlinear models, especially where training data are limited [34] [35]. The underlying assumption is that the linear model can provide a reasonable approximation to the regression function [35].

7.2.5 <u>Ensemble</u>

The combination of basic independent predictions learners into a meta-predictor is called ensemble learning and usually performs better than a single learner. Some of its declinations are stacking, model averaging, bagging, and boosting, they can be used to combine data from different sources [36]. By performing this combination, a more advanced model is created taking the strengths of the individual ones and this helps in the reduction of the prediction's bias since the dependence on a single model is not present anymore. The merging of the base models is performed using two major methods: stacking and aggregating. The former determines which are the reliable models by using a meta-learner concept, while the latter based on a majority voting scheme combines the individual results. For the meta-model to give the final results, it uses as input the individual models' results. If the target is only getting the best prediction accuracy, ensemble models are always better than single models but this will cause the cost of computation and storage to increase and will make the model less interpretable [37].

Ensemble prediction methods could be divided into two categories: heterogeneous ensemble and homogeneous ensemble, depending on the model generation technique used. In the heterogeneous, also called competitive, ensemble, the model is generated by using the same training dataset on the individual models but by changing the parameters and the final output will be the average of the base models, this method exploits the complementarity between the different learning algorithms. On the other hand, in a homogeneous model, or cooperative model, different datasets resampled from the original one are used with the same parameters to train the models, by dividing the job into multiple subtasks where each one is completed on a single algorithm depending on the nature and characteristics of this subtask and finally the prediction will be the sum of the different subtasks, this method optimizes the performance of the meta-model by training it multiple times with varying data and combining their predictions [38] [39].

Due to its high performance, accuracy, and ability to handle imbalanced and small-sized datasets ensemble technique is used more and more nowadays by researchers in many different fields [40]. Three factors should be considered during the model creation: identifying the best suited prediction algorithms for the application, understanding how many base models are to be used to have better accuracy, and selecting the combination techniques able to provide the better single model outputs merging, to get the best individual prediction [39].

8 Methodology of Work

In this section, the methodology of work will be discussed from data acquisition, processing, and model implementation. The goal is to generate a predictive model for the failure of the blower in the top drive system where low-pressure alarms were triggered. Moreover, to find some key performance indicators (KPIs) for the lube oil in the top drive system and the hydraulic pumps, the heat exchanger, and the hydraulic oil filter in the hydraulic power unit. These KPIs will help in predicting the performance of these components in the future to know when a failure might occur.

8.1 ProRig for Data Acquisition

ProRig is a data management system developed by Drillmec. It is an industrial Internet of Things platform for the monitoring, visualization, and analysis of the data coming from sensors mounted on the rig components in real-time. Each rig is sending data from its subcomponents such as hydraulic power units, mud systems, rotary systems, and electrical systems. The quality and amount of sensors affect the quality and quantity of the received data which in turn has a direct effect on the output of the system and its efficiency. Thus the better the sensors, the better will be the acquired data. This data received from the sensors is still considered raw and can't be used for any purpose unless it is cleaned, filtered, and verified. The platform is not only capable of displaying live data on dashboards with easy interfaces but can also apply calculations and extract important information from it. All the data is stored in the cloud, to which users and clients have direct access from anywhere in the world. The historical data are also accessible, allowing the revision of some operational incidents and the analysis of events or problems.

Several rigs can be connected simultaneously to the system as shown in Figure 10, this helps in comparing their performance and efficiency. Hence, any important feature or analysis extracted from one rig could be applied to improve another, although not all rigs are the same, but potentially capable of sharing similar subcomponents.

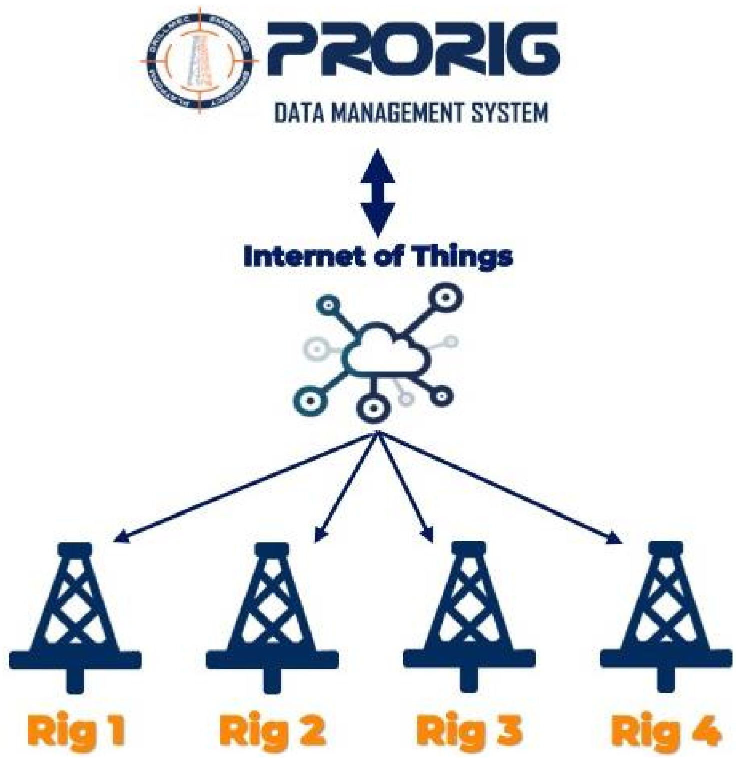

Figure 10: Rig Inter-Learning

Before arriving at the visualization and analysis part where the user interface is also present, the flow of data collected from the sensors on the rig passes through several steps. It first starts from the sensors mounted on the different components of the rig and this data is collected at different frequencies depending on the criticality of measuring a certain parameter every n seconds. After that the rig servers store and keep this data physically on the rig for as long as 30 days if needed, to avoid any loss of data in case of the interruption of the internet connection. If everything is running smoothly the next step will be for the data to be sent to the edge server in the company's headquarters. From there it is transferred and stored in the cloud before arriving to the last step of analysis and visualization, as the following Figure 11 shows.

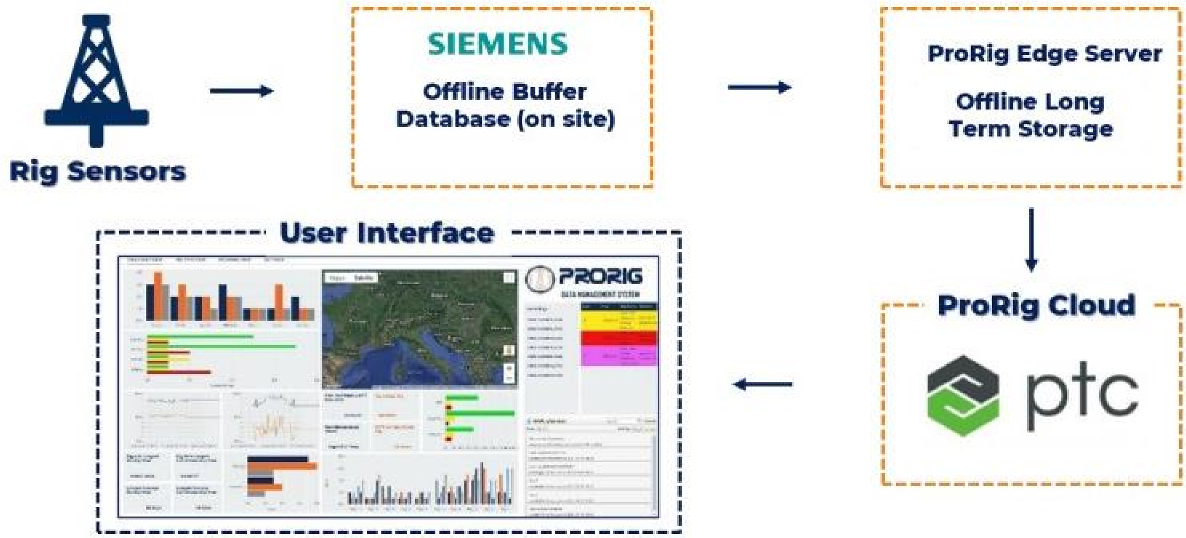

Figure 11: System Architecture.

One major problem that arises from this transfer of data is its security. When it comes to data sharing over a cloud, its transport from source to destination securely is an absolute priority. For this issue, the ProRig platform is equipped with a multilevel security and backup system. One of them is the one mentioned earlier about storing on the rig's servers for up to 30 days. Another one is the encryption of the data offline on these servers which is also re-encrypted when it reaches the online cloud before being sent to the headquarters. The data is transferred by using private VPN tunnels to ensure its safe and secure arrival. All these levels of security are reported in Figure 12.

Figure 12: Data Secure Flow From Source to Destination.

In addition to all these security measures, access to this data is only authorized to specific personnel chosen by the client. From Drillmec's side, only the support engineers that help in the system's development are allowed to have access to it with unique usernames and passwords.

8.2 Raw Data Collected

By using the ProRig data management system the initial raw data from the data acquisition system (DAS), the hydraulic power unit (HPU), and the top drive system (TDS) used in this work were collected and are showcased in this part.

8.2.1 DAS Raw Data

The data acquisition system is the one with the highest frequency of data recording as a new value is recorded every 30 seconds. By using this system which is part of the ProRig, data from the bit depth (Figure 13), well depth (Figure 14), surface revolutions per minute (RPM) (Figure 15), the three mud pumps' strokes per minute (SPM), SPM pump 1 (Figure 16), SPM pump 2 (Figure 17) and SPM pump 3 (Figure 18), the surface torque (Figure 19), the weight on bit (WOB) (Figure 20), the hook load (Figure 21), the average rate of penetration (ROP) of every hour (Figure 22), the maximum ROP of every hour (Figure 23) and the instantaneous ROP (Figure 24) were recorded and collected for a period of one month. In the following plots, each parameter's raw data will be displayed before any cleaning or filtering has occurred.

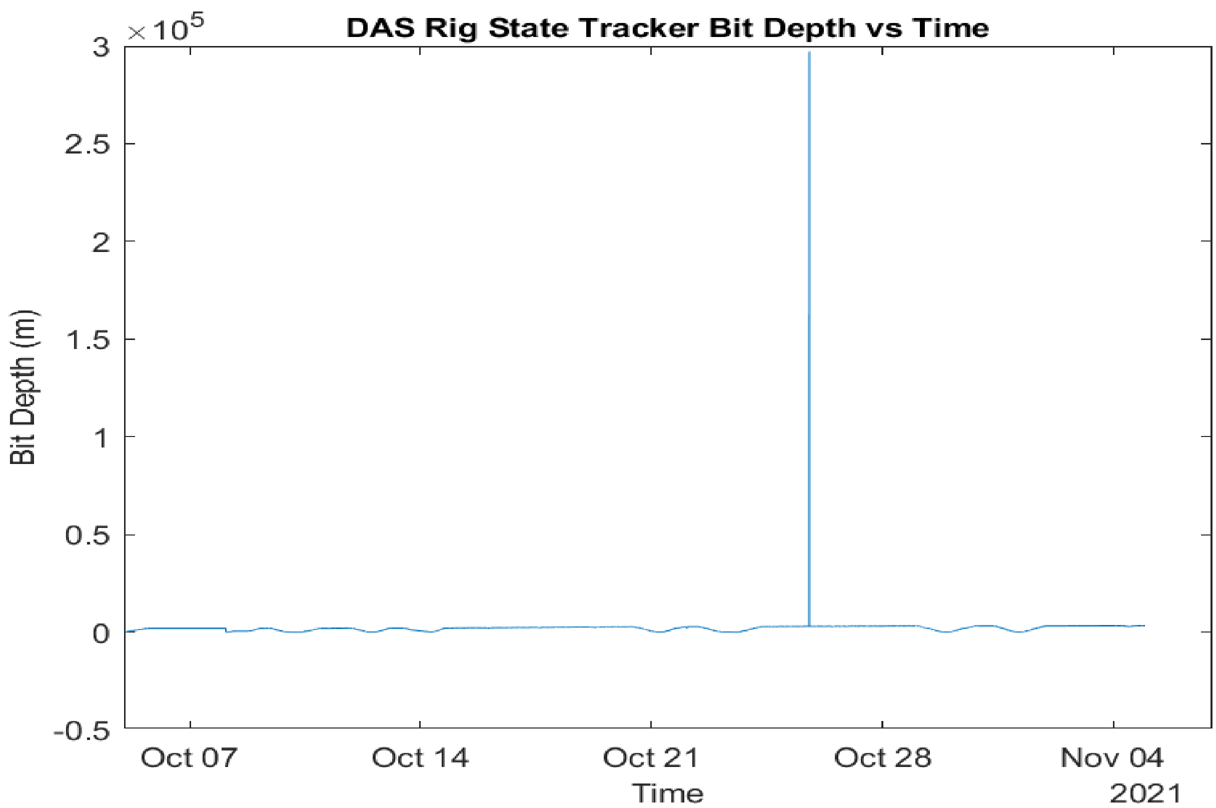

Figure 13: Bit Depth vs. Time (Raw Data)

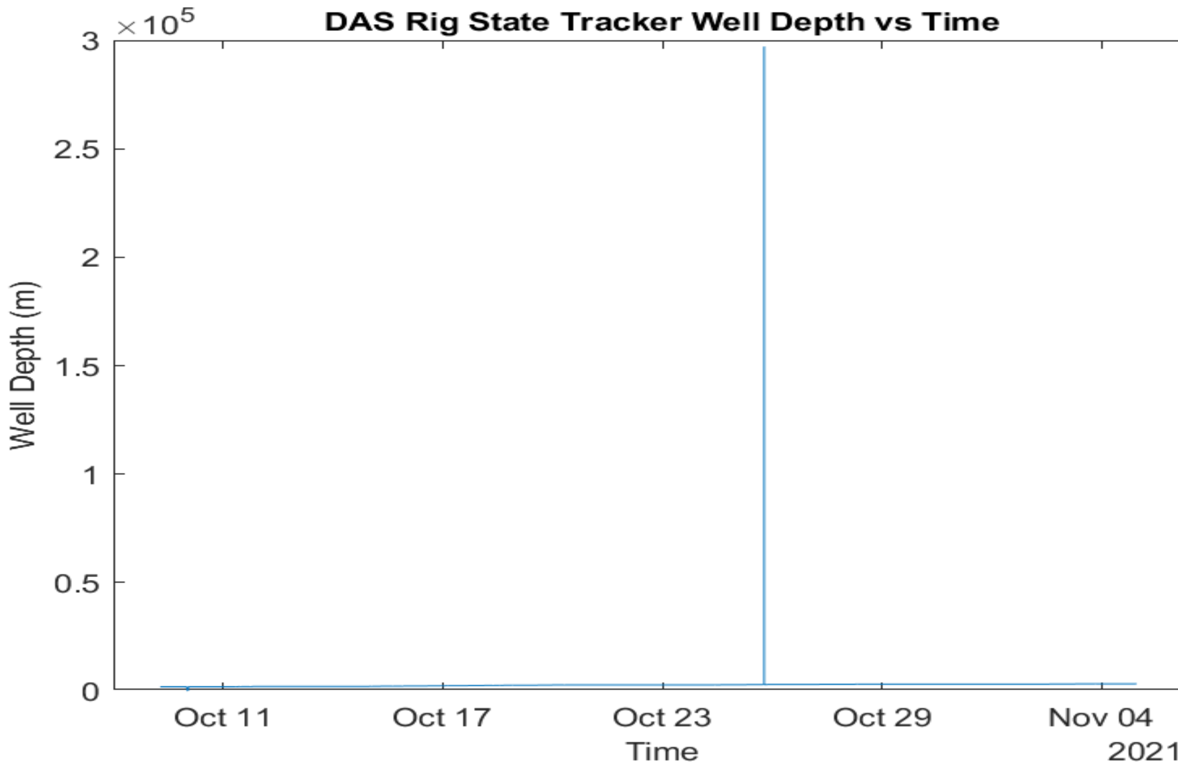

Figure 14: Well Depth vs. Time (Raw Data)

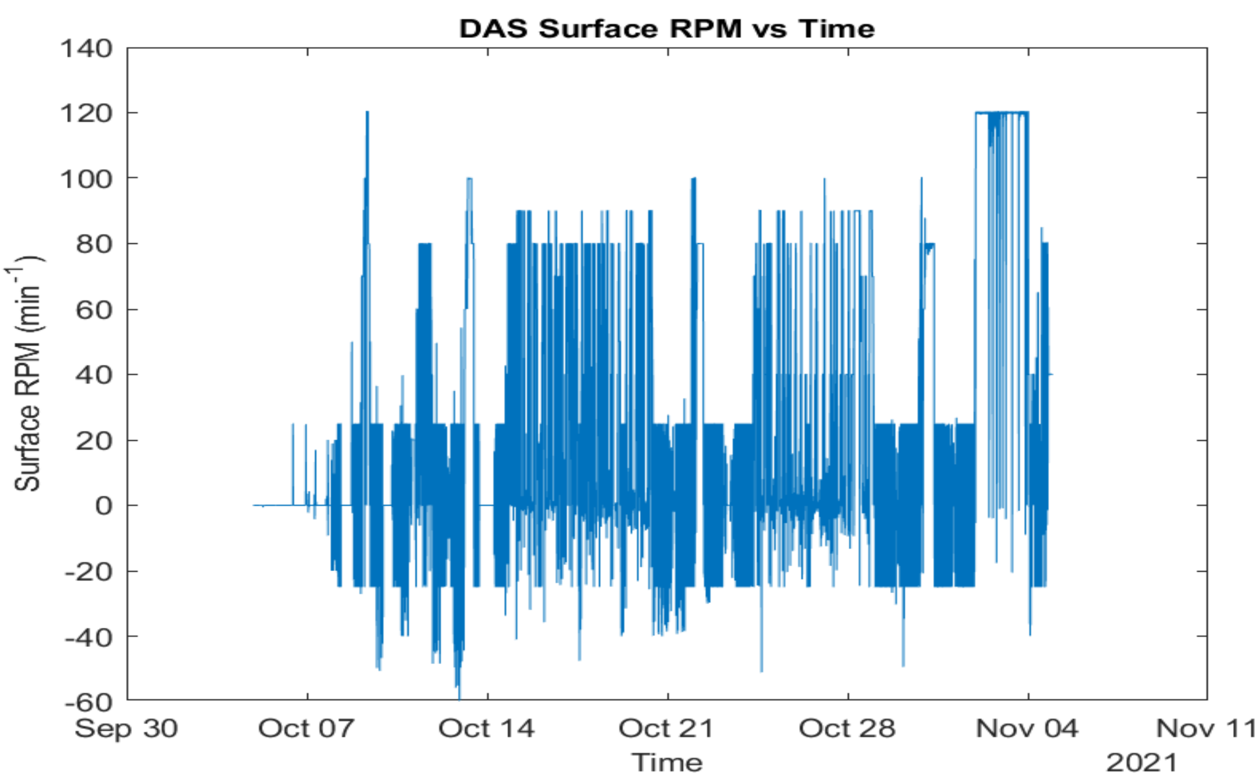

Figure 15: Surface RPM vs. Time (Raw Data)

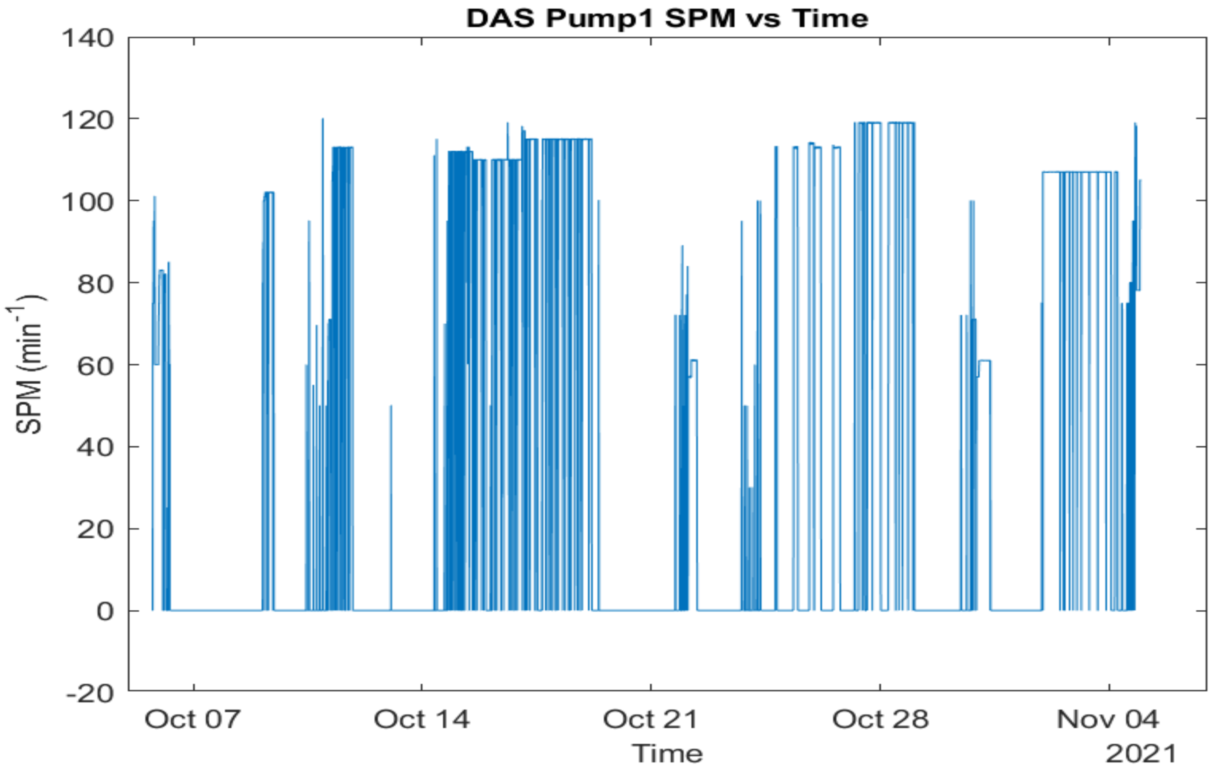

Figure 16: SPM Pump 1 vs. Time (Raw Data)

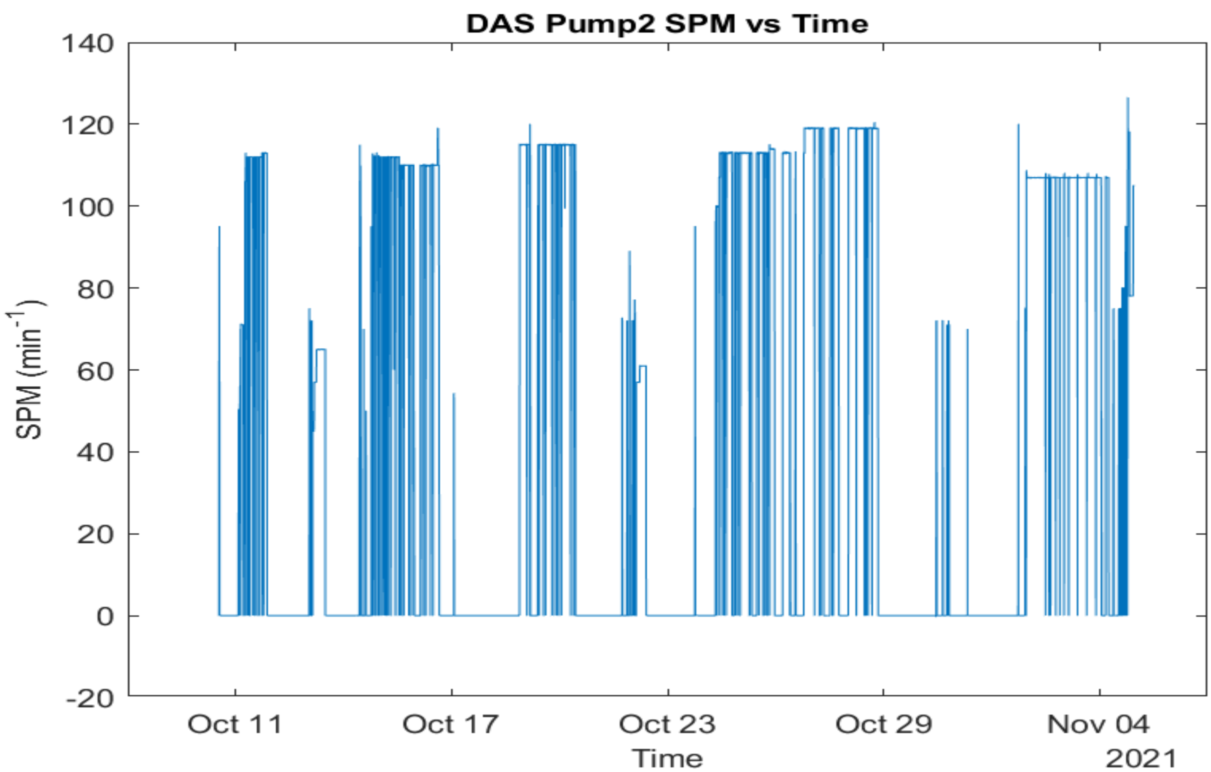

Figure 17: SPM Pump 2 vs. Time (Raw Data)

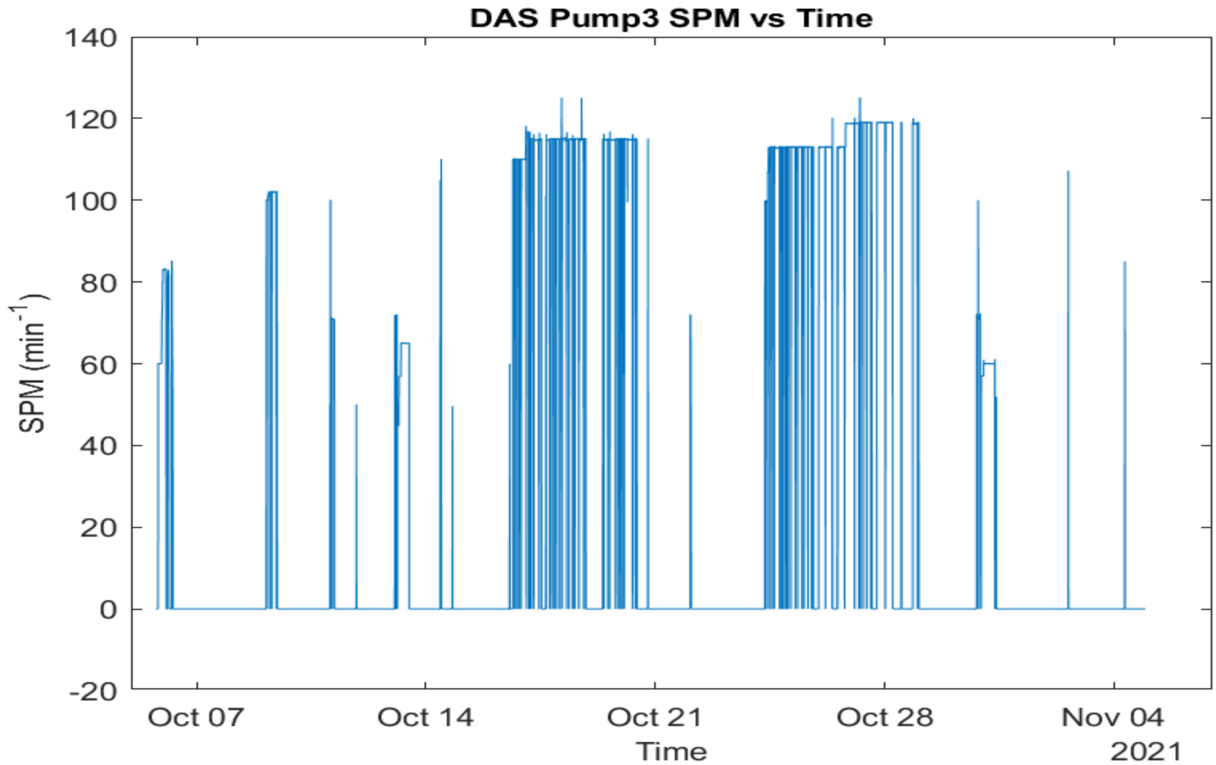

Figure 18: SPM Pump 3 vs. Time (Raw Data)

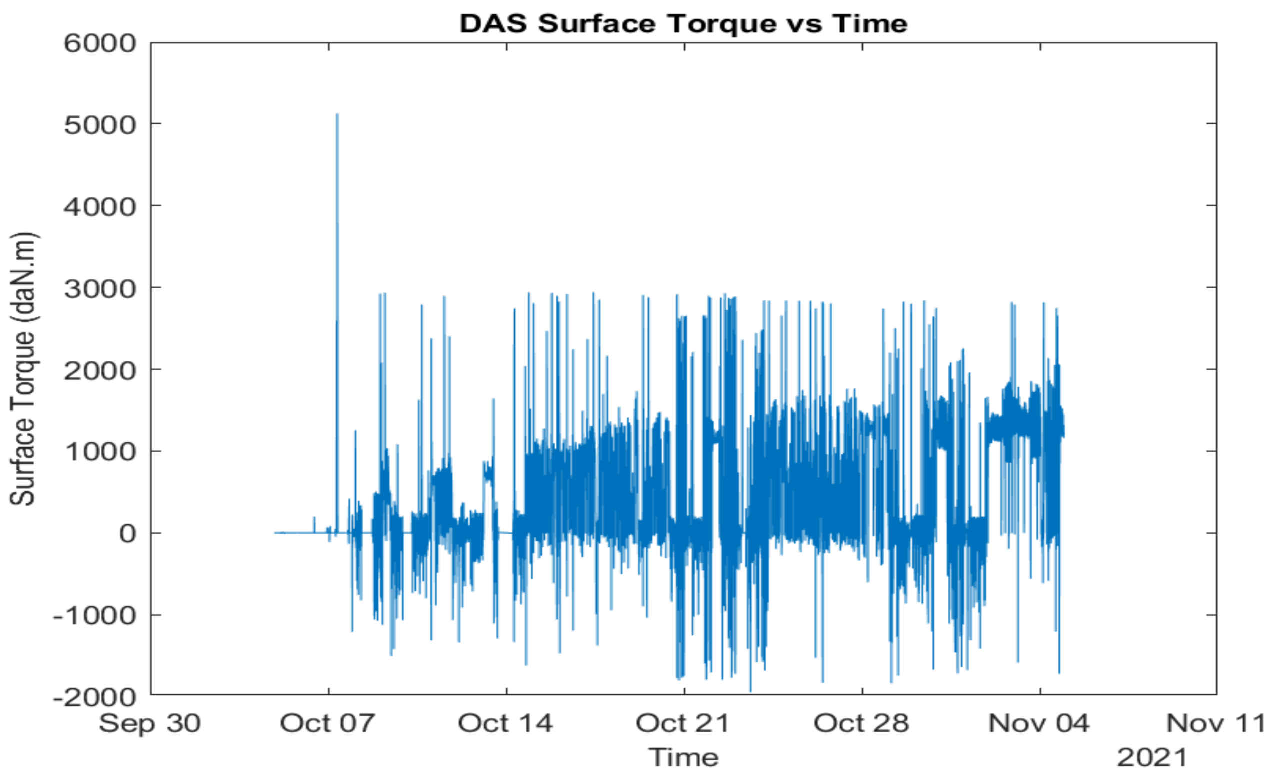

Figure 19: Surface Torque vs. Time (Raw Data)

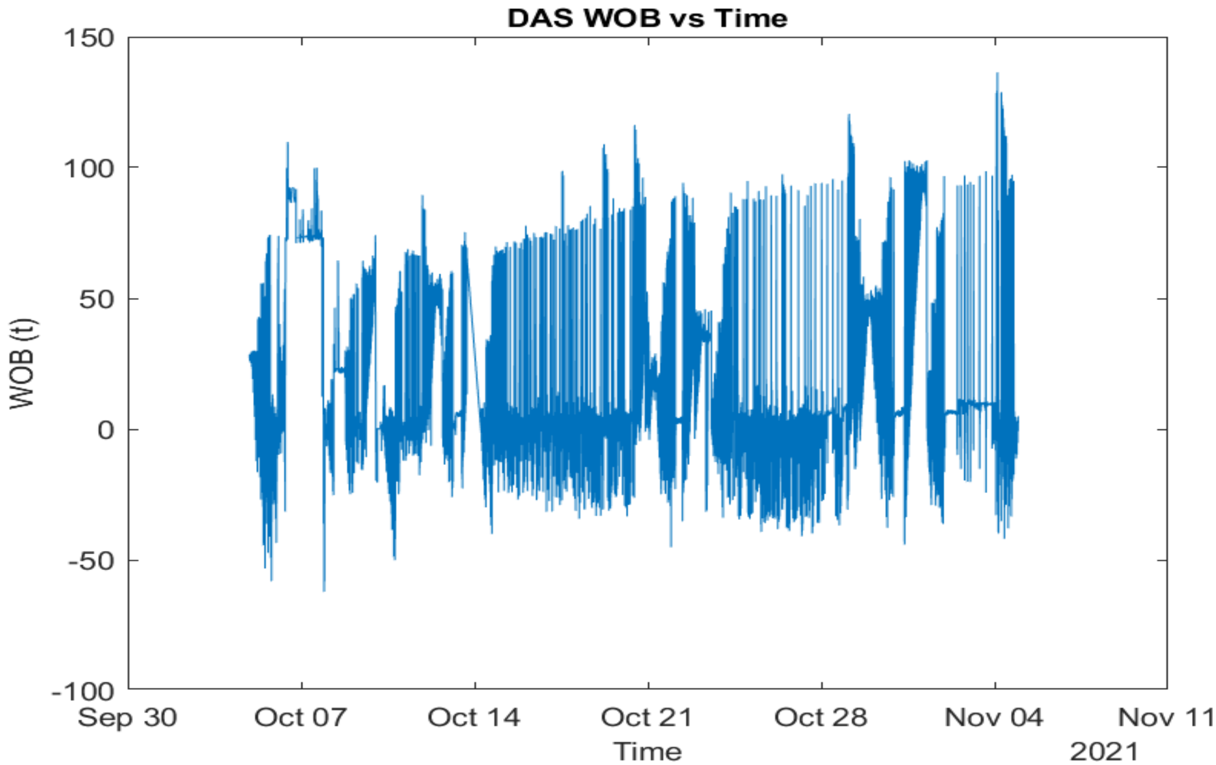

Figure 20: WOB vs. Time (Raw Data)

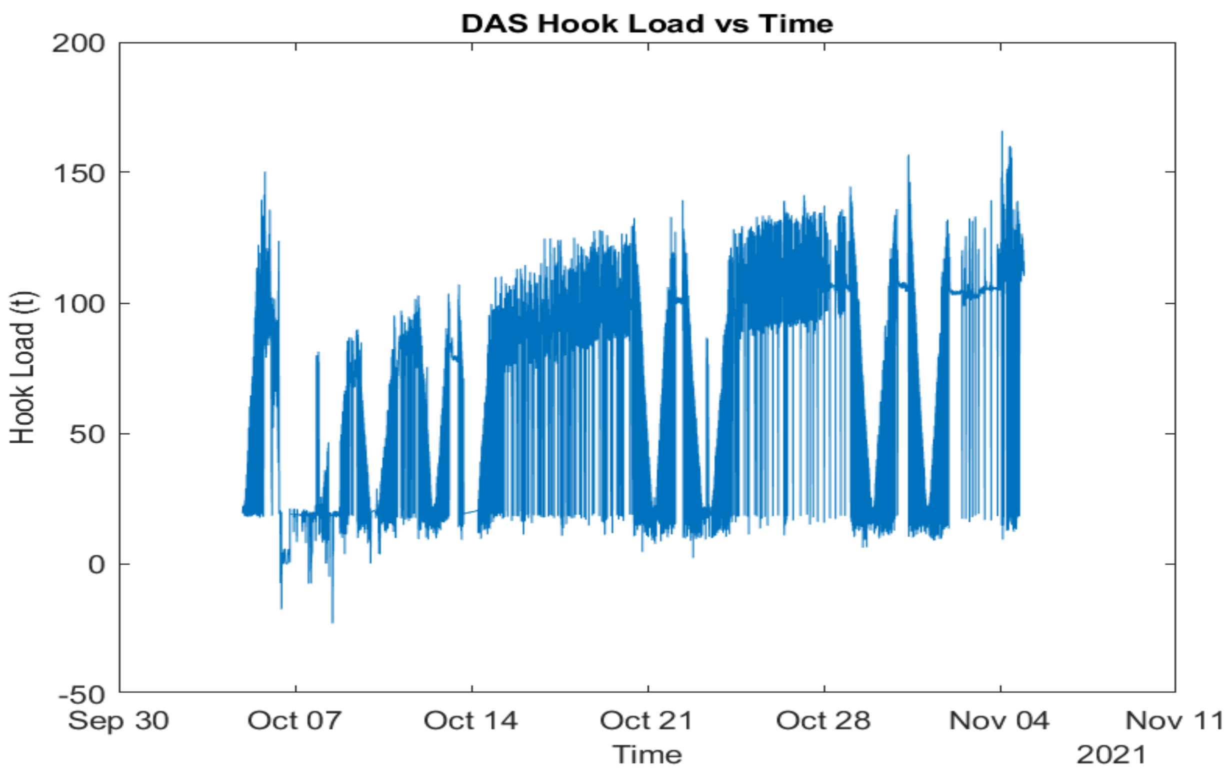

Figure 21: Hook Load vs. Time (Raw Data)

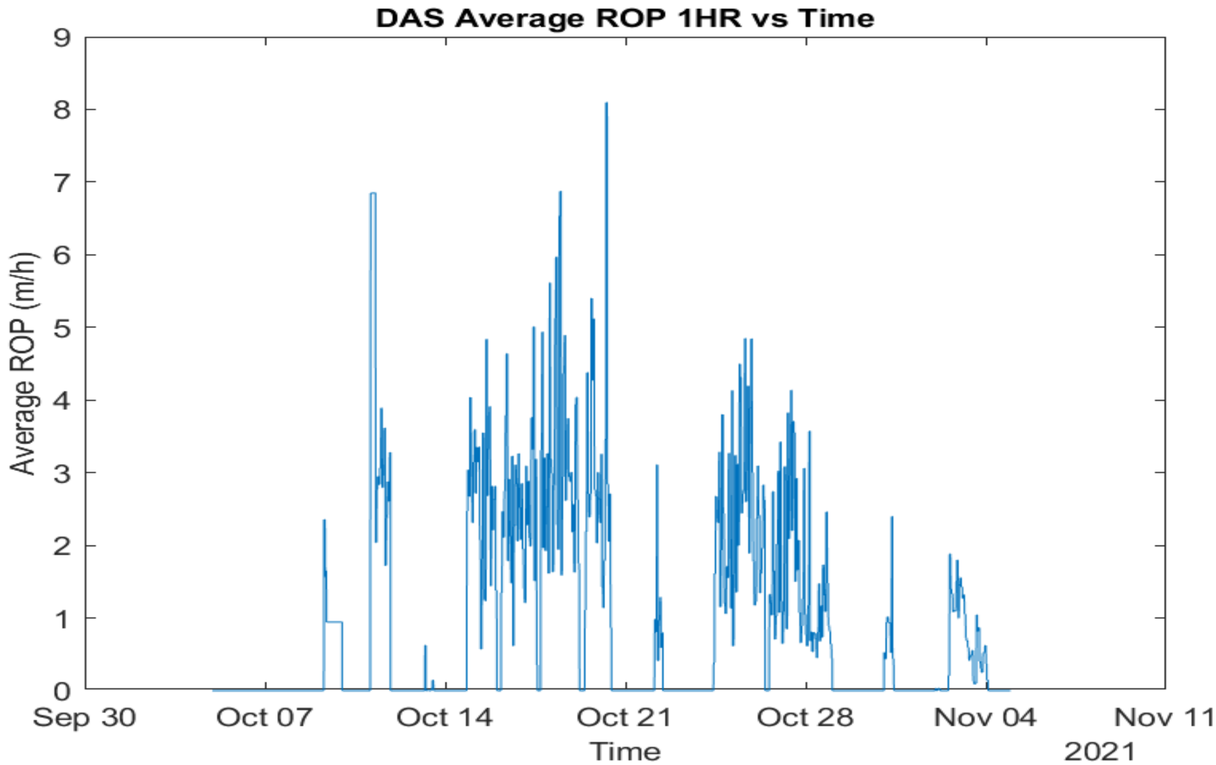

Figure 22: Average ROP every hour vs. Time (Raw Data)

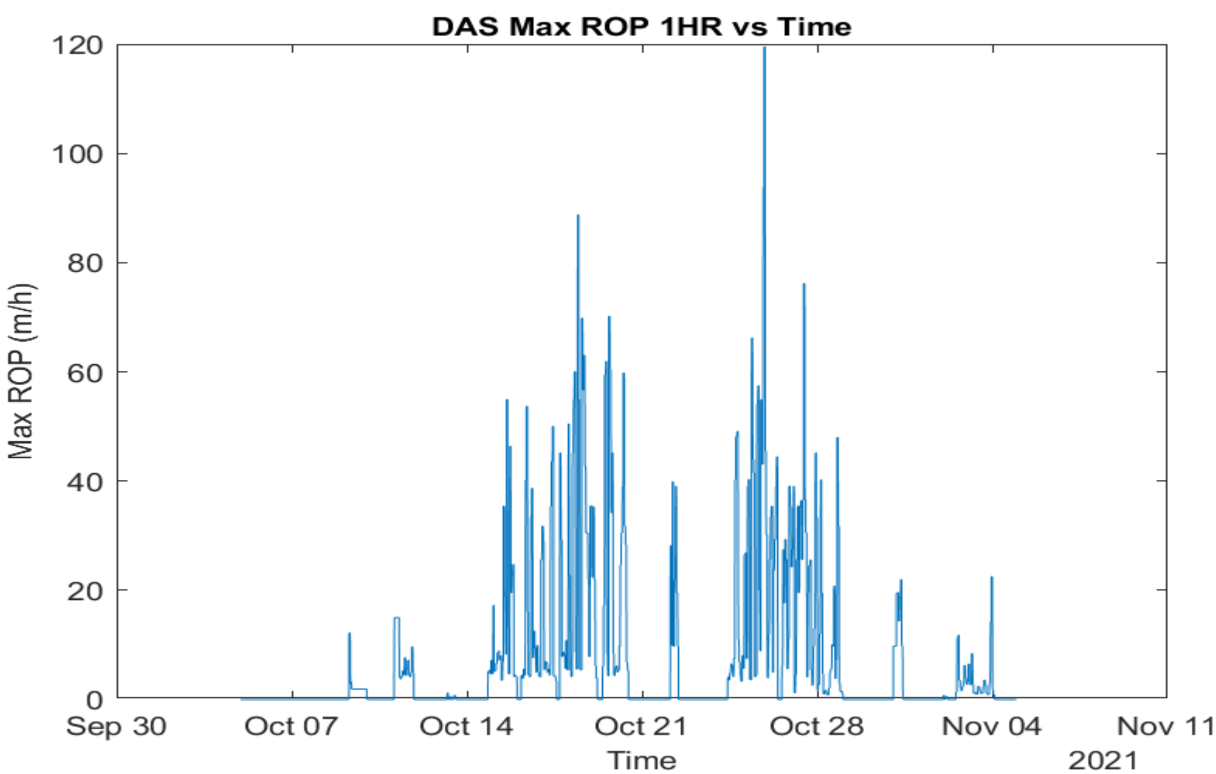

Figure 23: Max ROP every hour vs. Time (Raw Data)

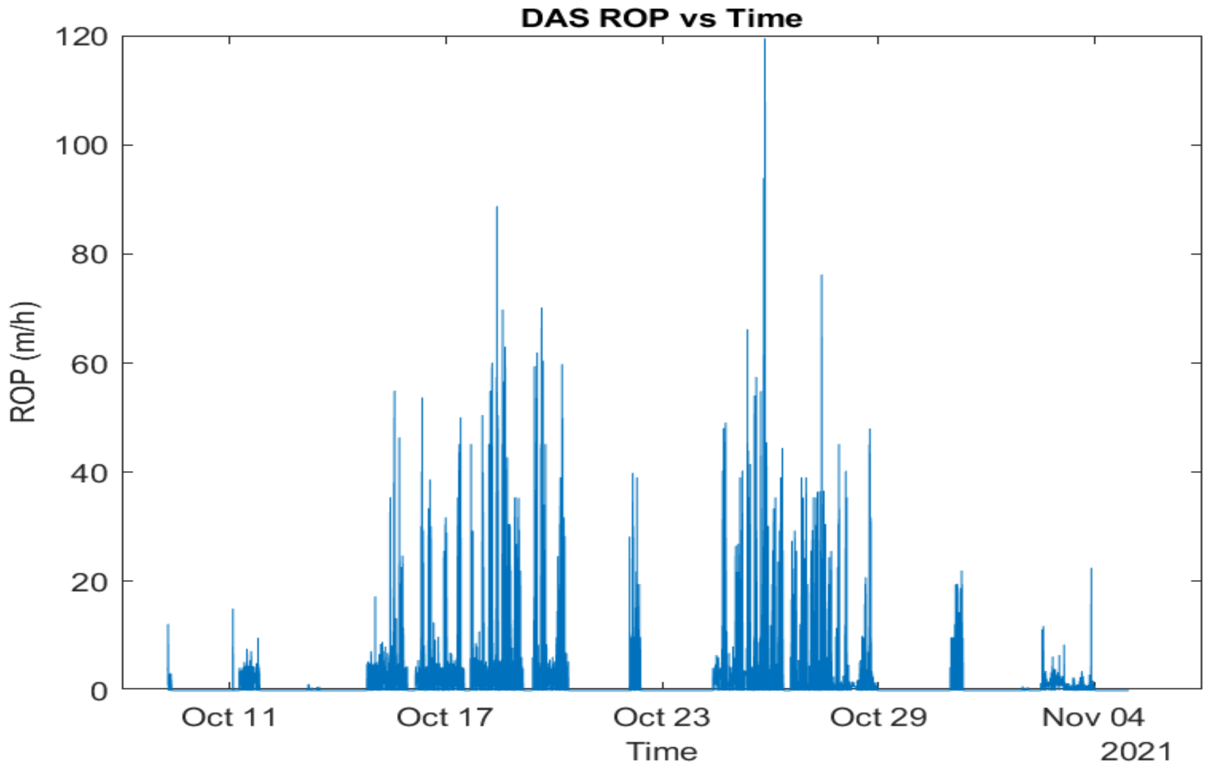

Figure 24: ROP vs. Time (Raw Data)

8.2.2 <u>HPU Raw Data</u>

The hydraulic power unit is the unit which is providing hydraulic power to all of the hydraulically operated equipment on the rig. This includes the raising and lowering of the mast, opening and closing of safety valves, top drives, power tongs, and much more. In this rig specifically, the top drive is electric, hence the HPU did not affect it. It is equipped with a huge number of sensors that captures many parameters with a frequency of five new values every hour. These parameters are, the pilot pressure of the HPU1 (Figure 25) and HPU2 (Figure 38), the average and maximum current (Figures 33 and 44) and the power (Figures 39 and 27) of one of the two electric motors of the hydraulic power unit, the average, and maximum flow out of the four pumps, PD1 (Figures 47 and 28), PD2 (Figures 43 and 37), TD1 (Figures 26 and 50) and TD2 (Figures 48 and 42), as well as their average and maximum pressure out (Figures 34, 40, 35, 41, 36, 46, 31 and 45), the temperature of the hydraulic oil present in the tank (Figure 29) and its actual level in it (Figure 49), the working hours of the electric motor of HPU1 (Figure 30) and the heat exchanger used to cool the hydraulic oil (Figure 32) were recorded for a month and in these following graphs their raw initial data will be represented as a function of time.

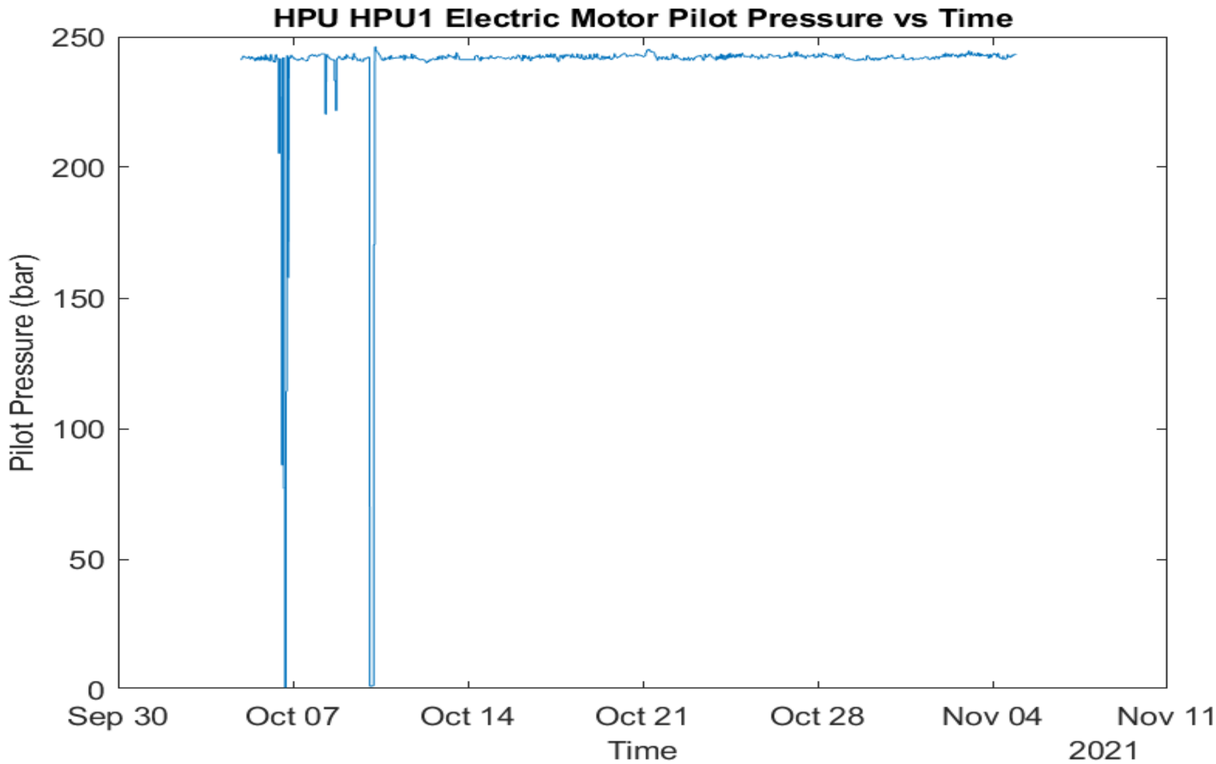

Figure 25: HPU1 Pilot Pressure vs. Time (Raw Data)

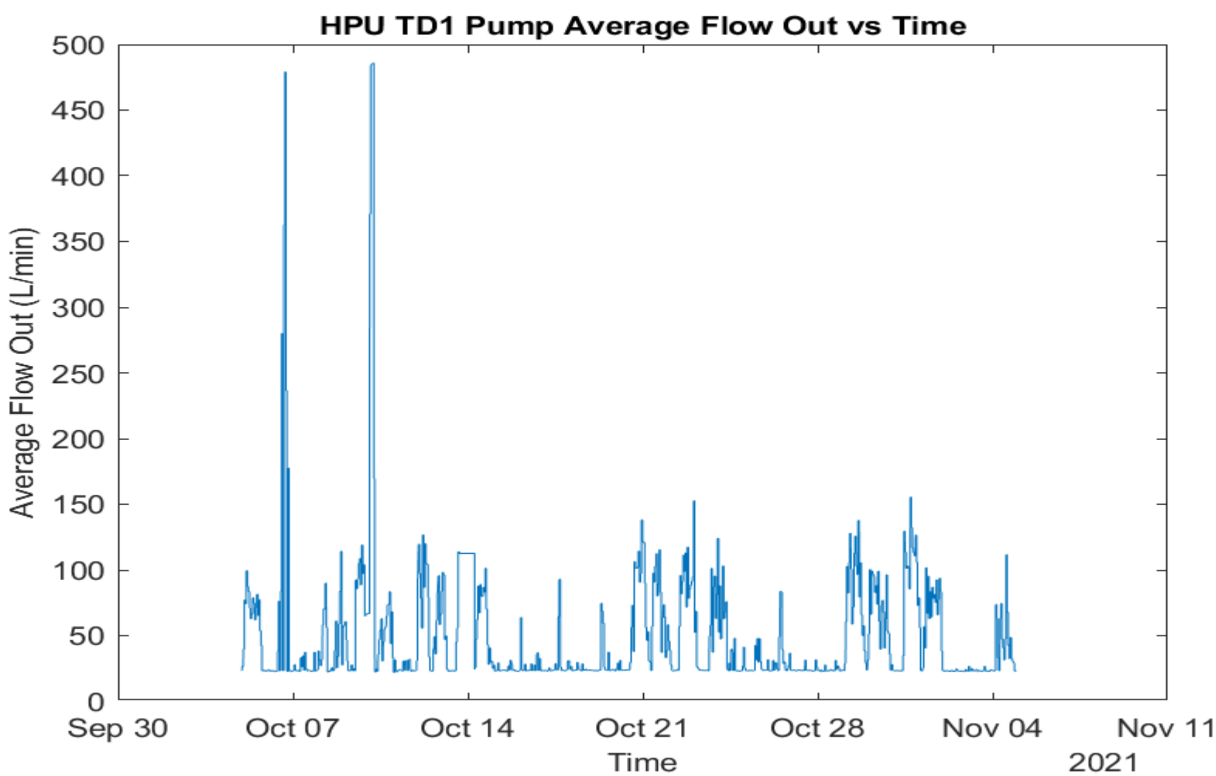

Figure 26: TD1 Average Flow Out vs. Time (Raw Data)

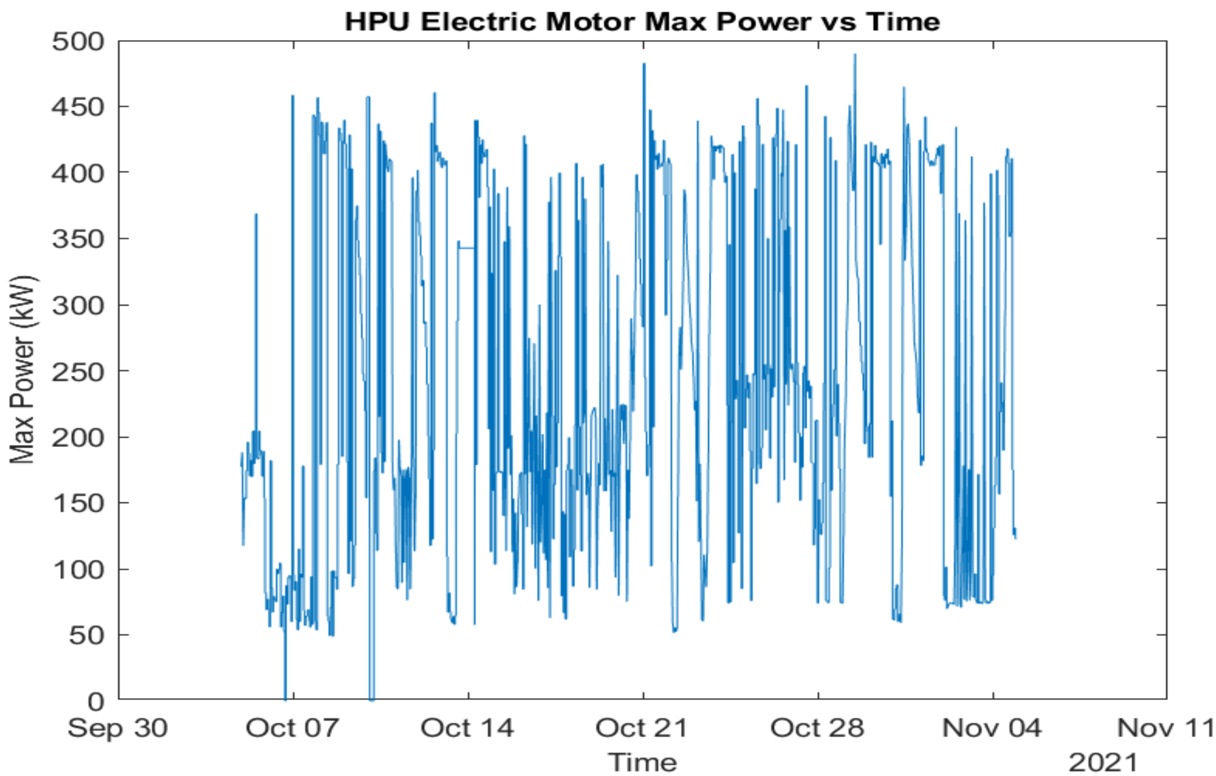

Figure 27: Max Power vs. Time (Raw Data)

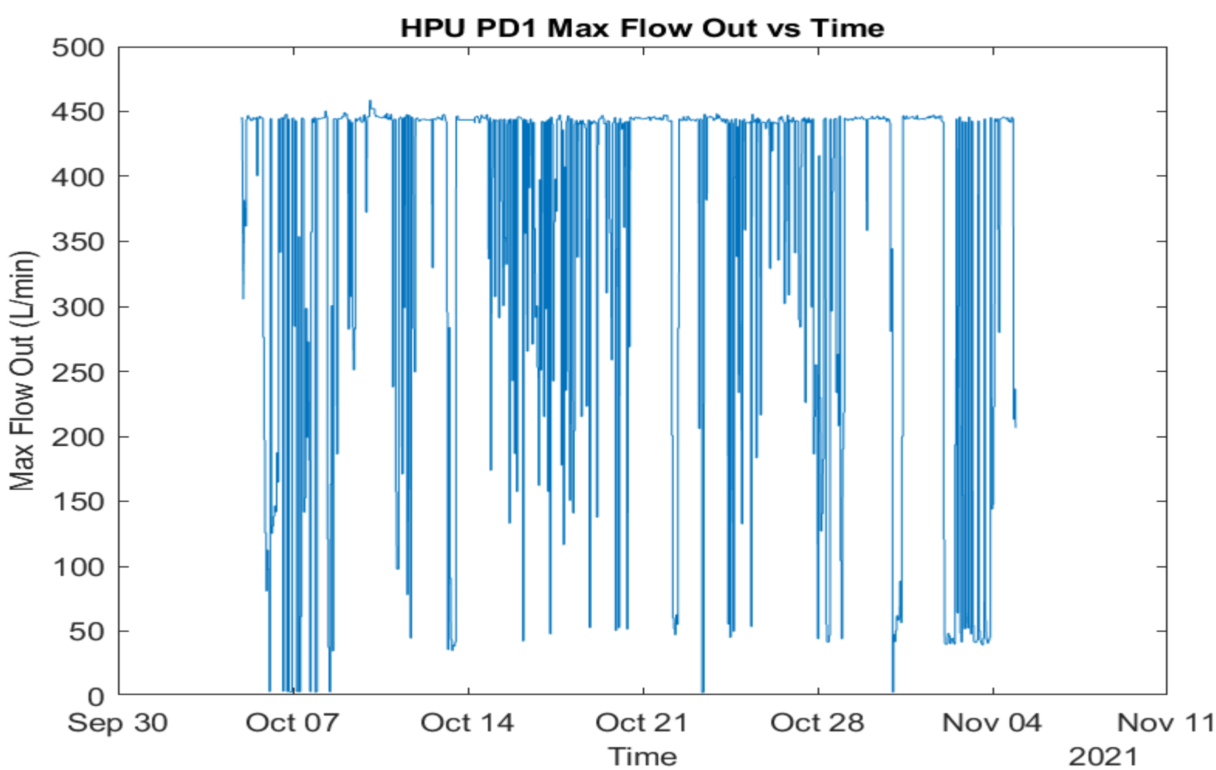

Figure 28: PD1 Max Flow Out vs. Time (Raw Data)

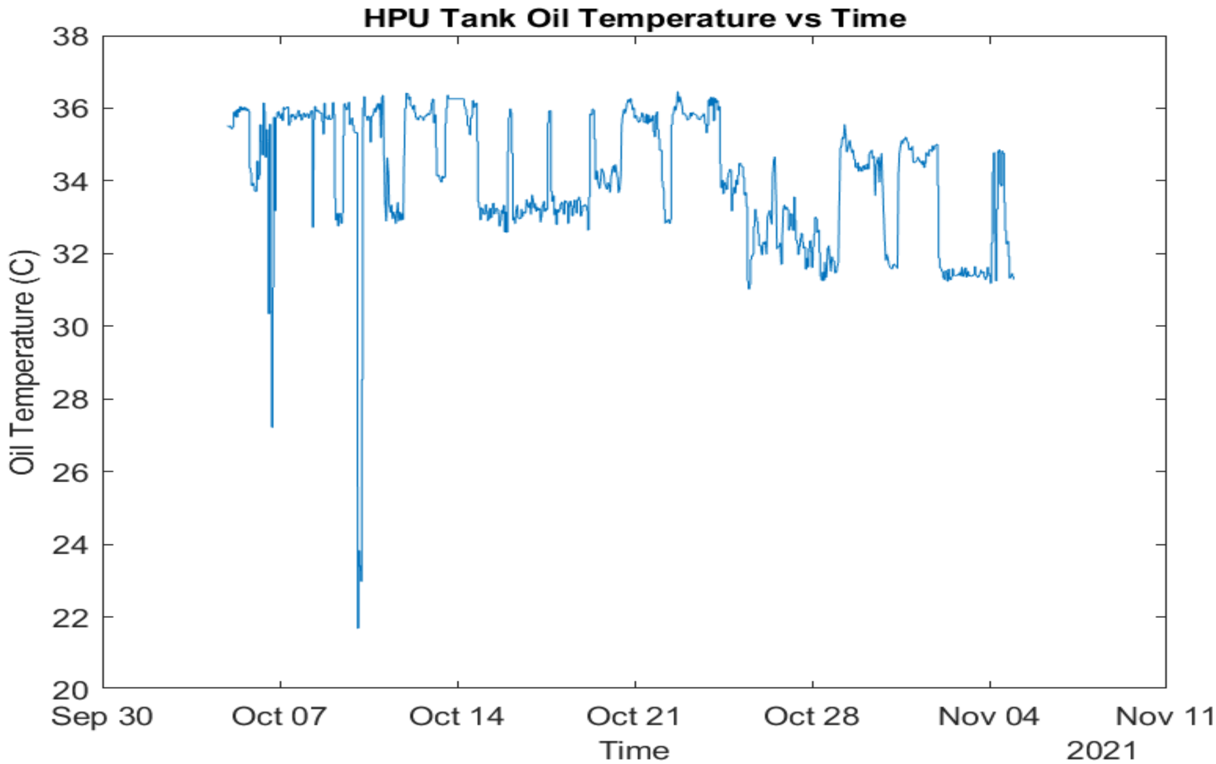

Figure 29: Oil Temperature vs. Time (Raw Data)

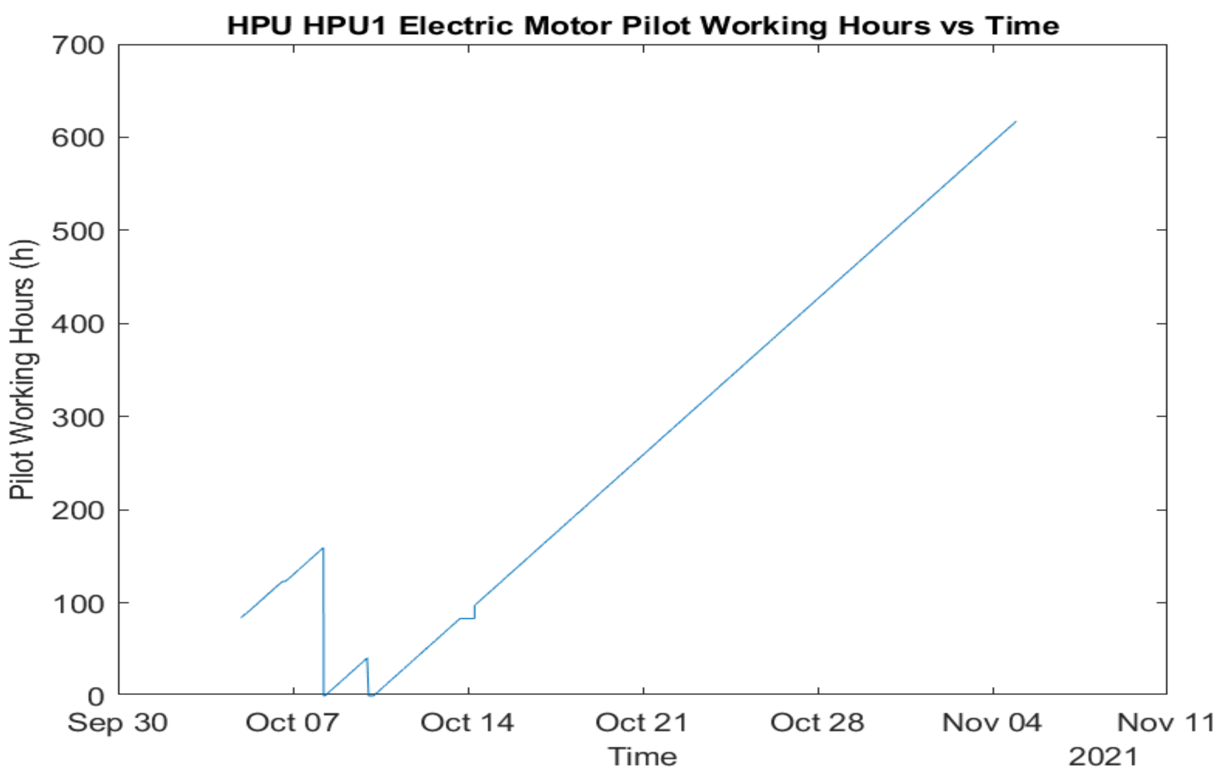

Figure 30: HPU1 Working Hours vs. Time (Raw Data)

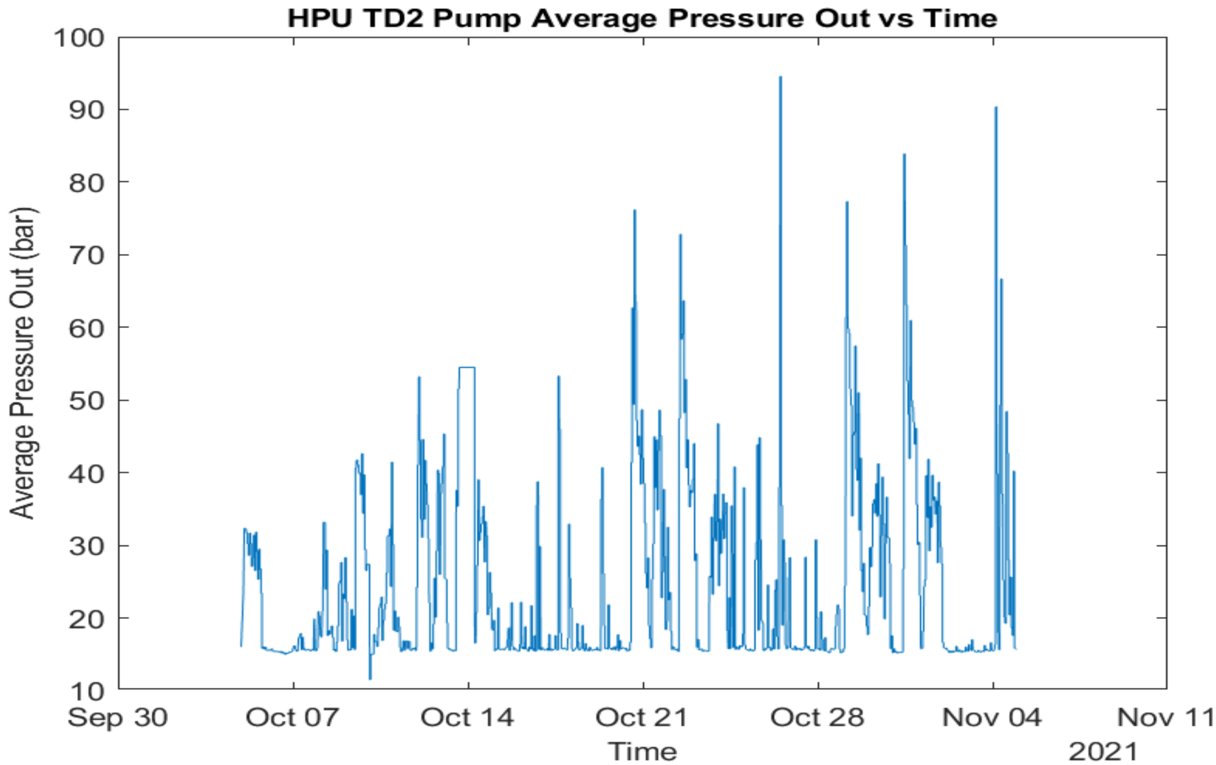

Figure 31: TD2 Average Pressure Out vs. Time (Raw Data)

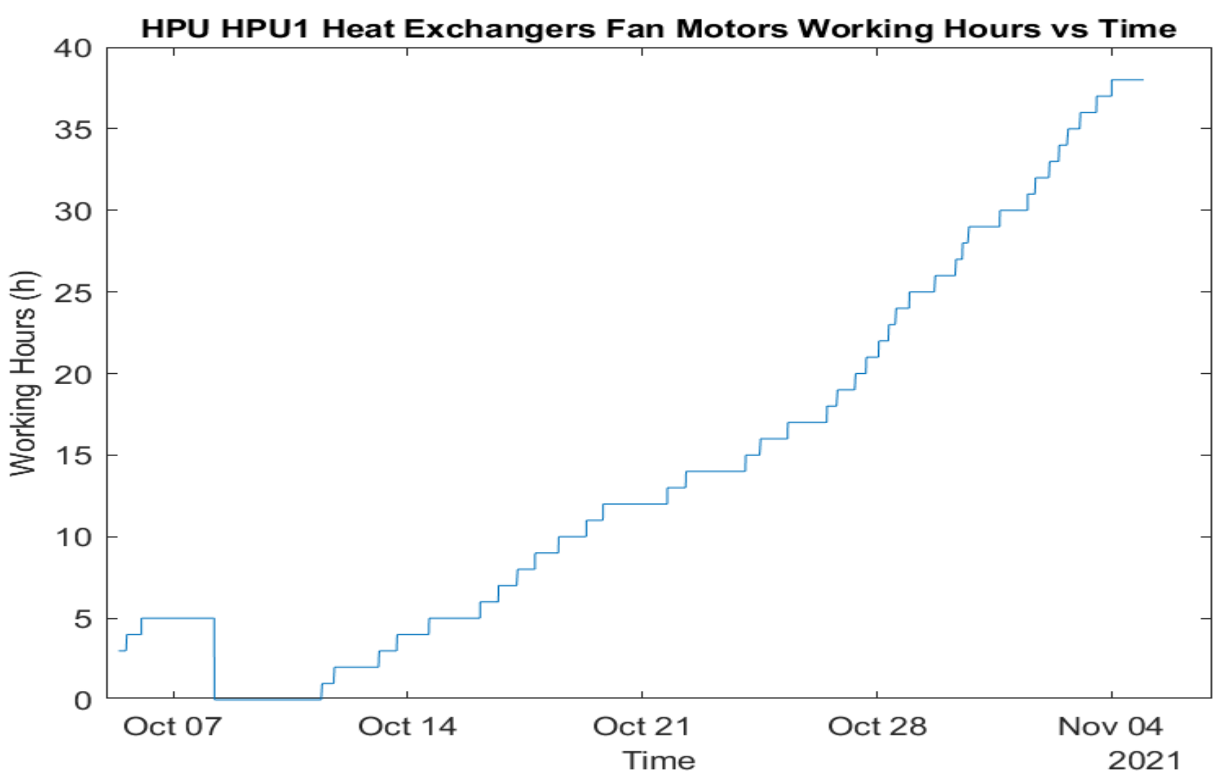

Figure 32: Heat Exchanger Working Hours vs. Time (Raw Data)

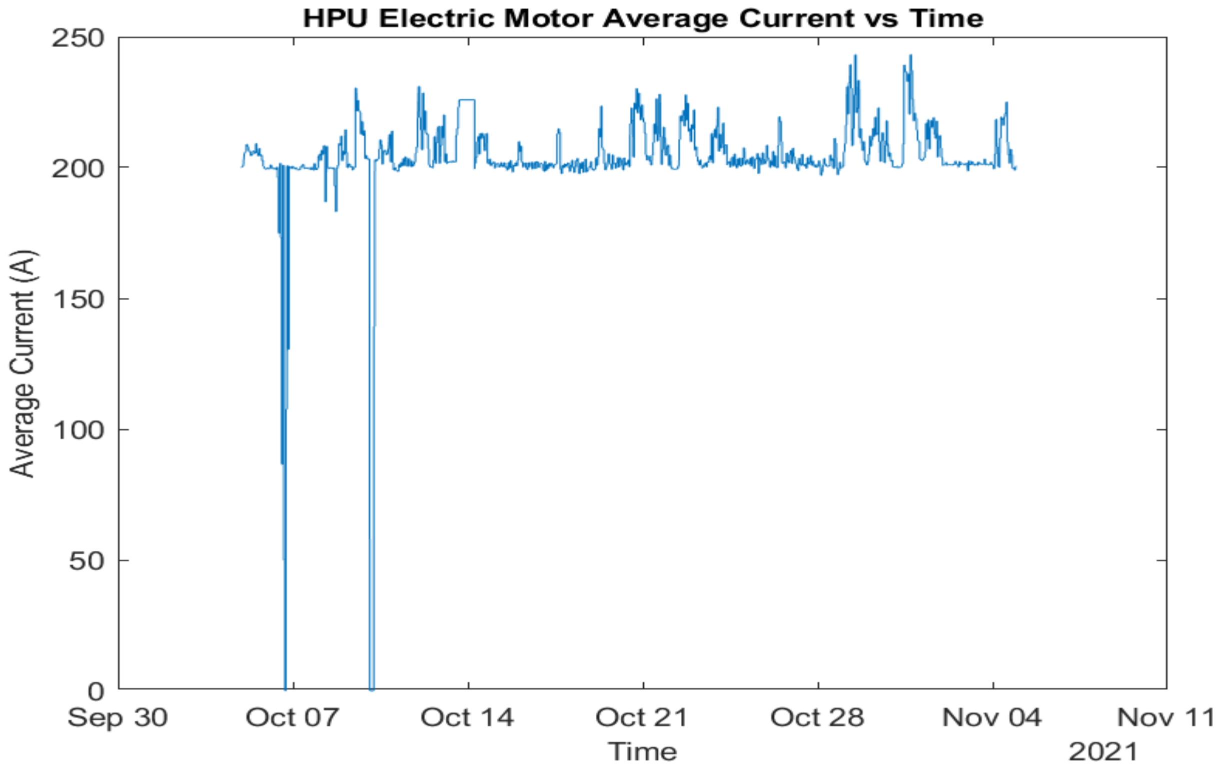

Figure 33: Average Current vs. Time (Raw Data)

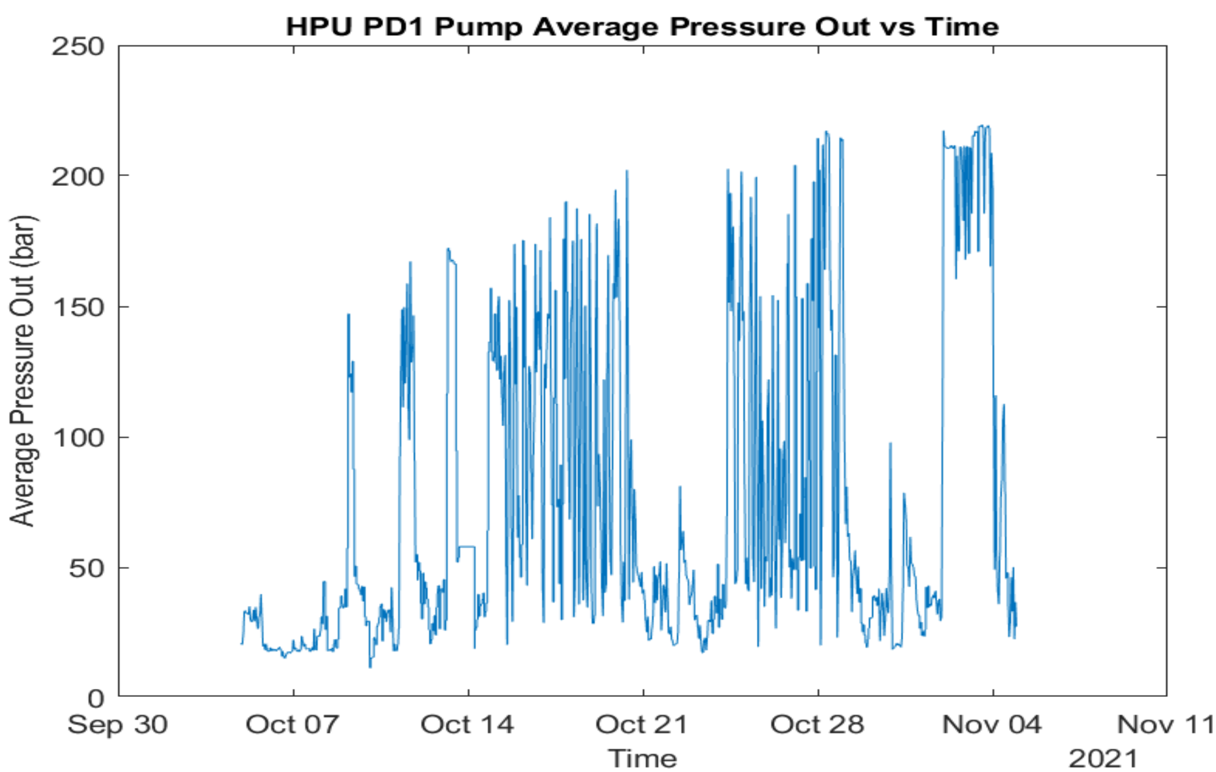

Figure 34: PD1 Average Pressure Out vs. Time (Raw Data)

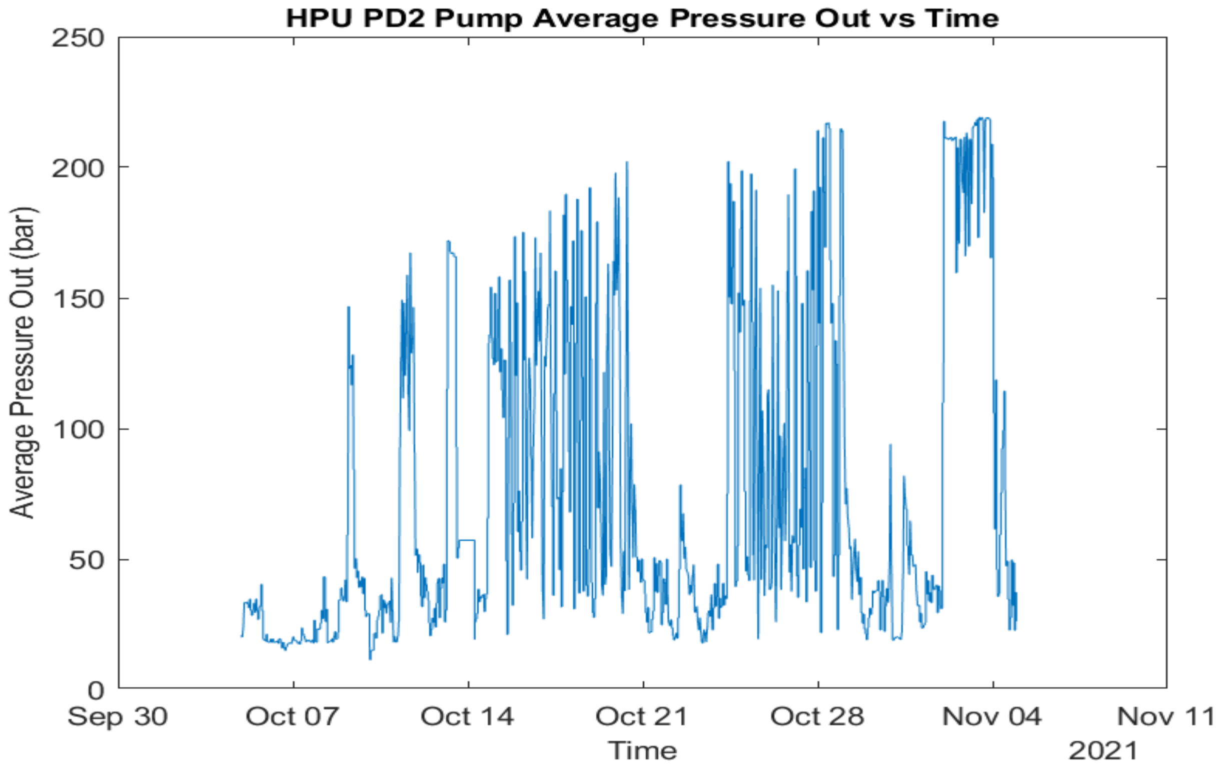

Figure 35: PD2 Average Pressure Out vs. Time (Raw Data)

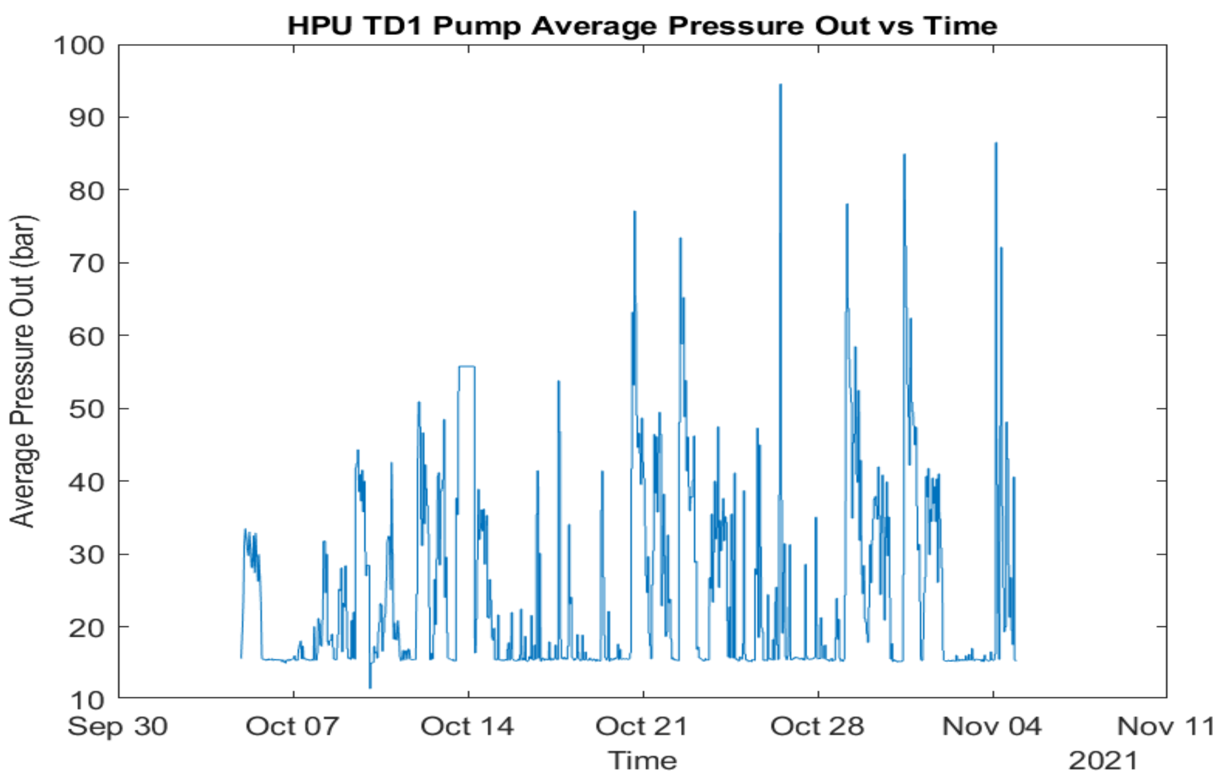

Figure 36: TD1 Average Pressure Out vs. Time (Raw Data)

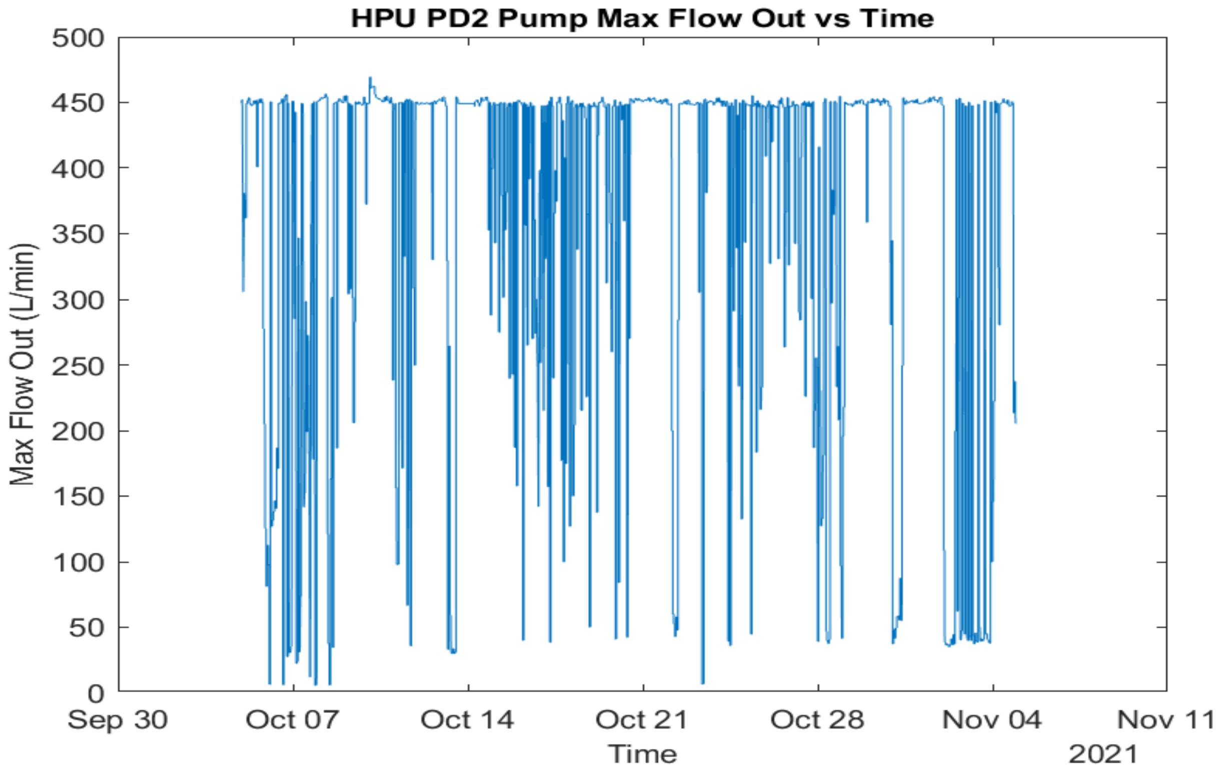

Figure 37: PD2 Max Flow Out vs. Time (Raw Data)

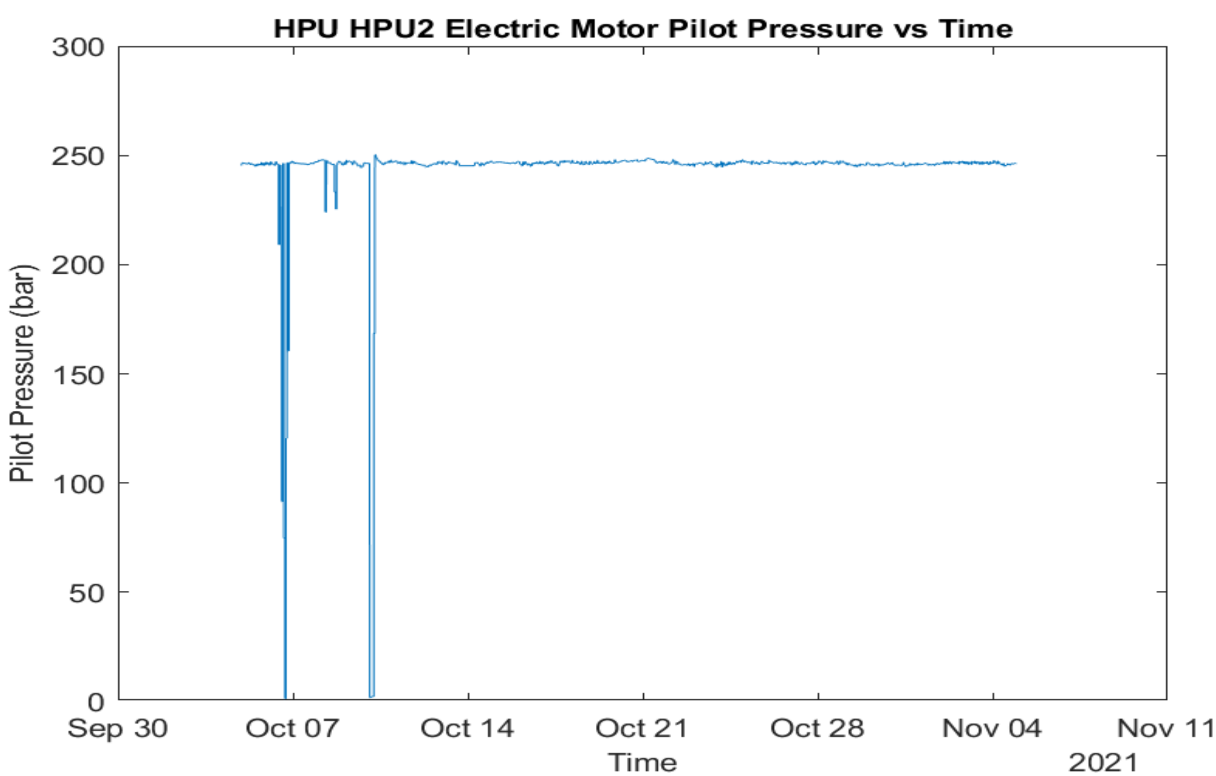

Figure 38: HPU2 Pilot Pressure vs. Time (Raw Data)

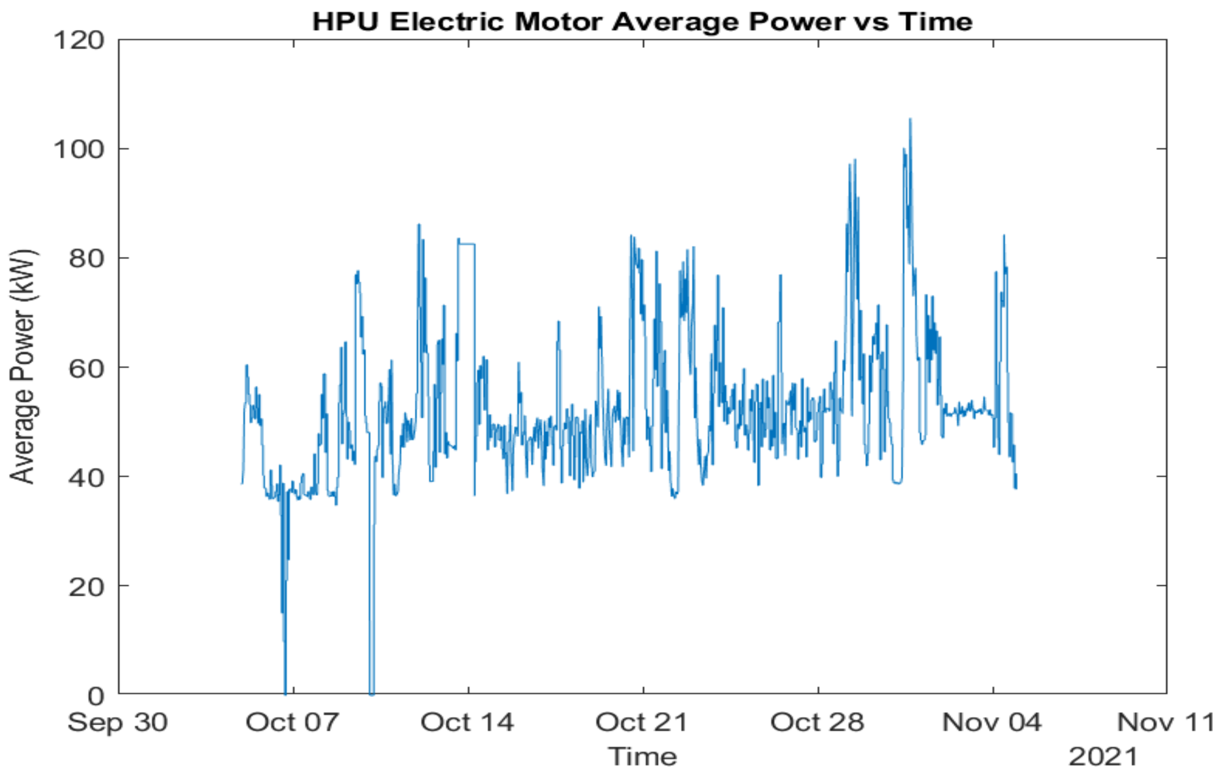

Figure 39: Average Power vs. Time (Raw Data)

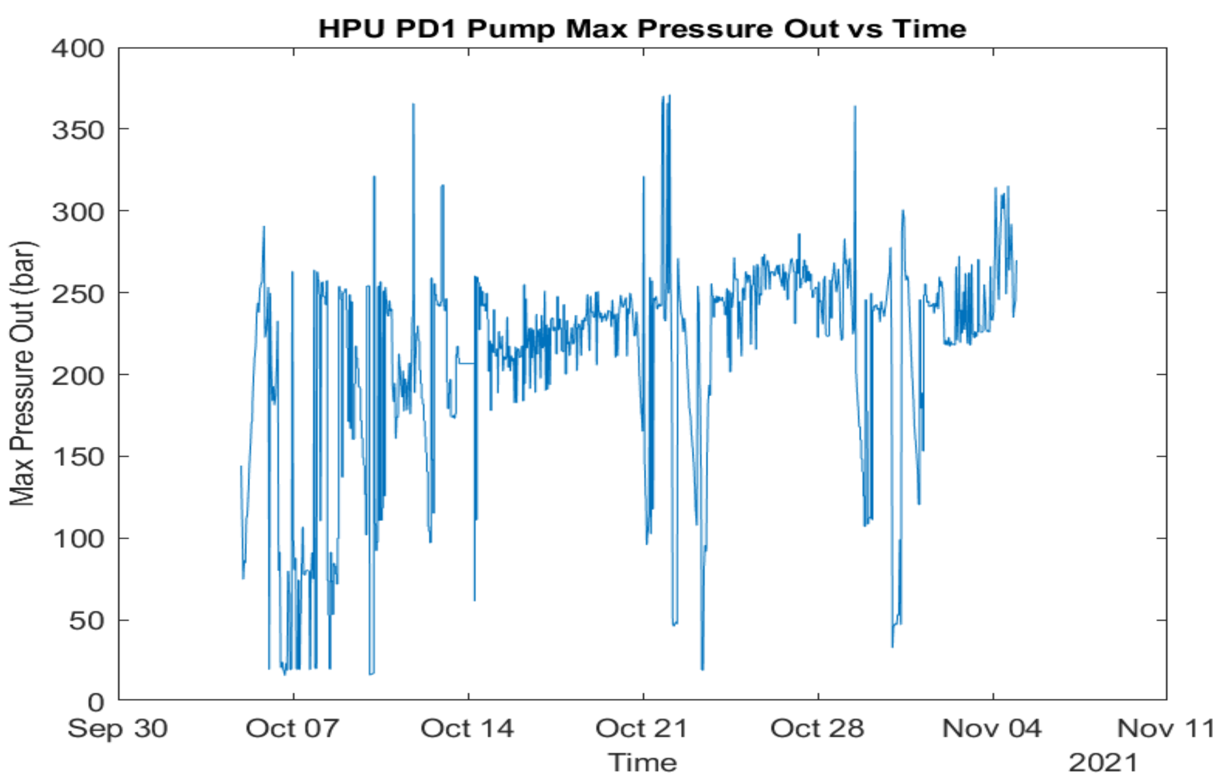

Figure 40: PD1 Max Pressure Out vs. Time (Raw Data)

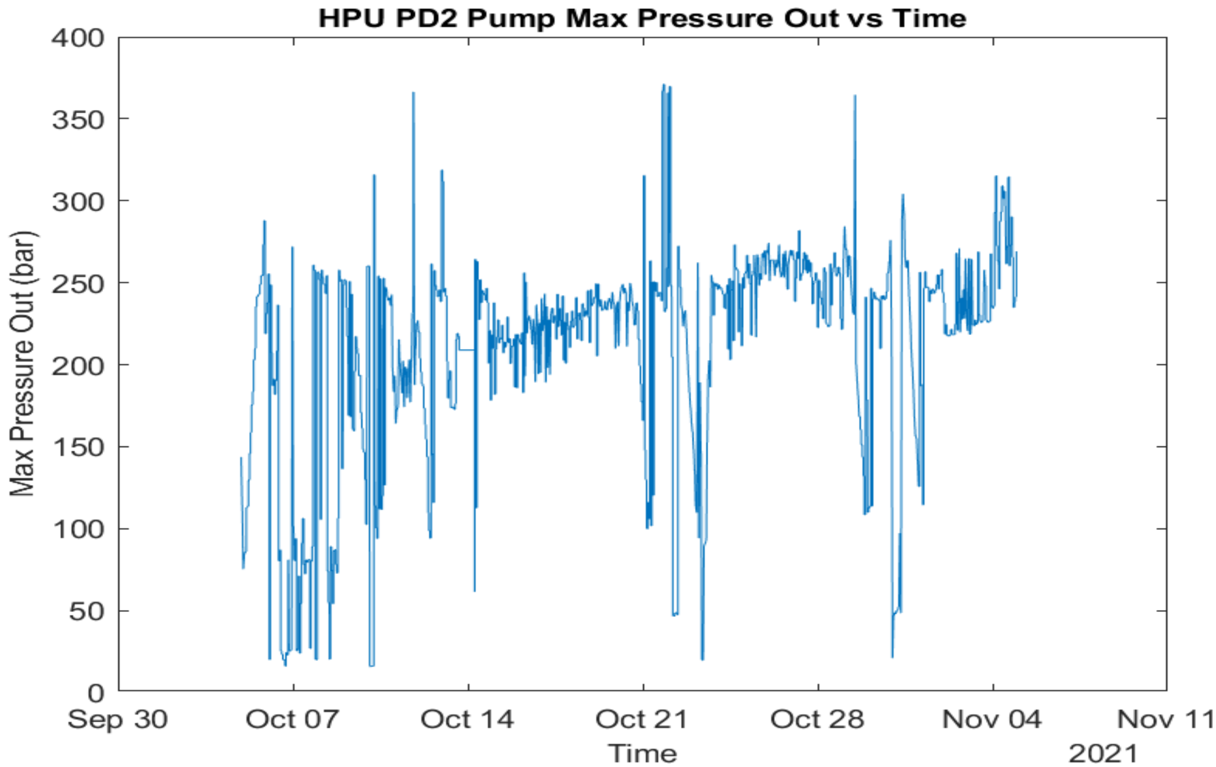

Figure 41: PD2 Max Pressure Out vs. Time (Raw Data)

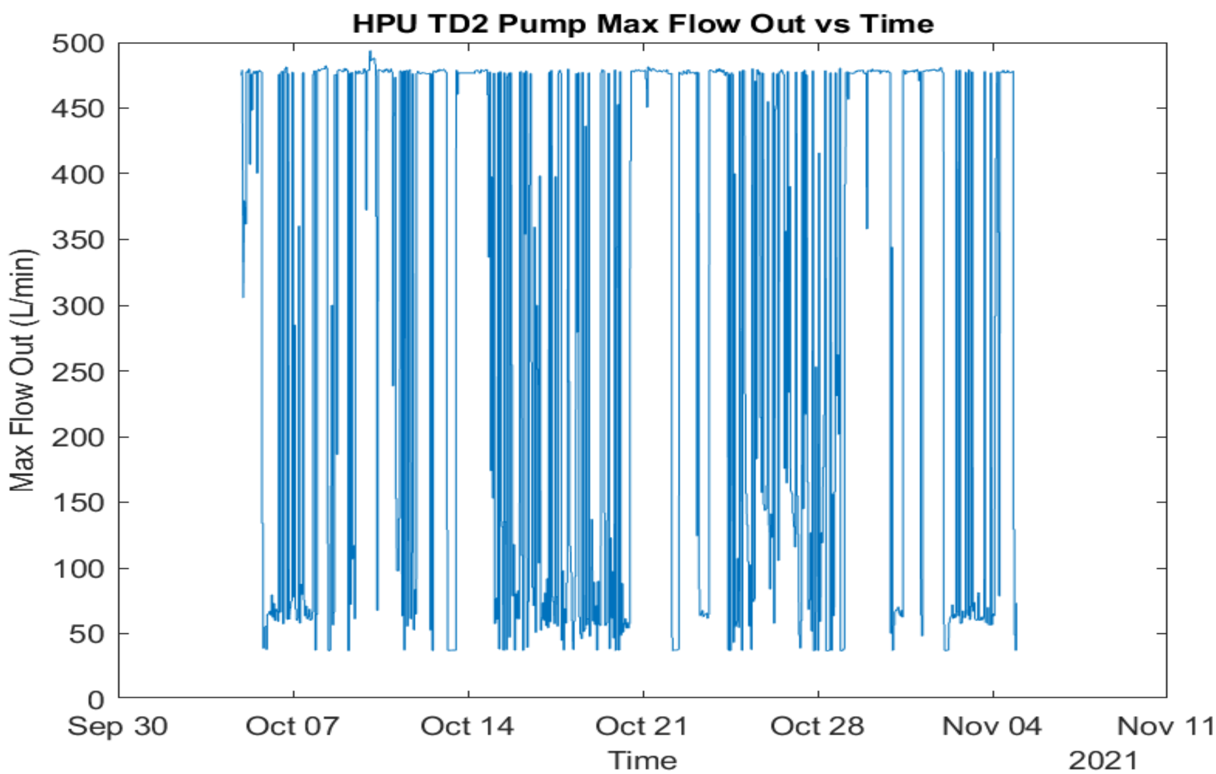

Figure 42: TD2 Max Flow Out vs. Time (Raw Data)

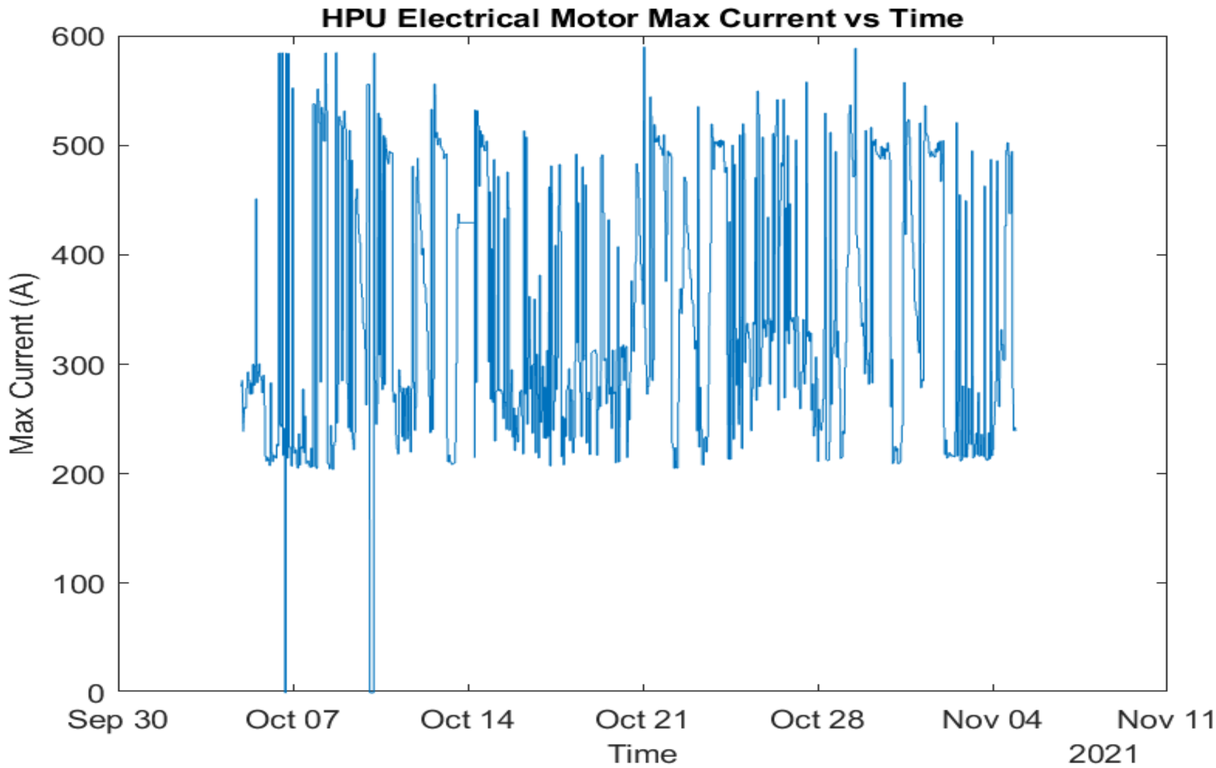

Figure 43: Max Current vs. Time (Raw Data)

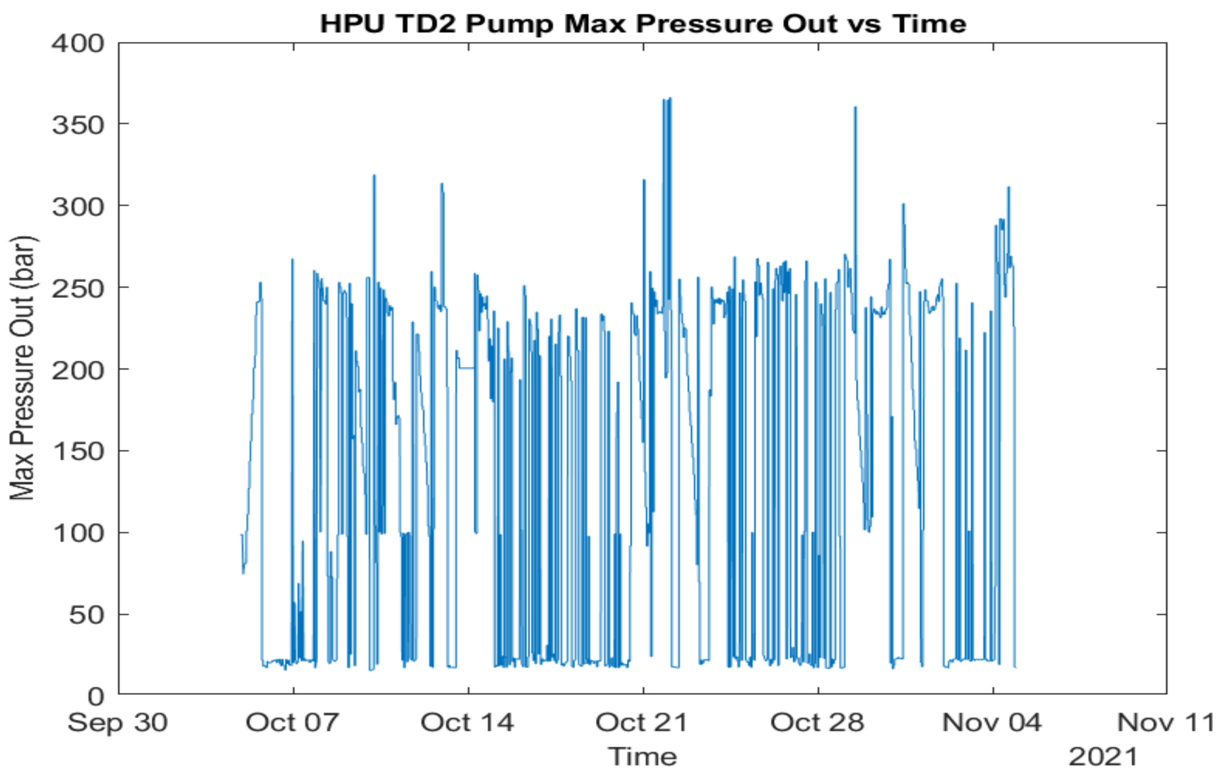

Figure 44: TD2 Max Pressure Out vs. Time (Raw Data)

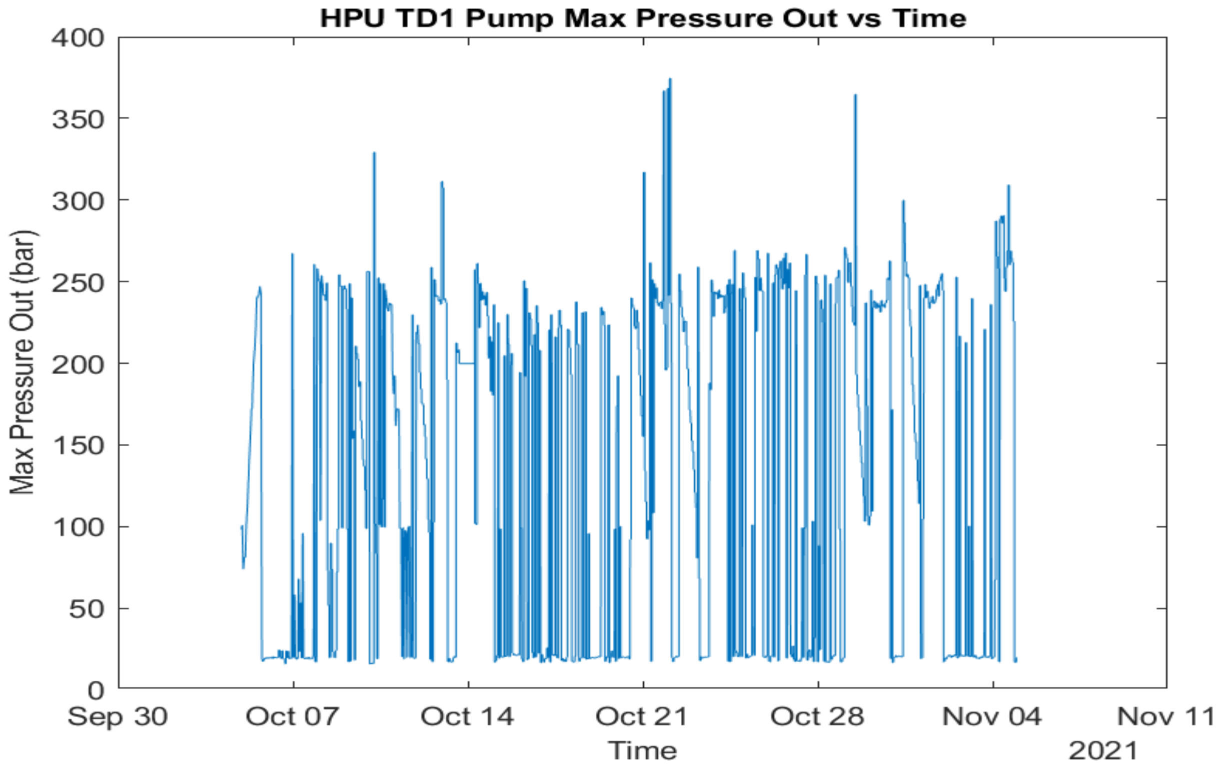

Figure 45: TD1 Max Pressure Out vs. Time (Raw Data)

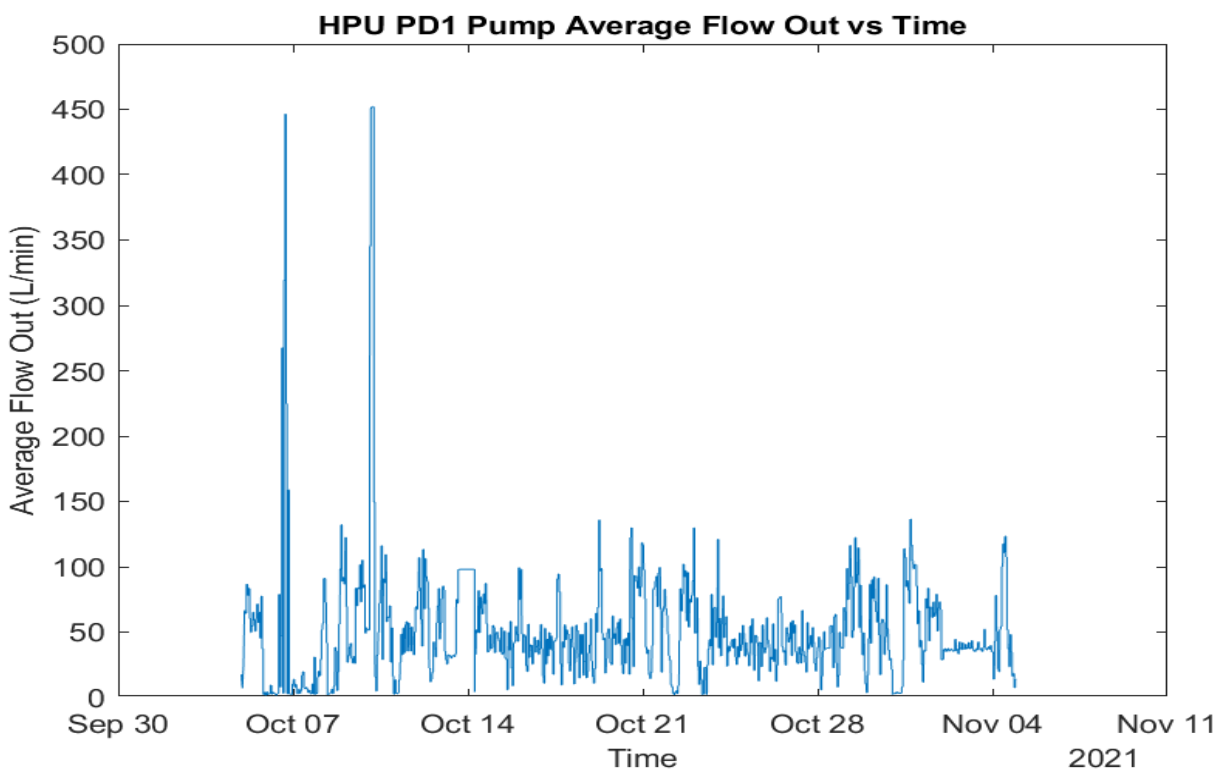

Figure 46: PD1 Average Flow Out vs. Time (Raw Data)

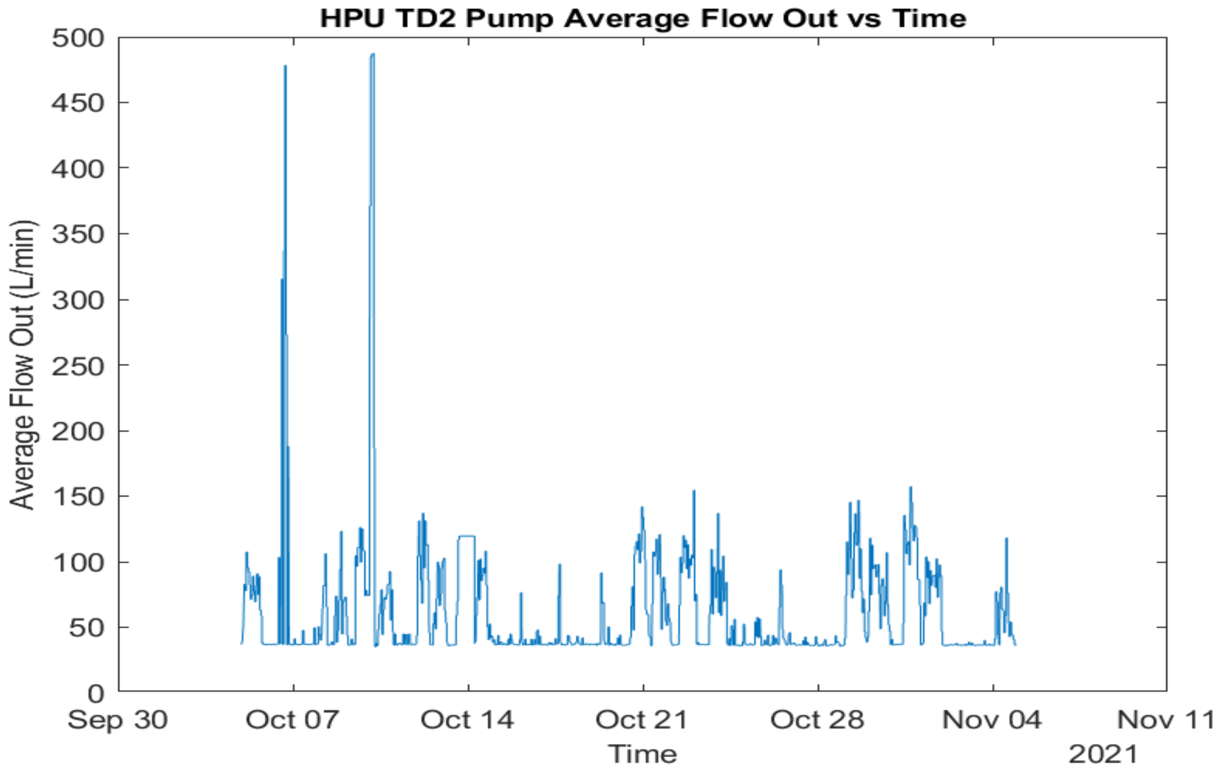

Figure 47: TD2 Average Flow Out vs. Time (Raw Data)

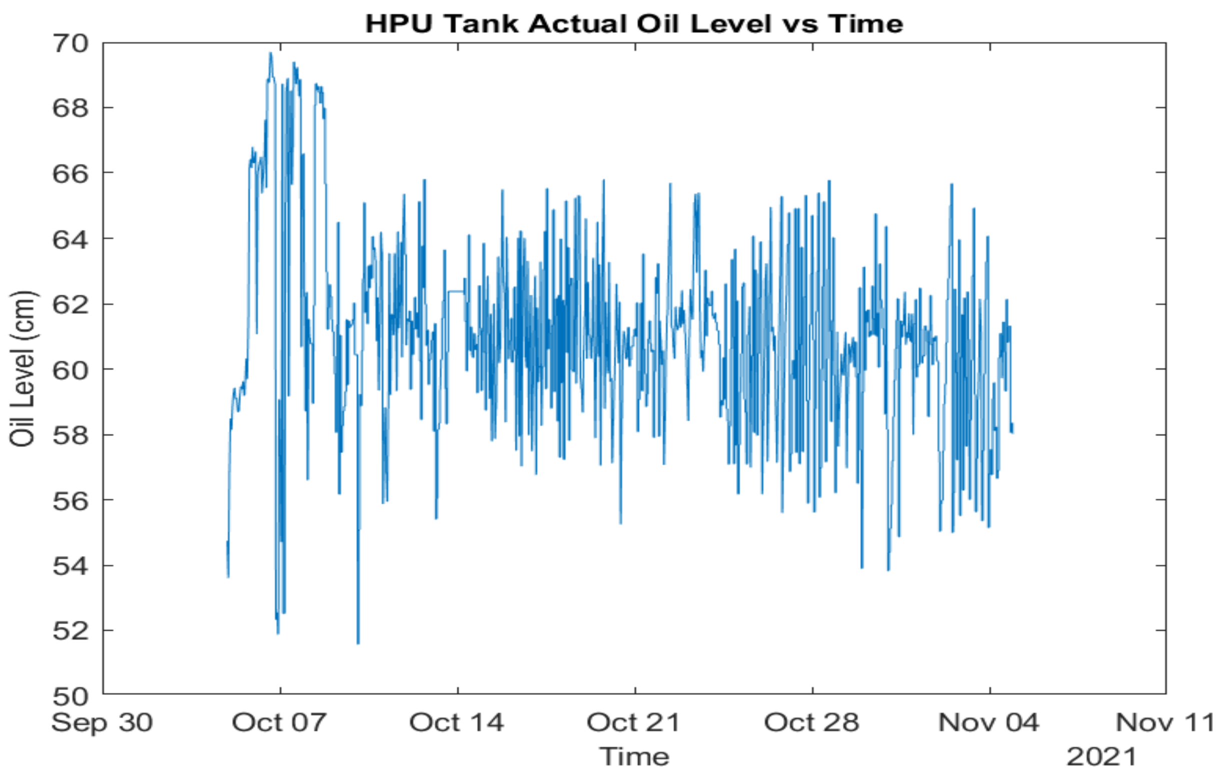

Figure 48: Actual Oil Level vs. Time (Raw Data)

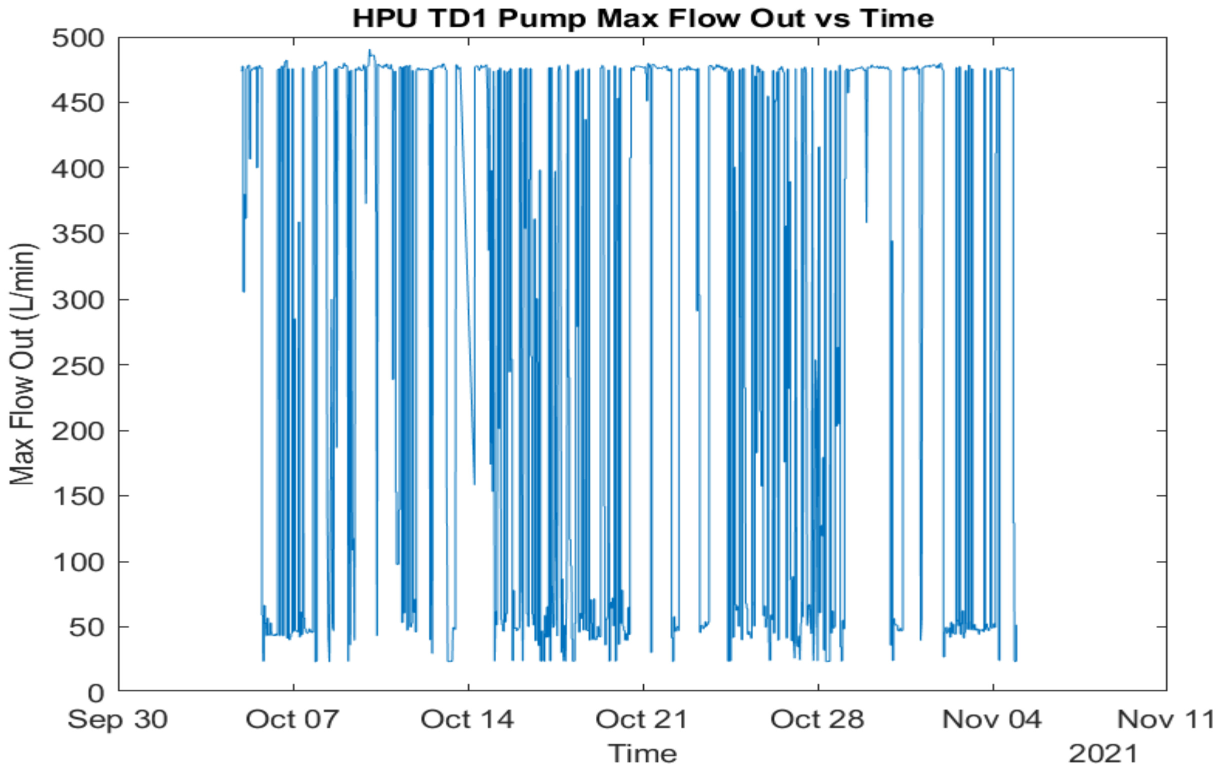

Figure 49: TD1 Max Flow Out vs. Time (Raw Data)

8.2.3 TDS Raw Data

The data coming from the top drive system, which is electrical on this rig, was very important for this study because it contained the goal failures of the predictive model. This data that will be presented in the following series of plots corresponds to the raw data that were cleaned in further steps and used to predict the number of low-pressure alarms of the blower. The blower is a subcomponent of the top drive system which is responsible for the cleaning and cooling of the top drive system. It has a critical role in removing all kinds of debris and dust by blowing cold air at high pressures, hence maintaining acceptable working conditions for the top drive. Every time the pressure of the blower is low, an alarm is triggered to alert that there is a malfunction in this device. This alarm was triggered a lot during this drilling activity. The sensors mounted on the different components of the top drive system collected this data within a period of 54 days with a frequency of one new value per hour. These sensors collected the following parameters, the electric motor's average and max current (Figures 57 and 51), the electric motor's average and max torque (Figures 56 and 52), the blower working hours (Figure 53), the electric motor's average and max RPM (Figures 54 and 62), the number of low-pressure alarms of the blower (Figure 55), the lube oil temperature (Figure 58), the hydraulic brake temperature (Figure 59), the electric motor working hours (Figure 60) and the hydraulic brake working hours (Figure 61).

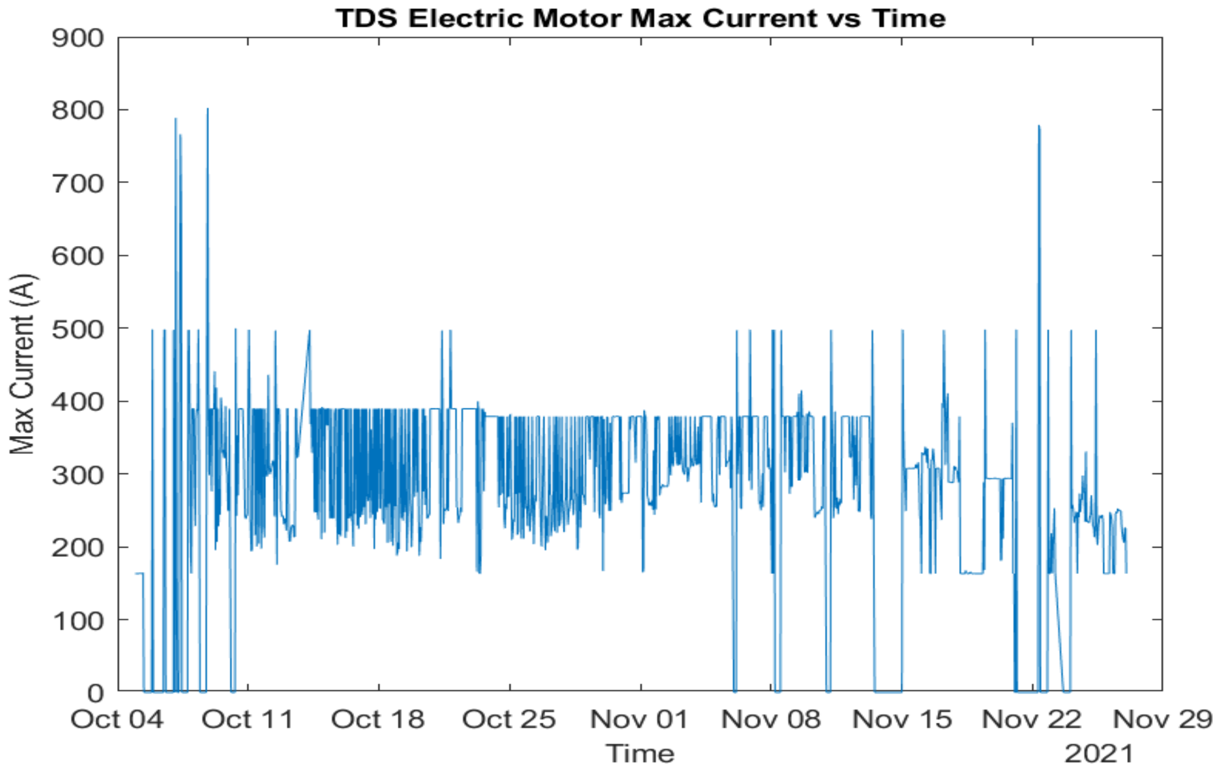

Figure 50: Max Current vs. Time (Raw Data)

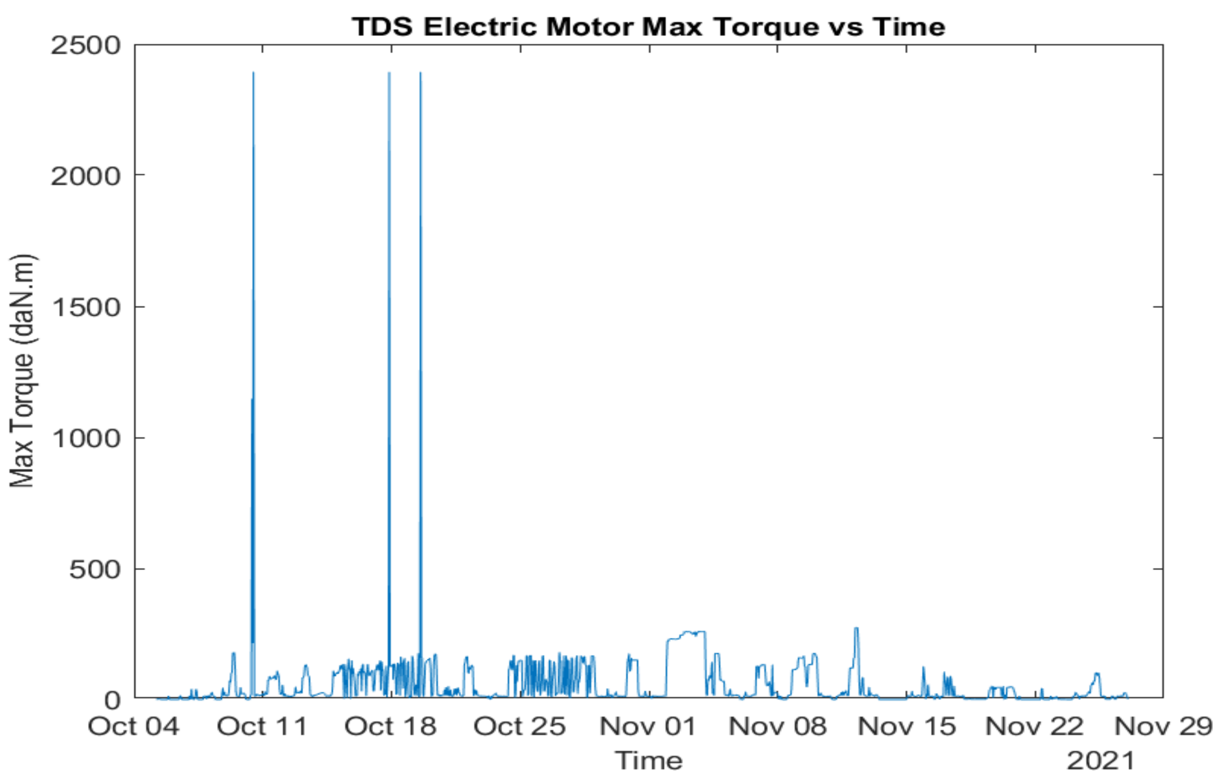

Figure 51: Max Torque vs. Time (Raw Data)

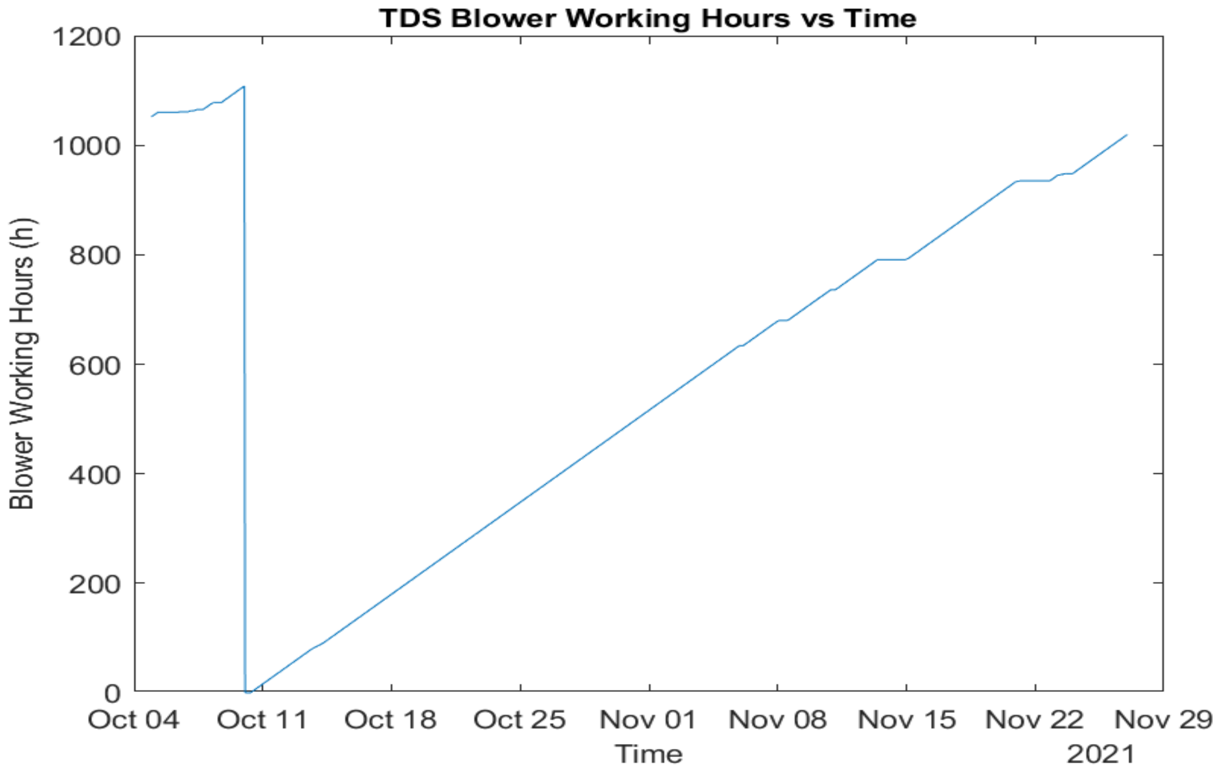

Figure 52: Blower Working Hours vs. Time (Raw Data)

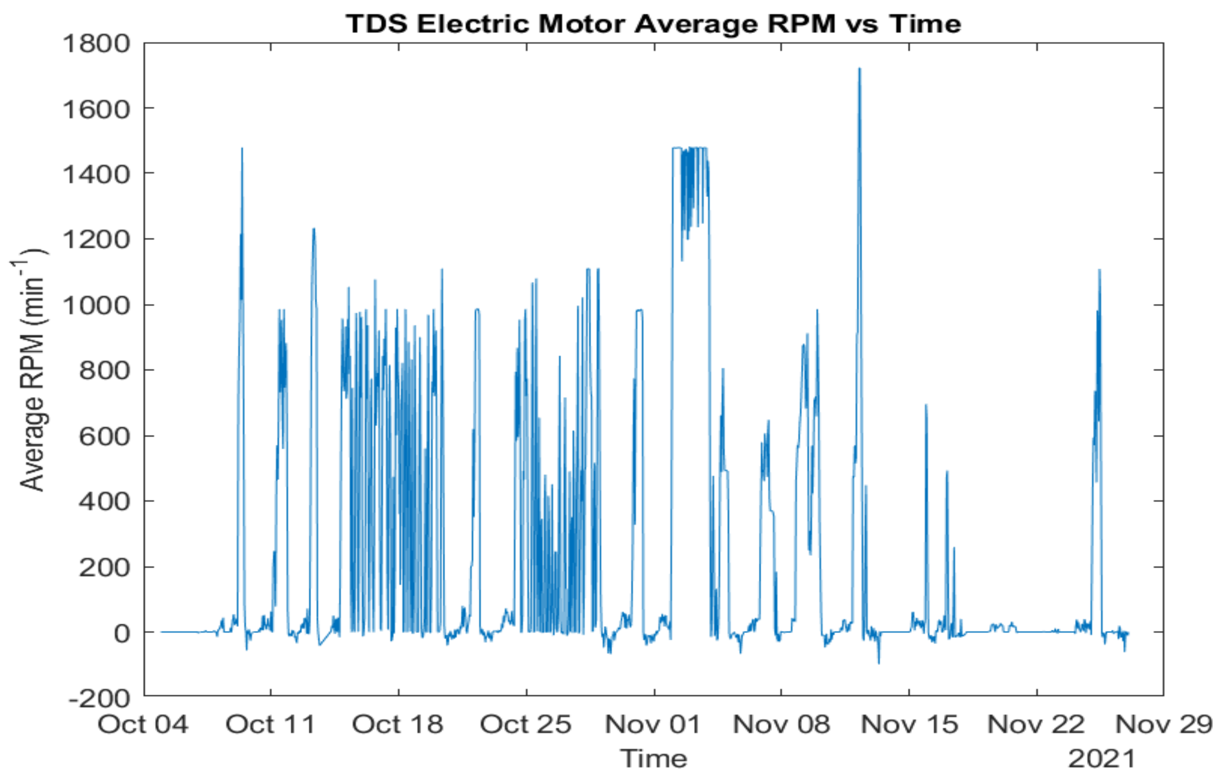

Figure 53: Average RPM vs. Time (Raw Data)

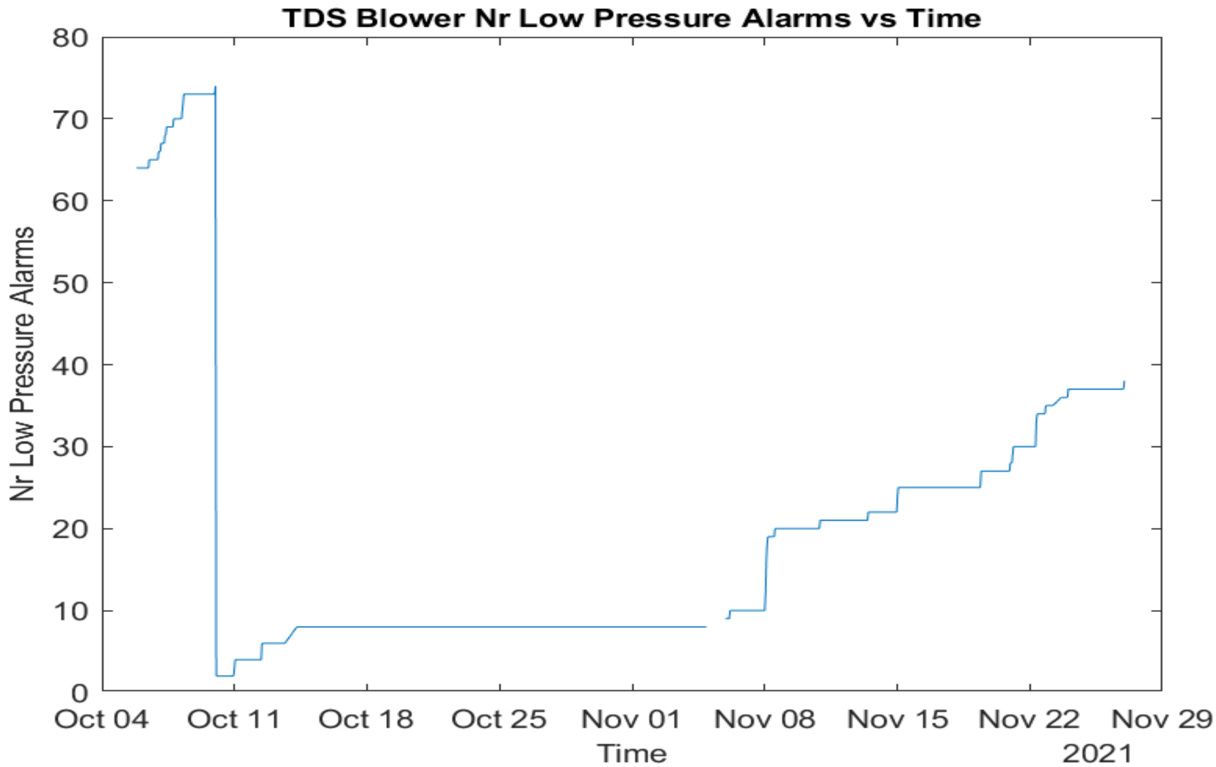

Figure 54: Blower Number of Low-Pressure Alarms vs. Time (Raw Data)

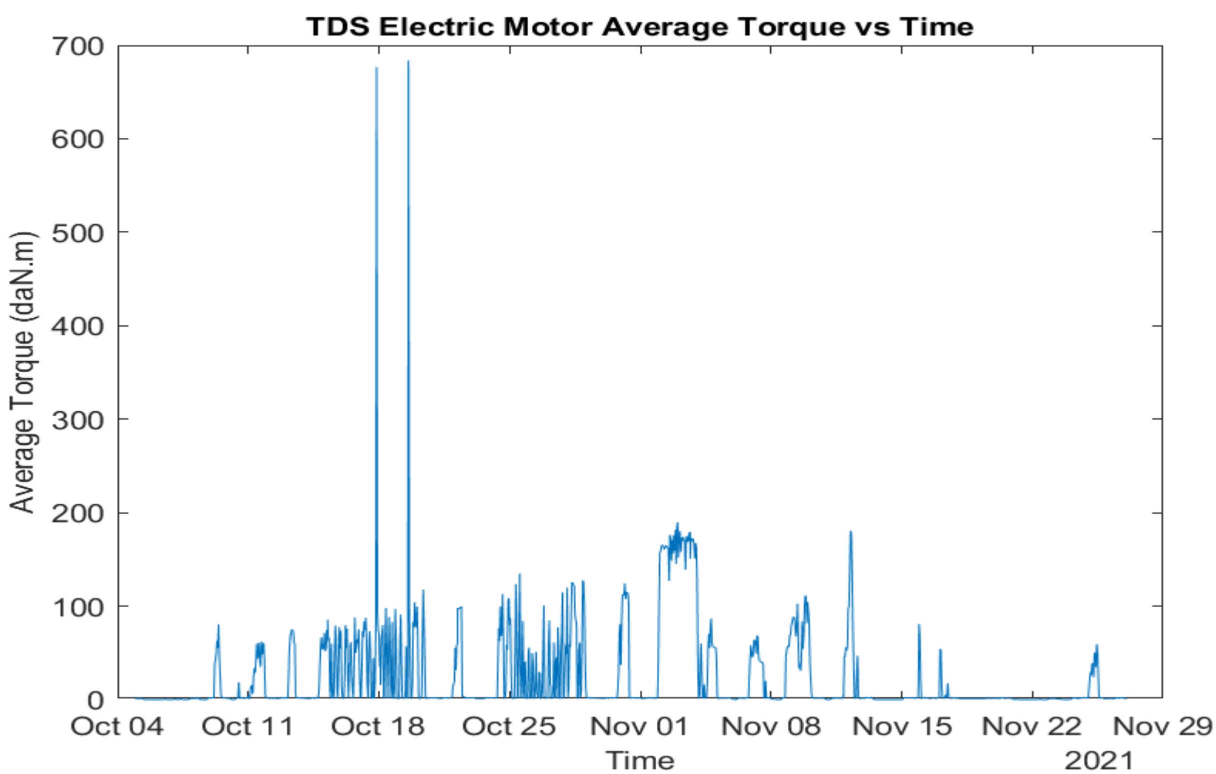

Figure 55: Average Torque vs. Time (Raw Data)

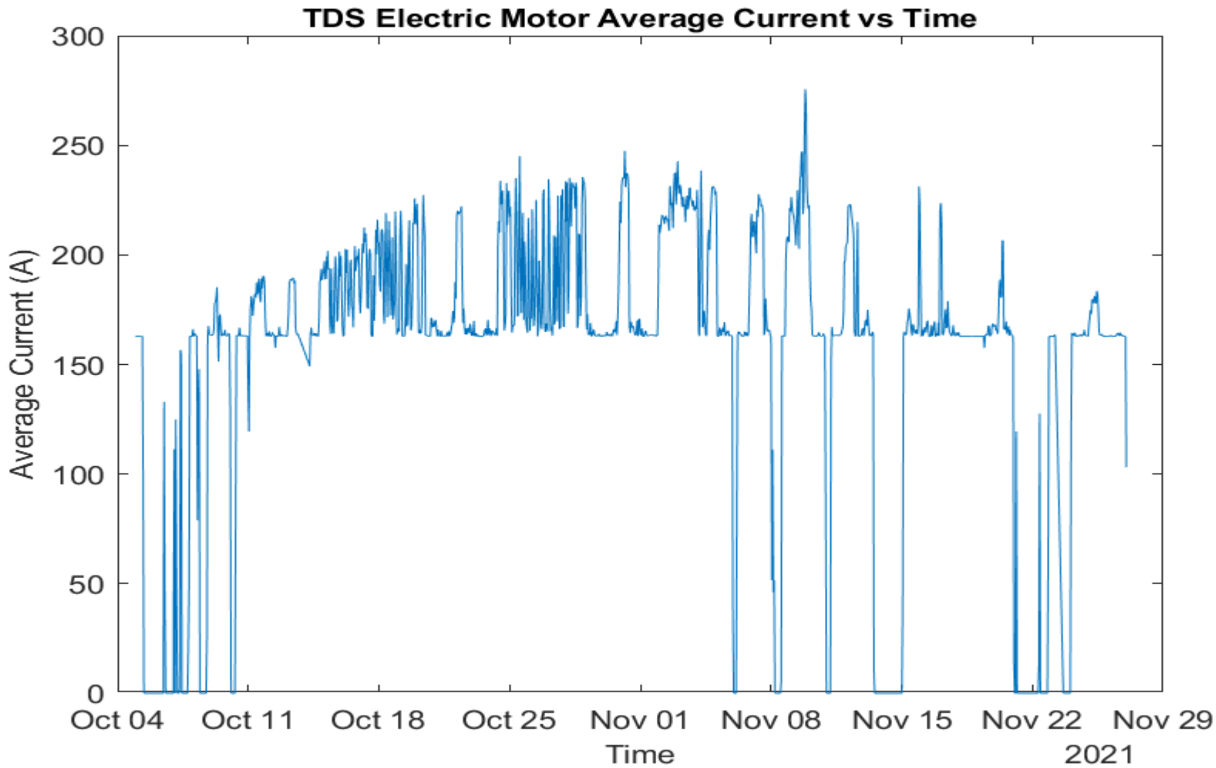

Figure 56: Average Current vs. Time (Raw Data)

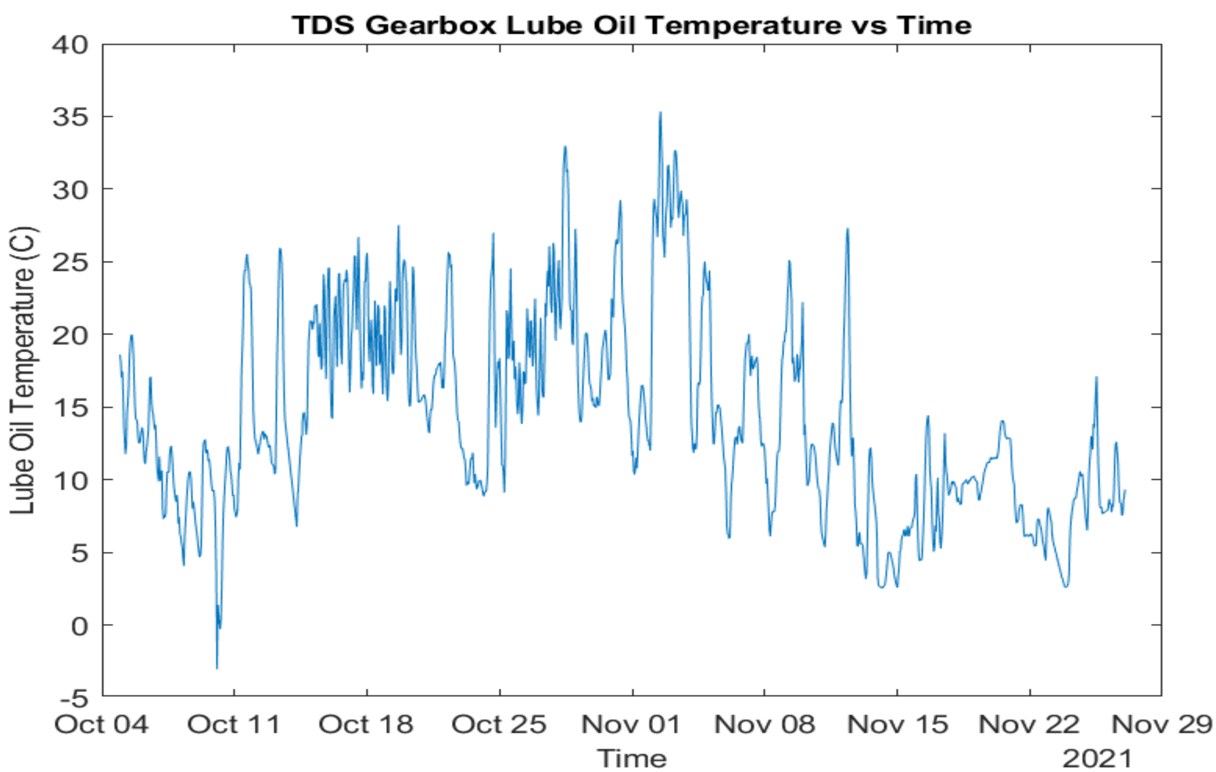

Figure 57: Lube Oil Temperature vs. Time (Raw Data)

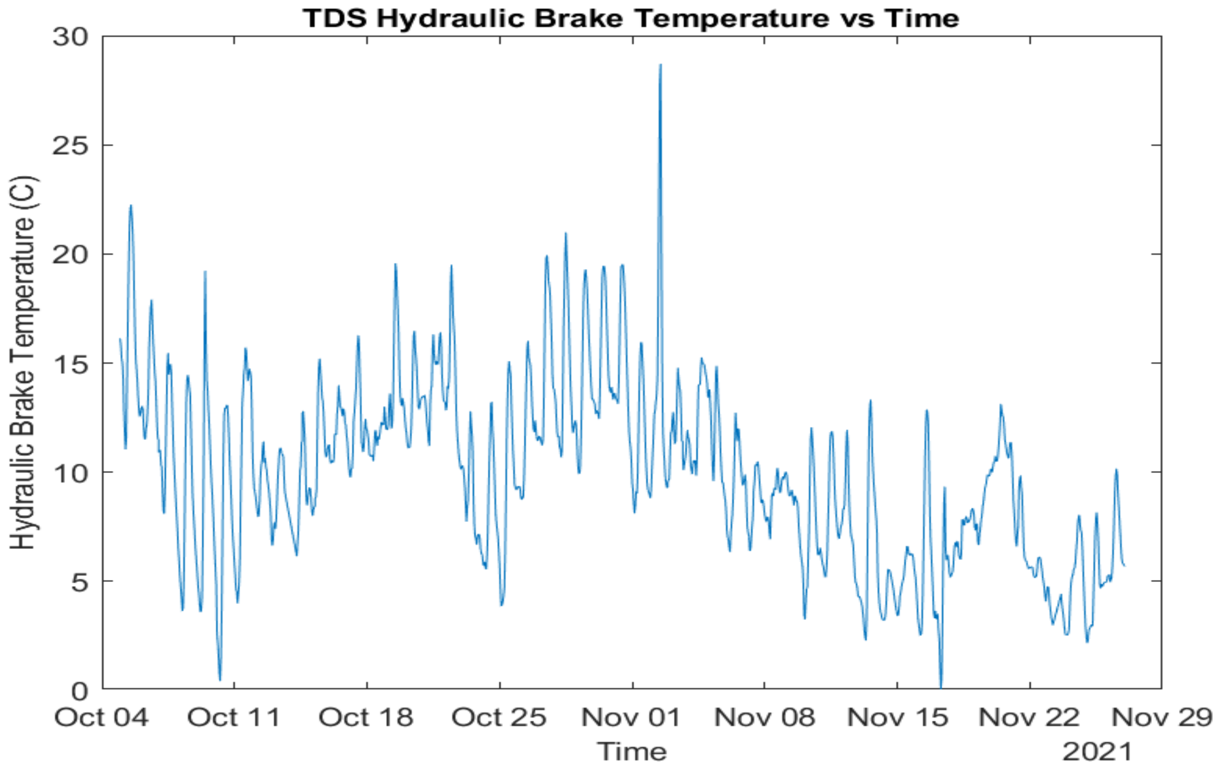

Figure 58: Hydraulic Brake Temperature vs. Time (Raw Data)

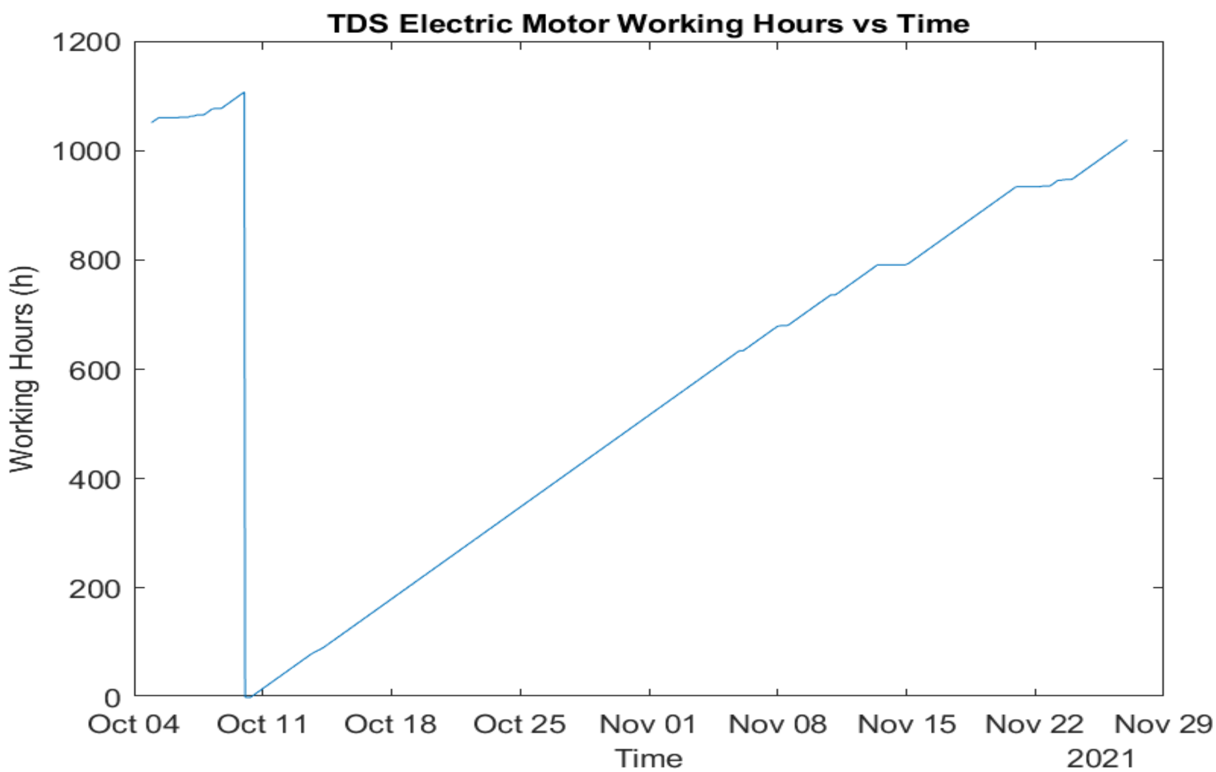

Figure 59: Electric Motor Working Hours vs. Time (Raw Data)

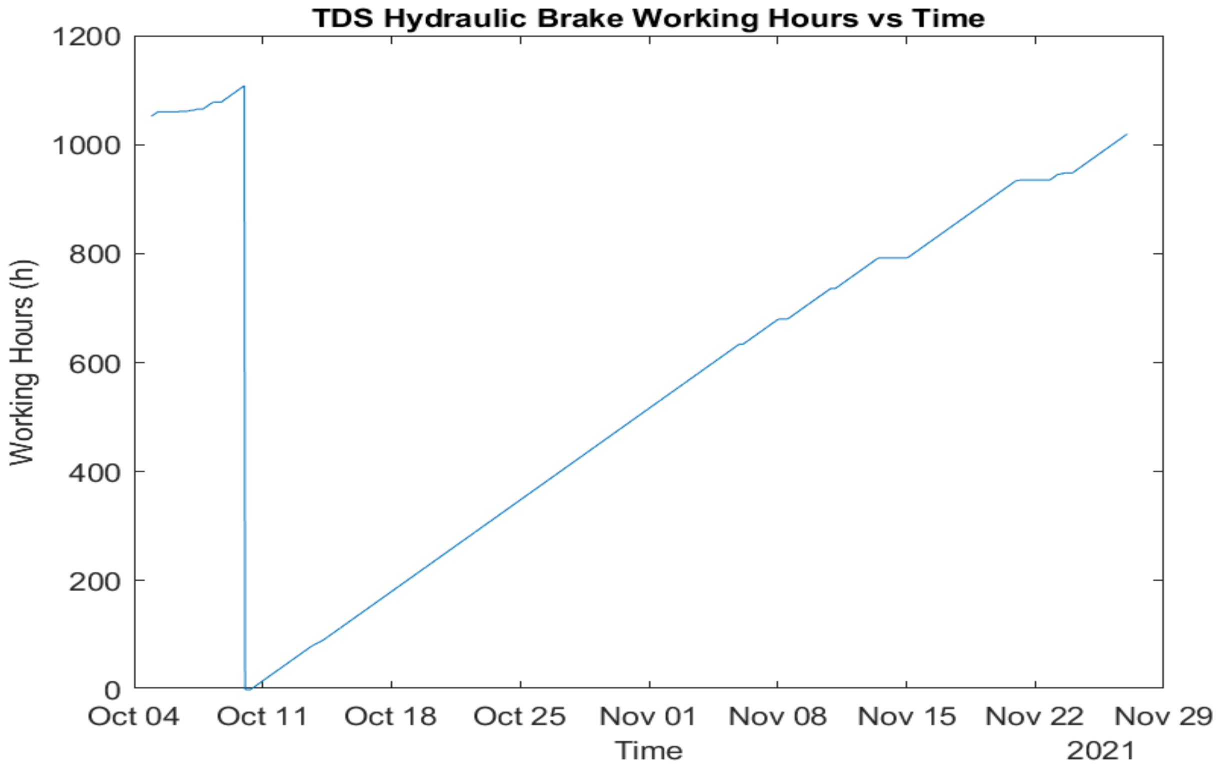

Figure 60: Hydraulic Brake Working Hours vs. Time (Raw Data)

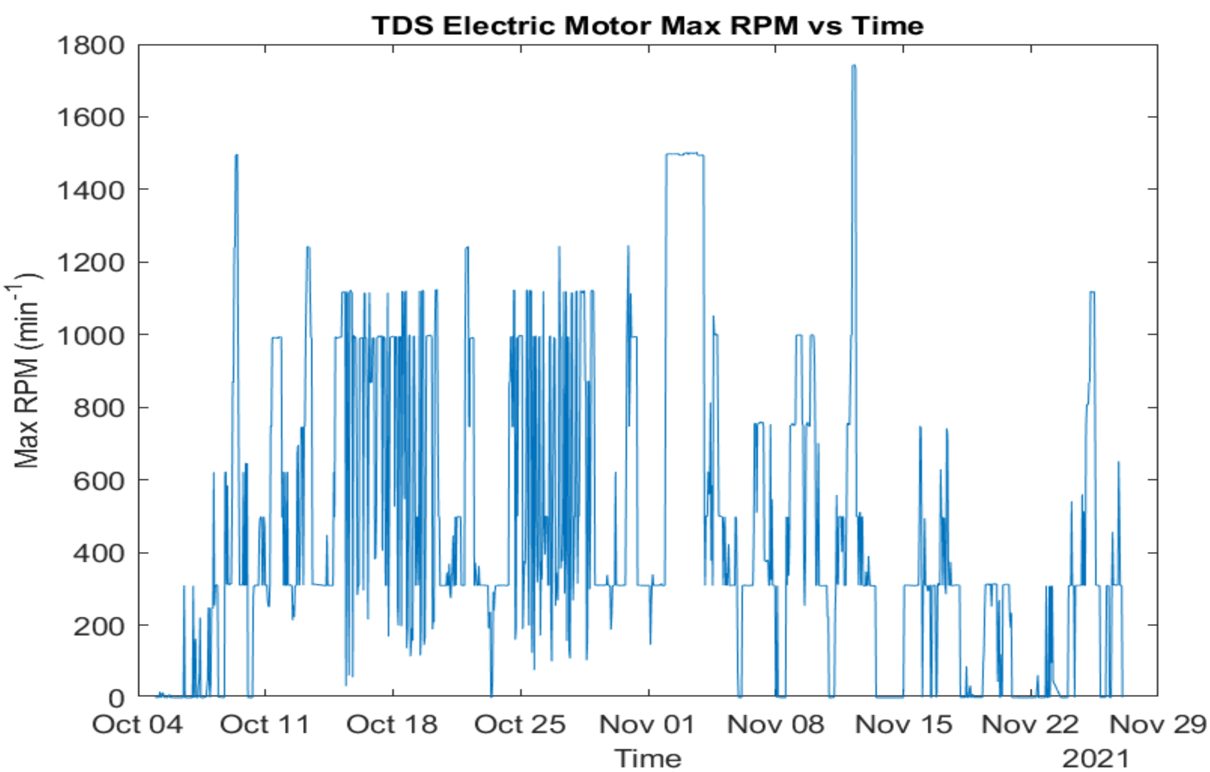

Figure 61: Max RPM vs. Time (Raw Data)

8.3 Data Cleaning

After having collected and stored all the initial raw data from all these sensors, the next and most important step of the work is data cleaning. At this stage of the process, all the plots showcased previously are seen in more detail one at a time to remove or replace any wrong values or behaviors, fill missing values and make sure that all the data is at the correct timestamp. Incorrect values do not mean failures since the goal of predictive models is to predict these failures but it means any abnormal behavior of the data recorded that could be linked to a sensor's reading error. The criticality of this procedure is linked to the reliability of the generated model at the end because if the model was trained with false data even if it gives a high accuracy value, this high value will be linked to a false prediction causing all the work to be fallacious. In that case, a low but correct accuracy is always preferred over a high but rather unreliable accuracy. In the following subparagraphs, every parameter will be discussed to show what and if any data cleaning was used on it in addition to the new and final plots of the data after all the filtering was done.

8.3.1 DAS Data Cleaning

For the data coming from the data acquisition system, where one common thing was done to all the parameters and that is the removal of all the data corresponding to the first seven days. This was done due to its bad quality since for the most part of that time the rig was not working which means these measurements were not useful in any way for this study and there were a lot of missing rows for long periods that could not be filled. Now every parameter will be discussed by itself:

- For the bit depth, one huge peak that was hiding all the real variation of this sensor's data was removed and replaced by a value calculated using linear interpolation. After that, some discontinuities in the plots and small abnormal peaks appeared which were also removed and replaced by using the same technique (Figure 63).

- For the well depth, similar to the bit depth, the huge peak was removed by using linear interpolation. Then there were only some missing values but nothing abnormal, these absent measurements were filled with the same technique (Figure 64).

- For the surface RPM no additional data cleaning was necessary (Figure 65).

- For the SPM of pumps 1, 2, and 3, outliers were removed and replaced by averaged values of the strokes per minute corresponding to the values in a close time interval. In addition to that sudden peaks that appeared where the pumps were not working were replaced by the value 0. Some negative values were all replaced by 0 (Figures 66, 67, and 68).

- For the surface torque, no additional data cleaning was performed (Figure 69).

- For the WOB first of all, all the negative values were replaced by 0 since the weight cannot be negative. Second of all, whenever the bit depth and the well depth were not equal, the WOB values corresponding to that were set to 0 (Figure 70).

- For the hook load, all the measurements lower than 18 t were set to 18. The reason behind doing that is, the weight of the top drive and whatever the mast is holding without any extra weight is 18 t (Figure 71).

- For the average ROP of one hour, the max ROP of one hour, and the instantaneous ROP only really small values were set to 0 by checking that the well depth was not varying at these points (Figures 72, 73, and 74).

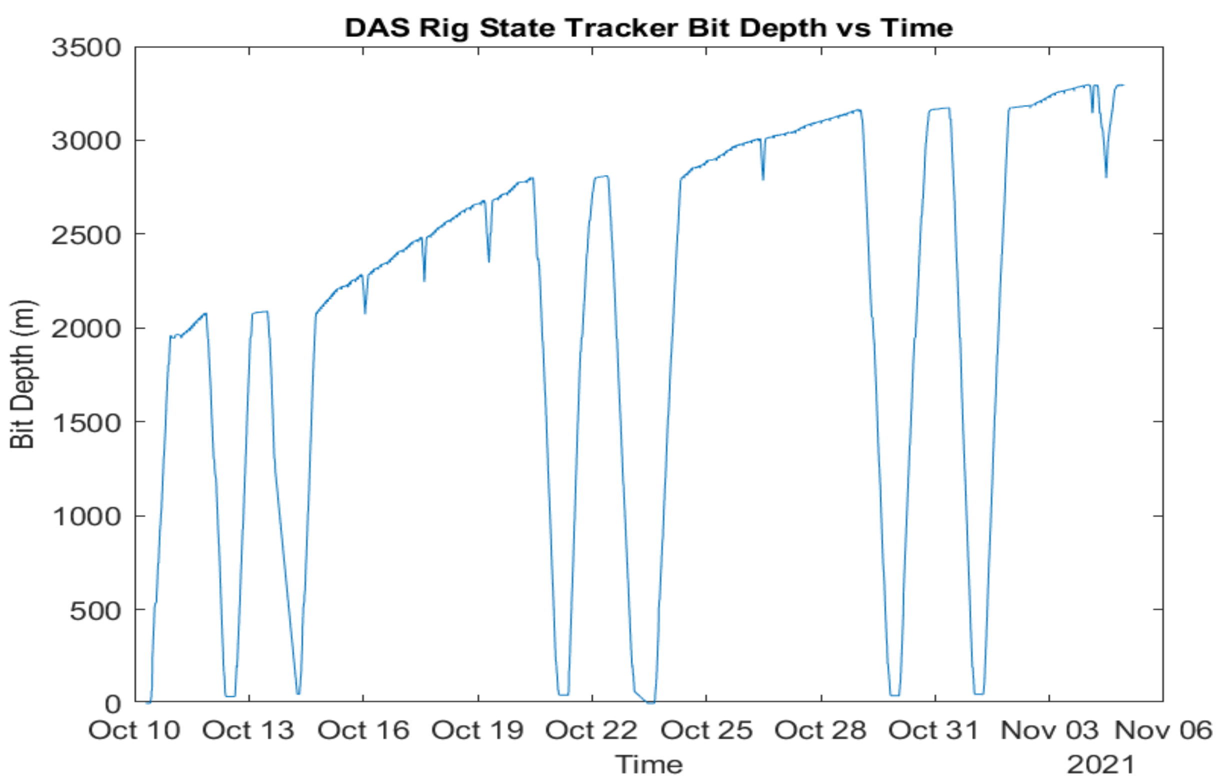

Figure 62: Bit Depth vs. Time (Cleaned Data)

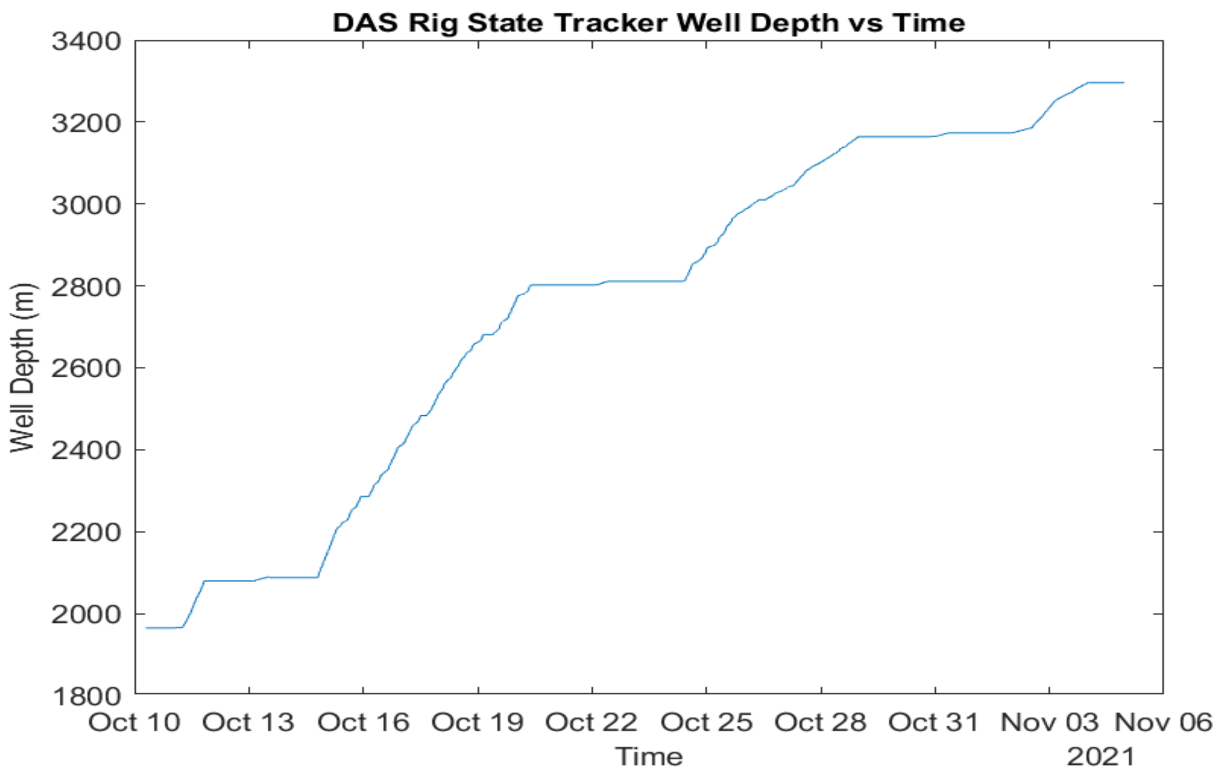

Figure 63: Well Depth vs. Time (Cleaned Data)

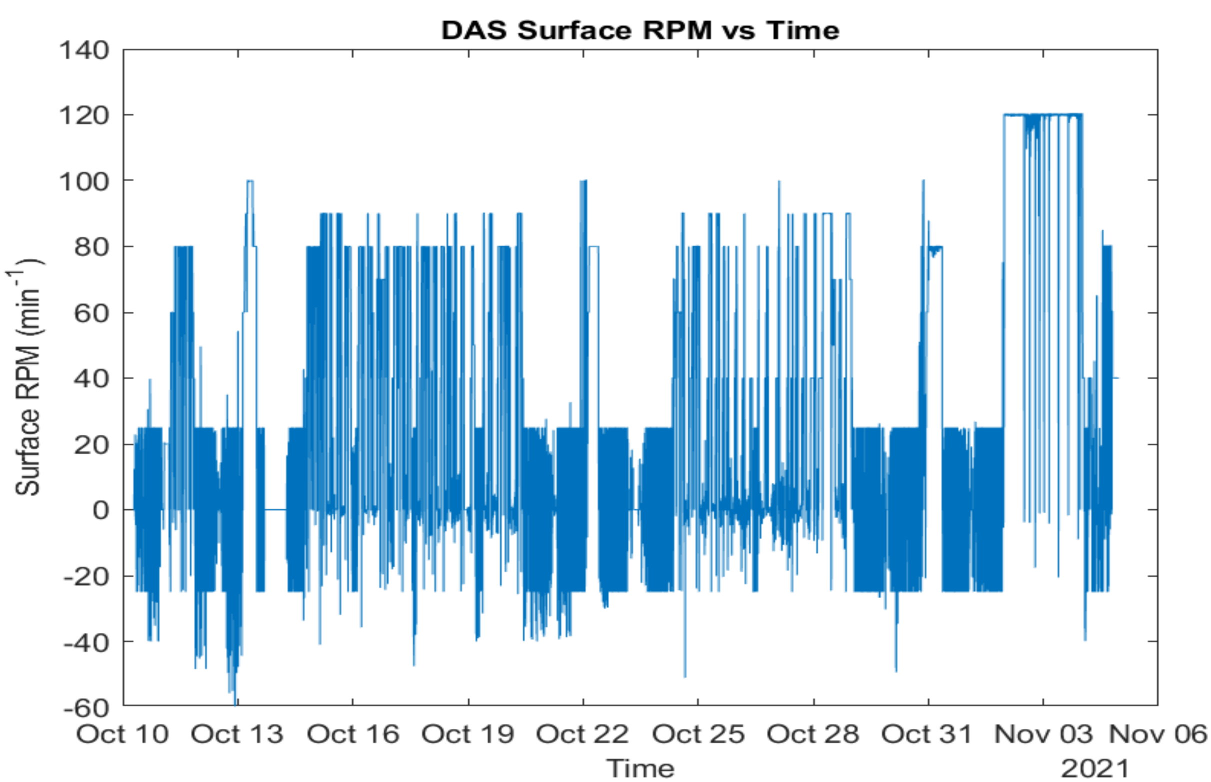

Figure 64: Surface RPM vs. Time (Cleaned Data)

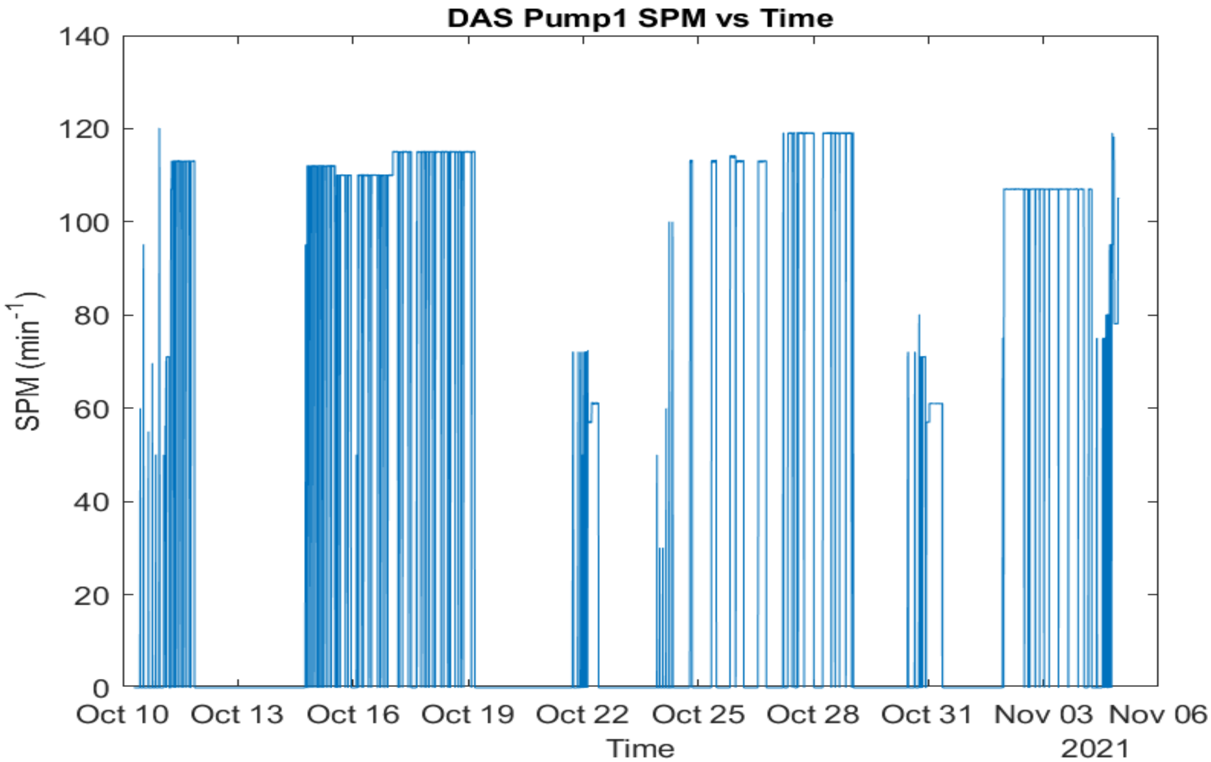

Figure 65: SPM Pump 1 vs. Time (Cleaned Data)

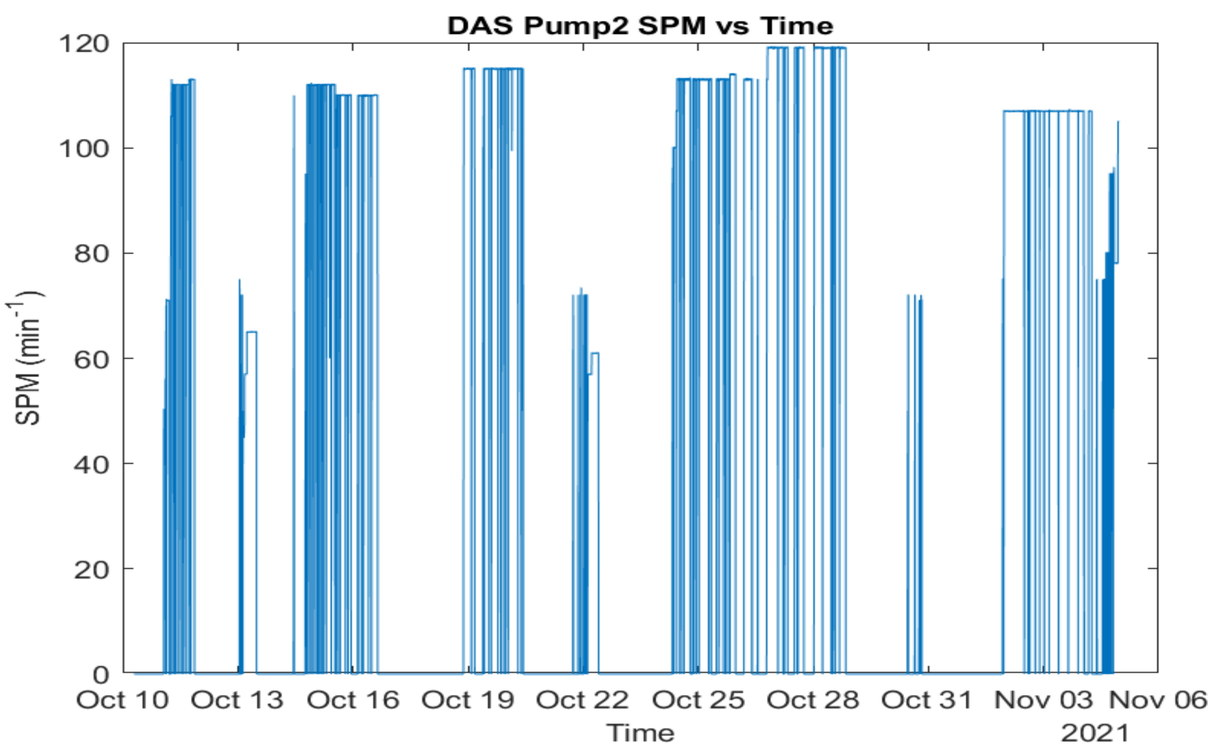

Figure 66: SPM Pump 2 vs. Time (Cleaned Data)

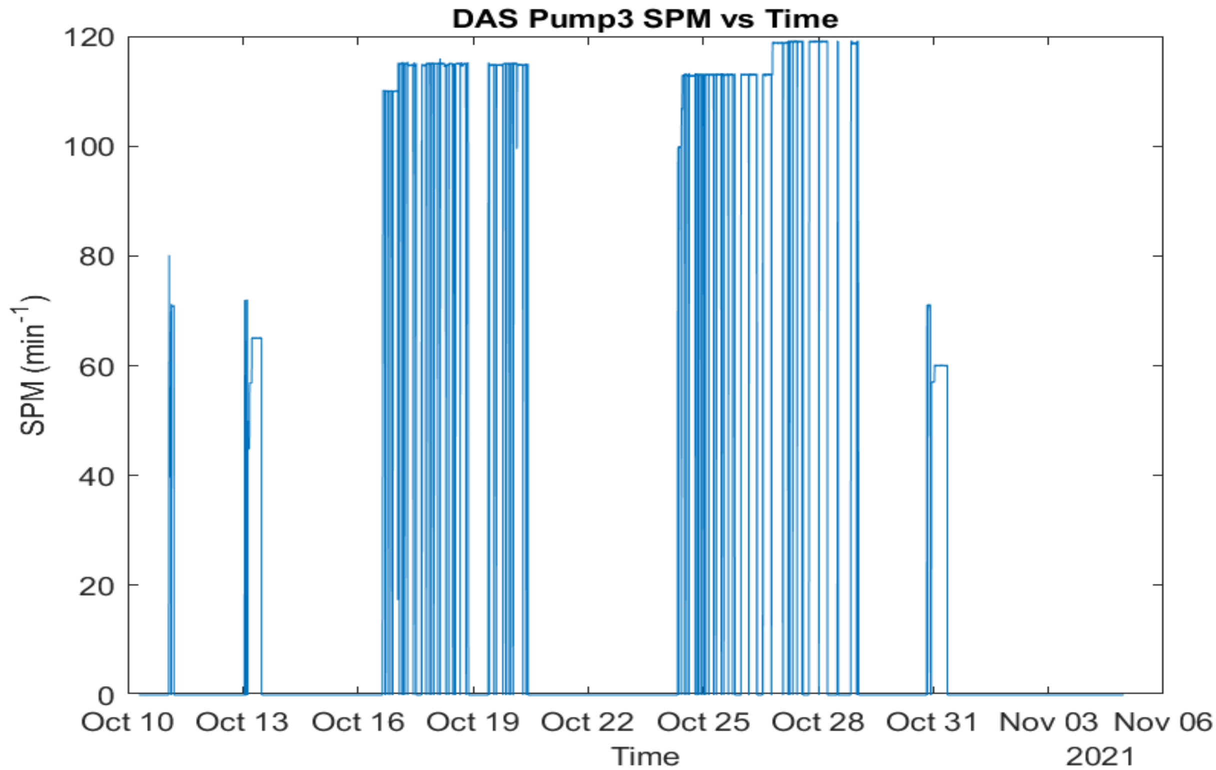

Figure 67: SPM Pump 3 vs. Time (Cleaned Data)

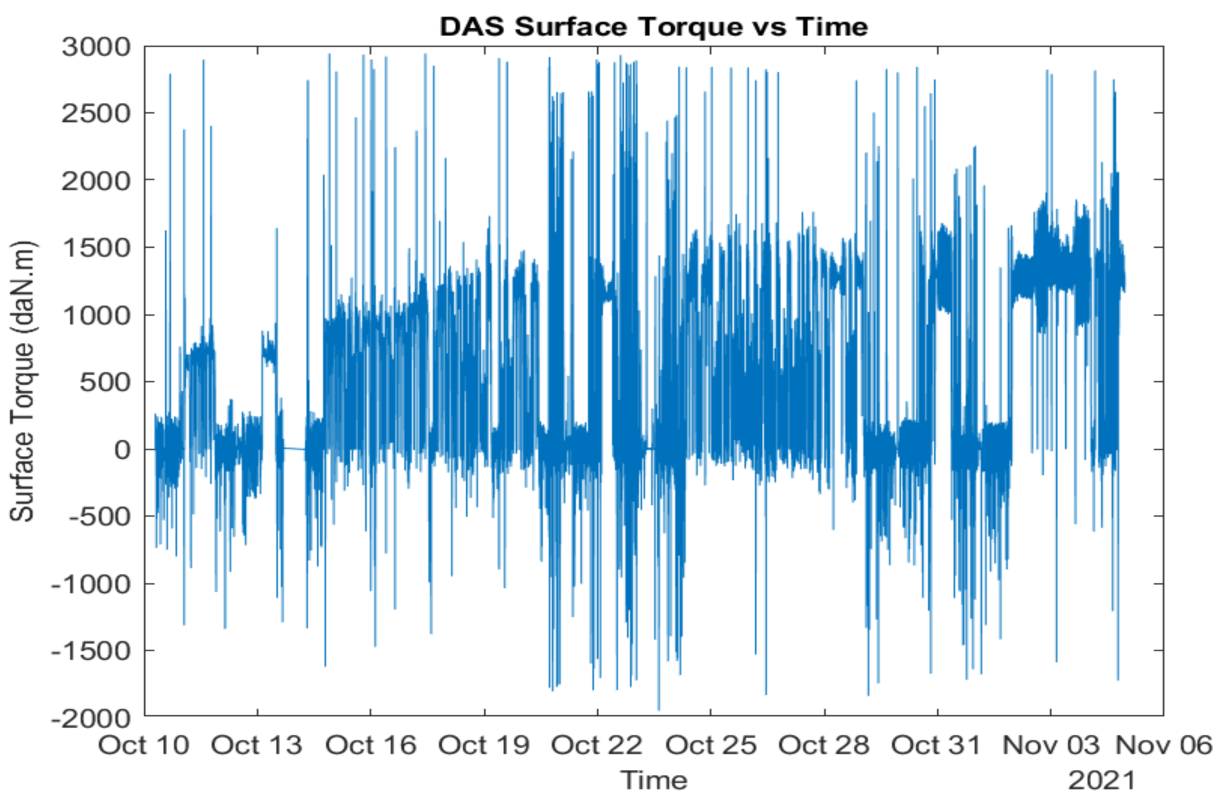

Figure 68: Surface Torque vs. Time (Cleaned Data)

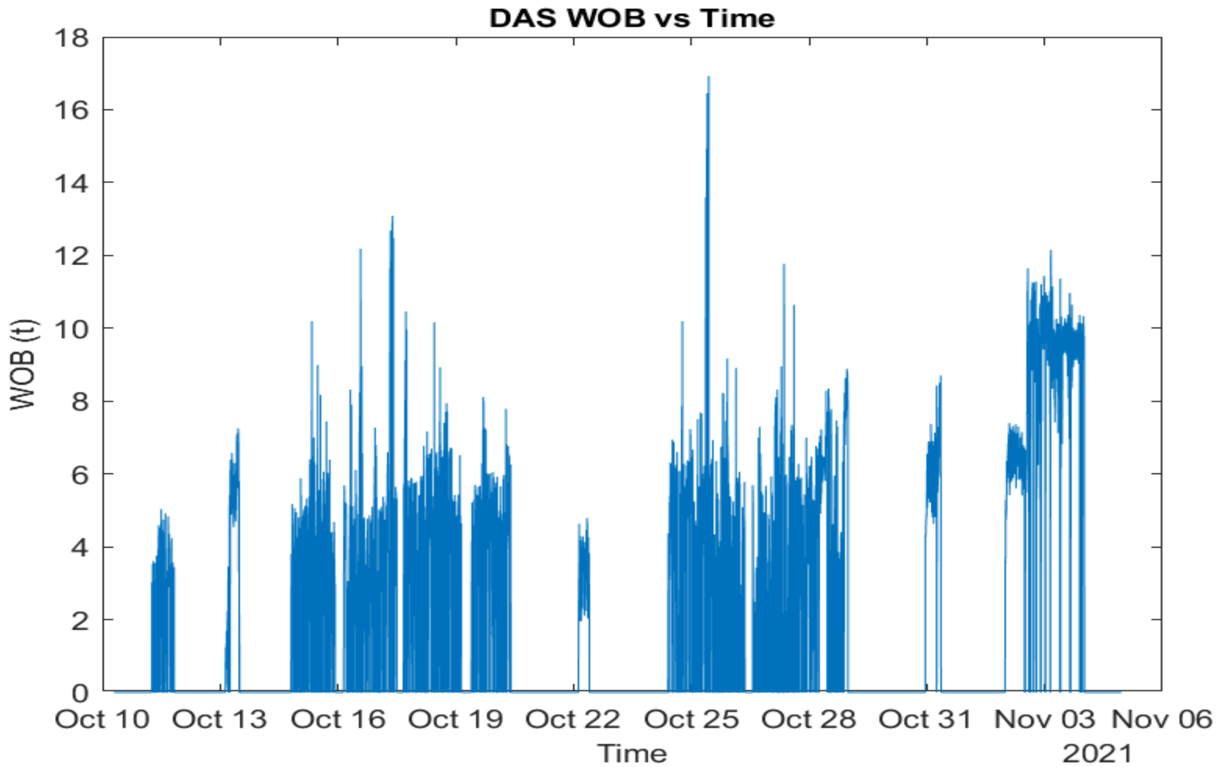

Figure 69: WOB vs. Time (Cleaned Data)

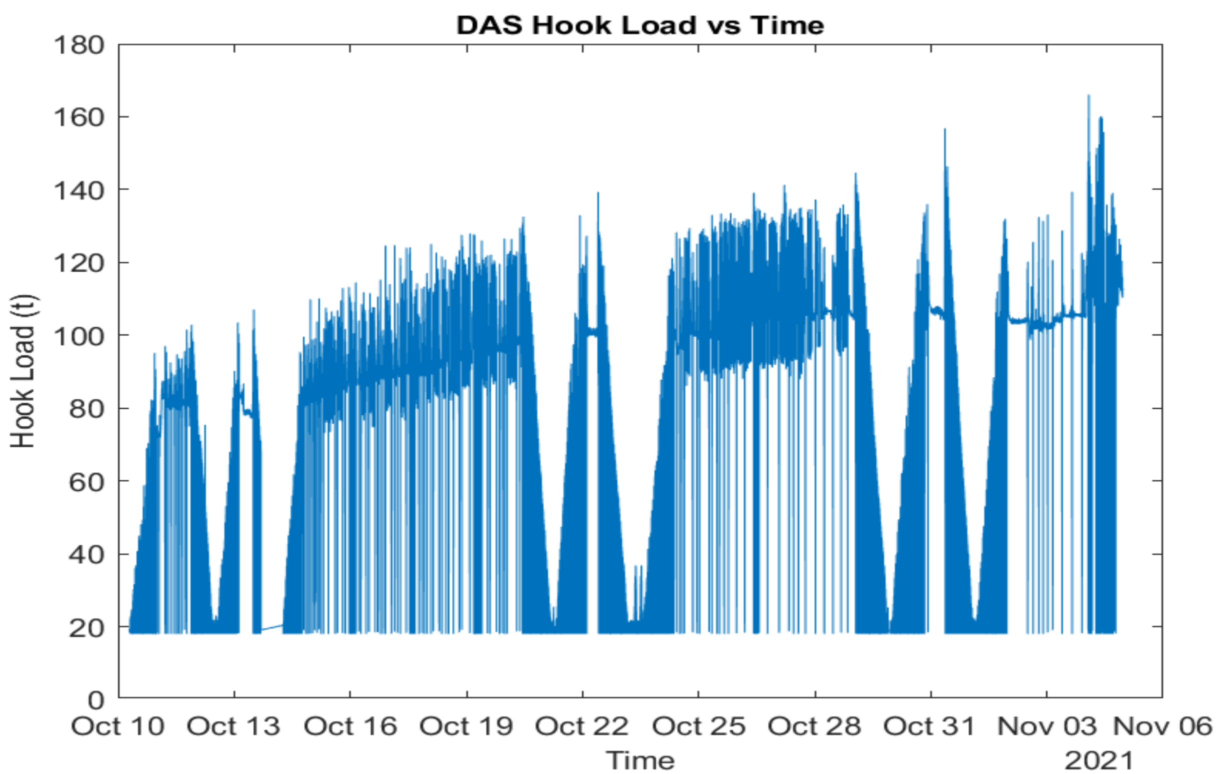

Figure 70: Hook Load vs. Time (Cleaned Data)

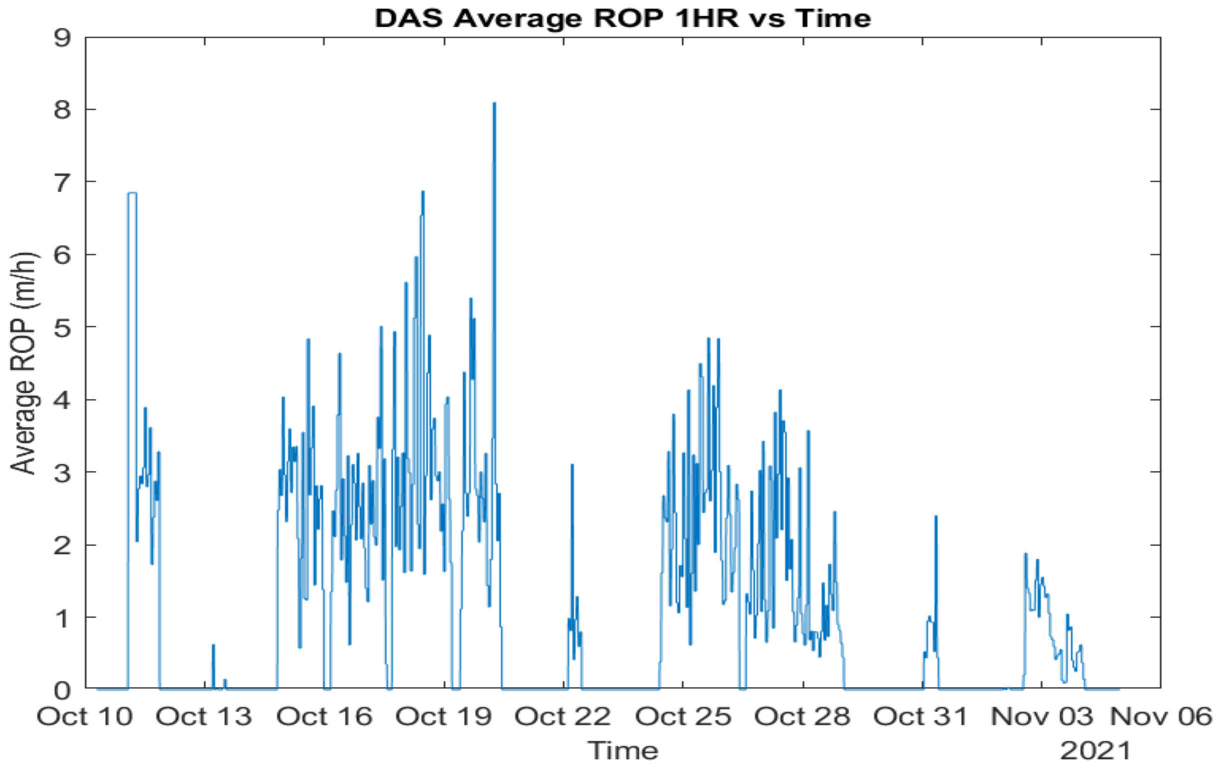

Figure 71: Average ROP every Hour vs. Time (Cleaned Data)

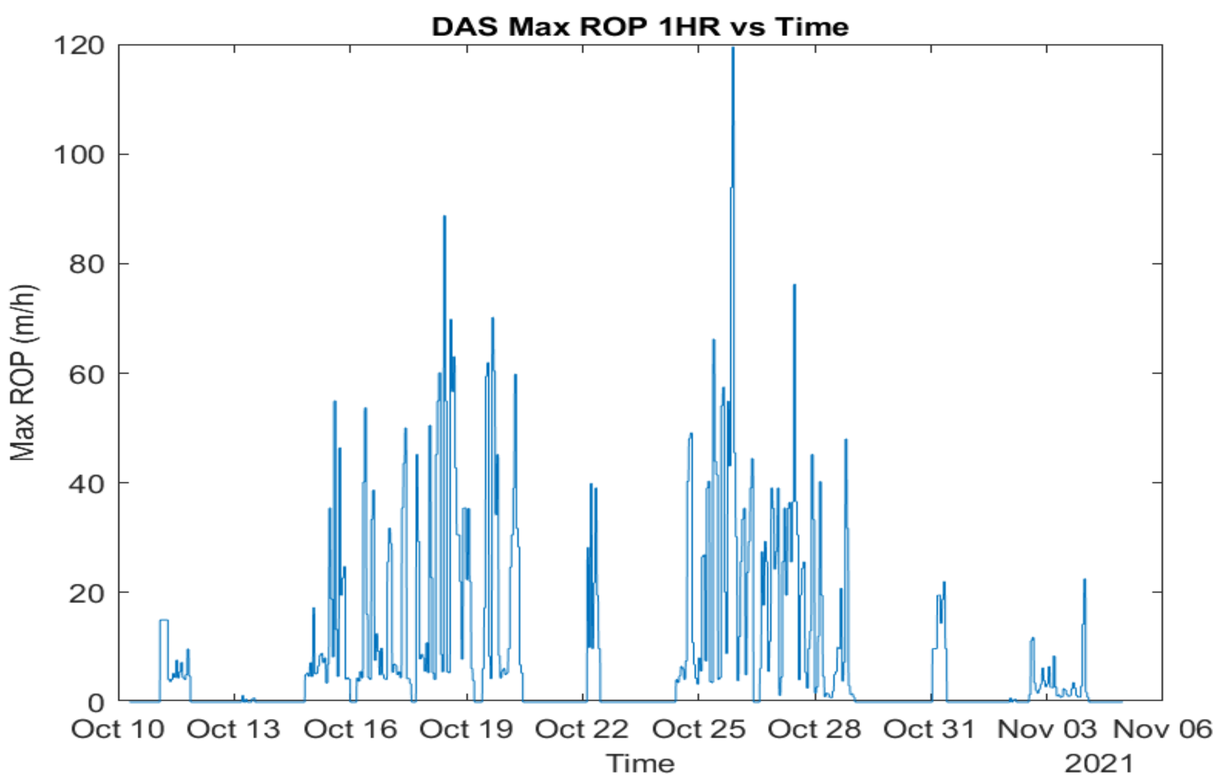

Figure 72: Max ROP every Hour vs. Time (Cleaned Data)

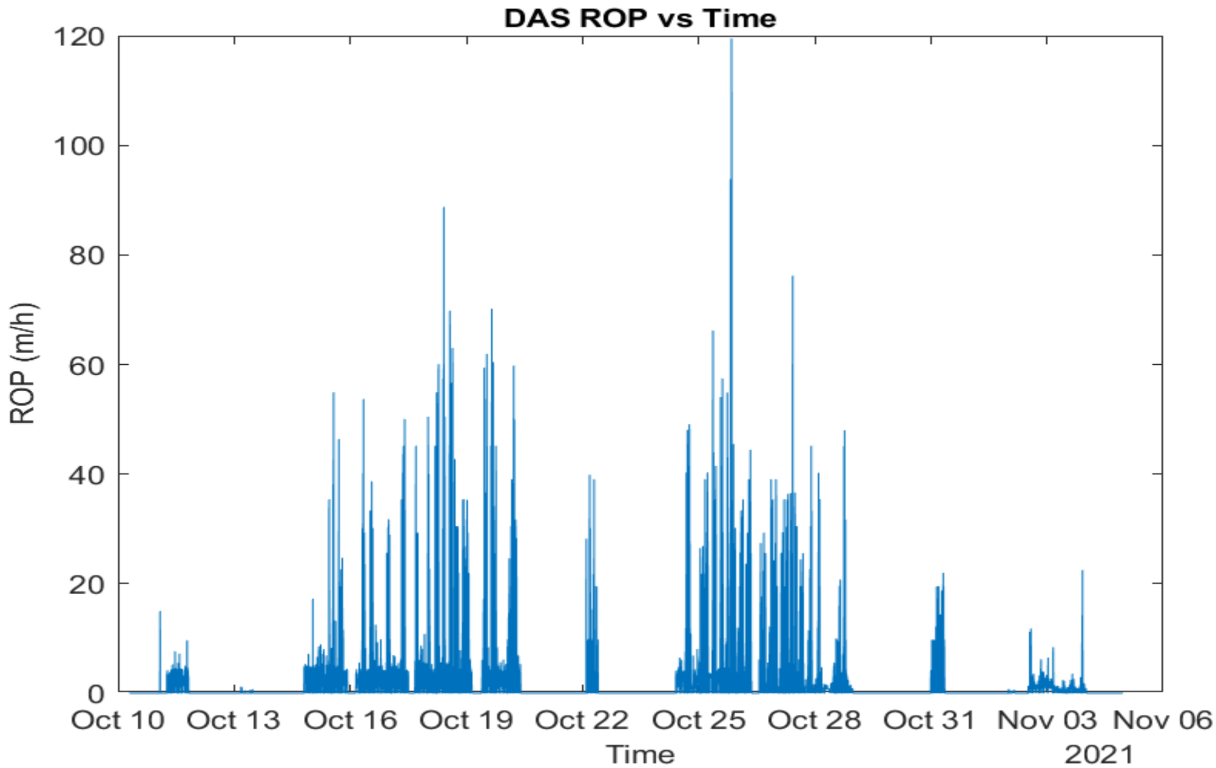

Figure 73: ROP vs. Time (Cleaned Data)

8.3.2 <u>HPU Data Cleaning</u>

Concerning the hydraulic power unit's data, it was the cleanest one and it only required the removal of redundant lines. In addition to that similarly to the DAS data, all the measurements related to the first week were also removed for all the same reasons mentioned earlier plus the working hours of the heat exchanger and the electrical motor were reset after the first seven days. Concerning the average and max power and current, their values were doubled because the sensors were only planted on one of the two motors of the HPU thus this was done to get their real total values (Figures 89, 77, 83, and 94). All the other parameters' cleaned data are shown in Figures 75, 76, 78, 79, 80, 81, 82, 84, 85, 86, 87, 88, 90, 91, 92, 93, 95, 96, 97, 98, 99 and 100.

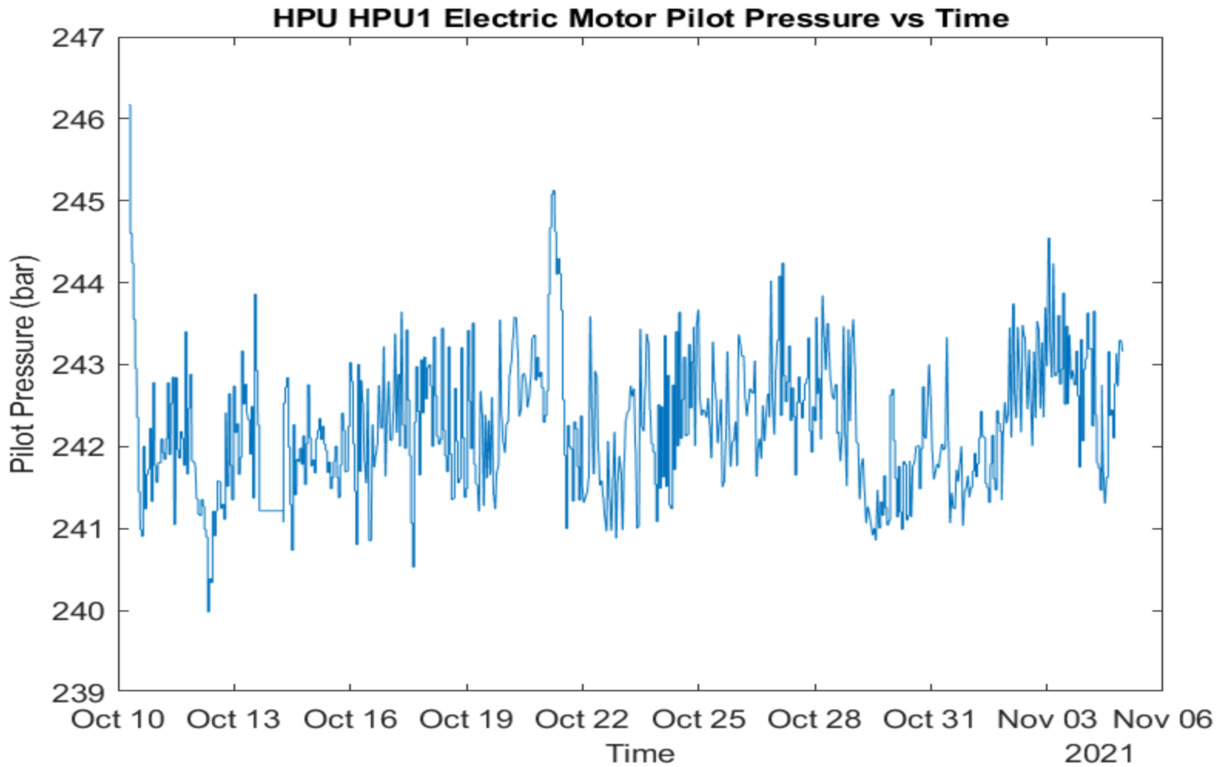

Figure 74: HPU1 Pilot Pressure vs. Time (Cleaned Data)

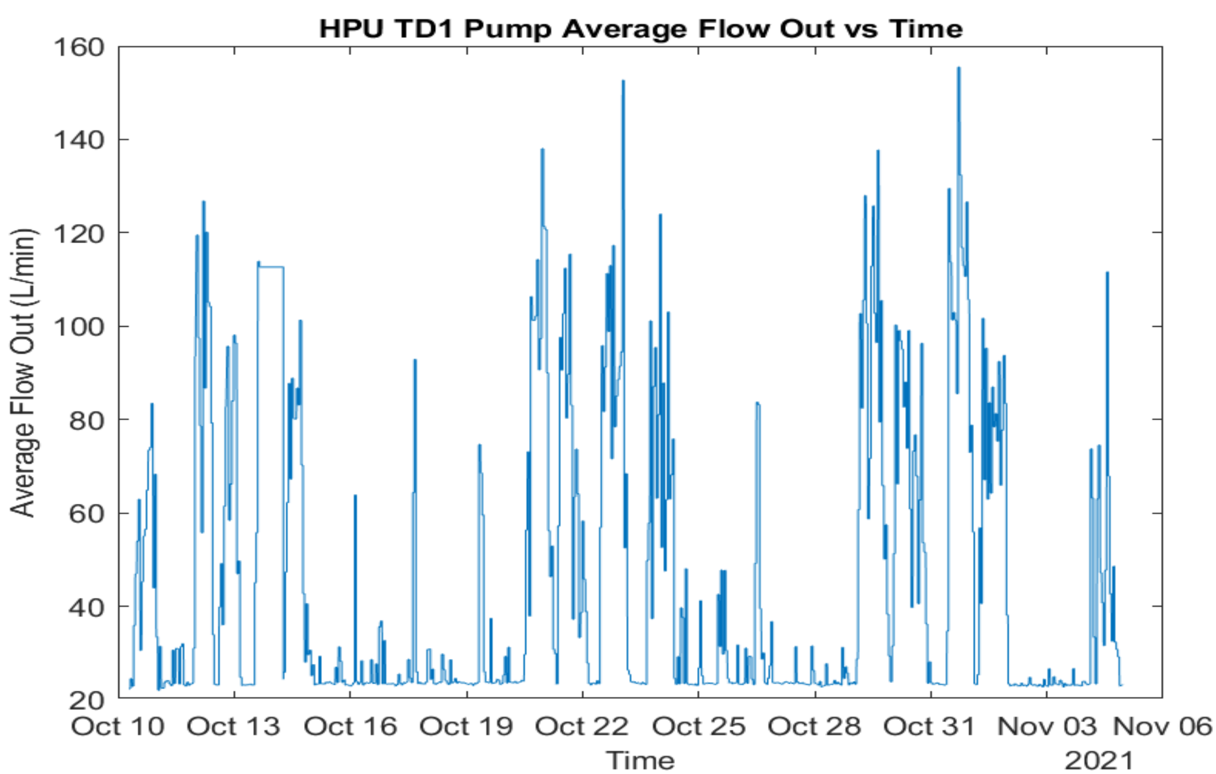

Figure 75: TD1 Average Flow Out vs. Time (Cleaned Data)

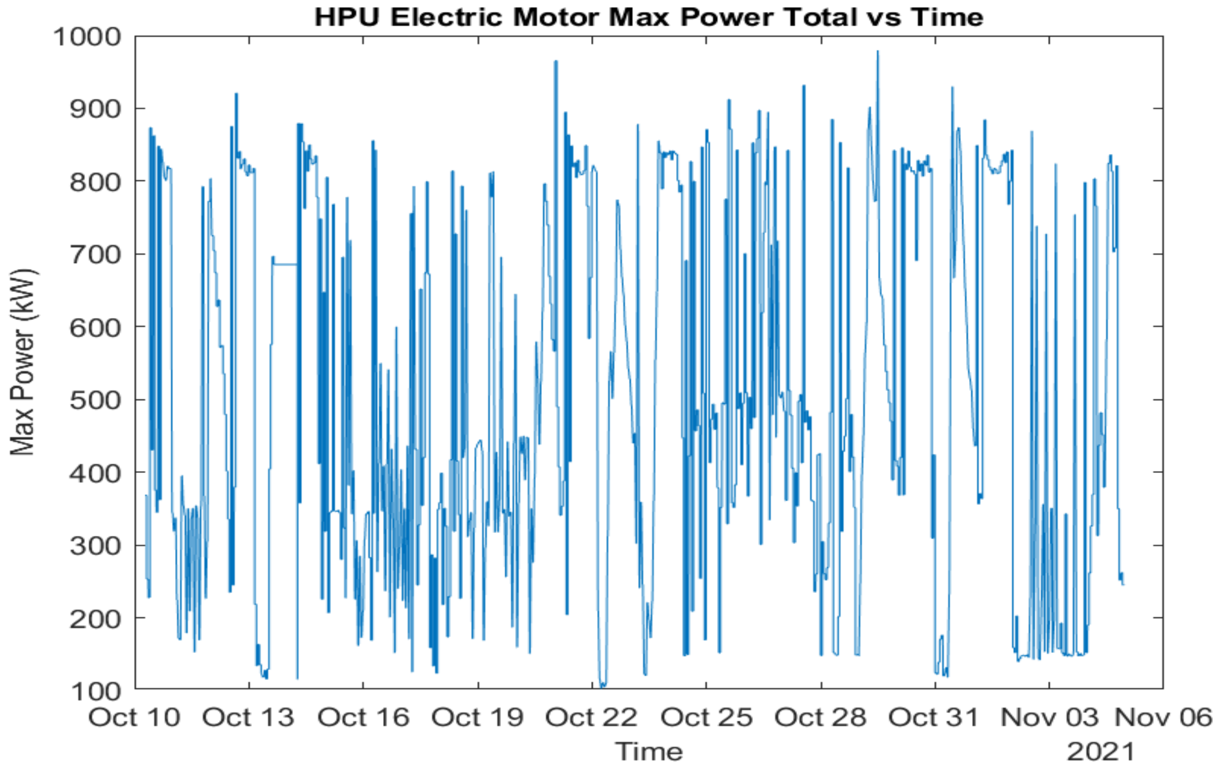

Figure 76: Total Max Power vs. Time (Cleaned Data)

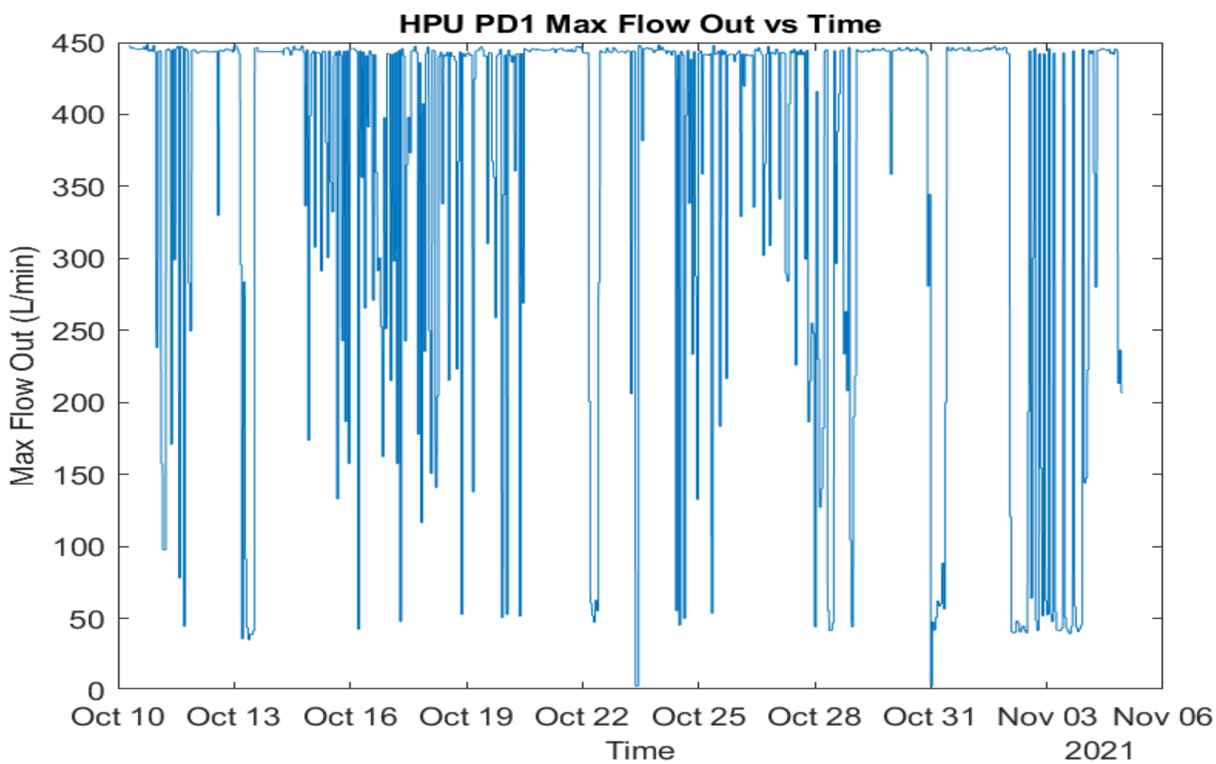

Figure 77: PD1 Max Flow Out vs. Time (Cleaned Data)

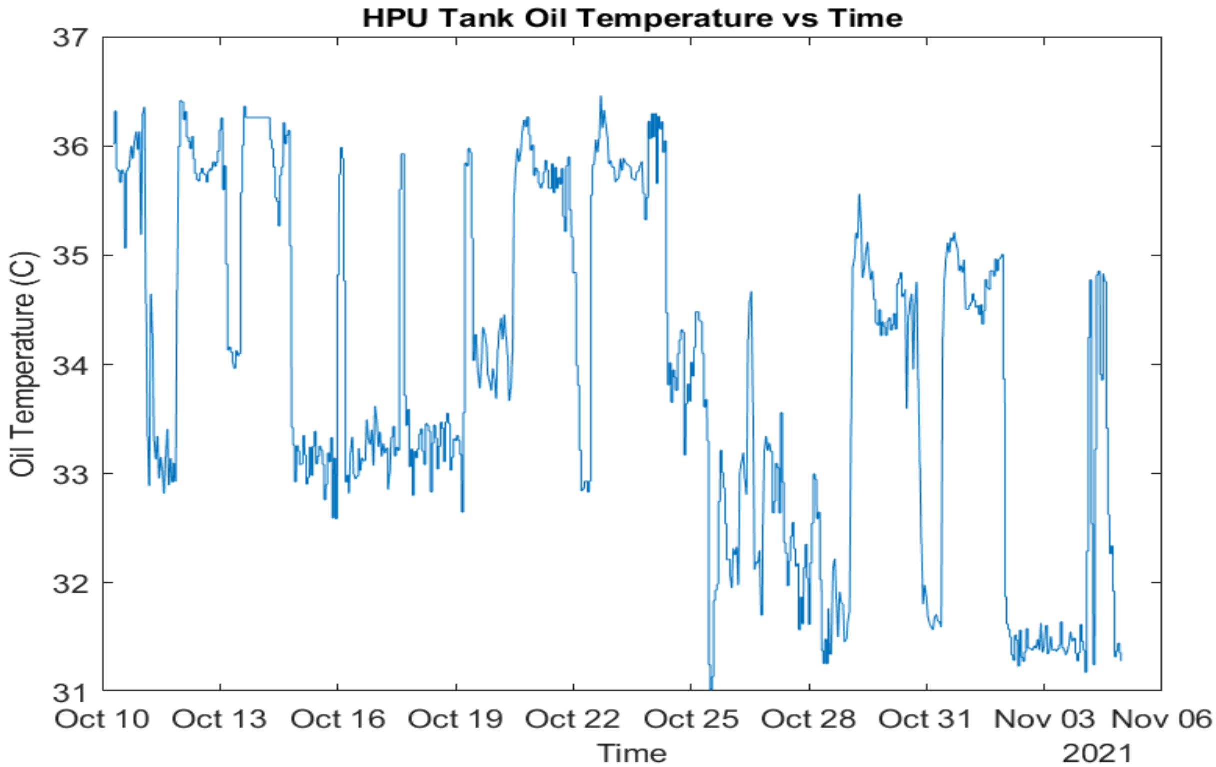

Figure 78: Oil Temperature vs. Time (Cleaned Data)

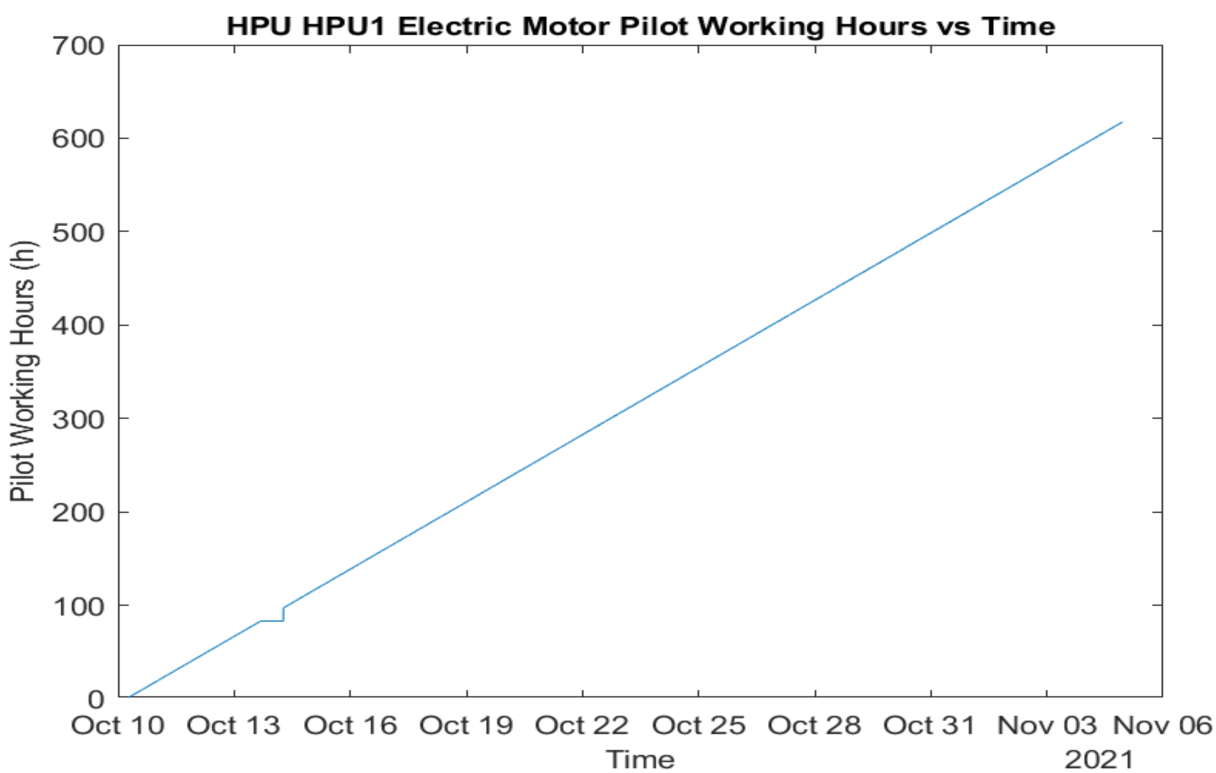

Figure 79: HPU1 Working Hours vs. Time (Cleaned Data)

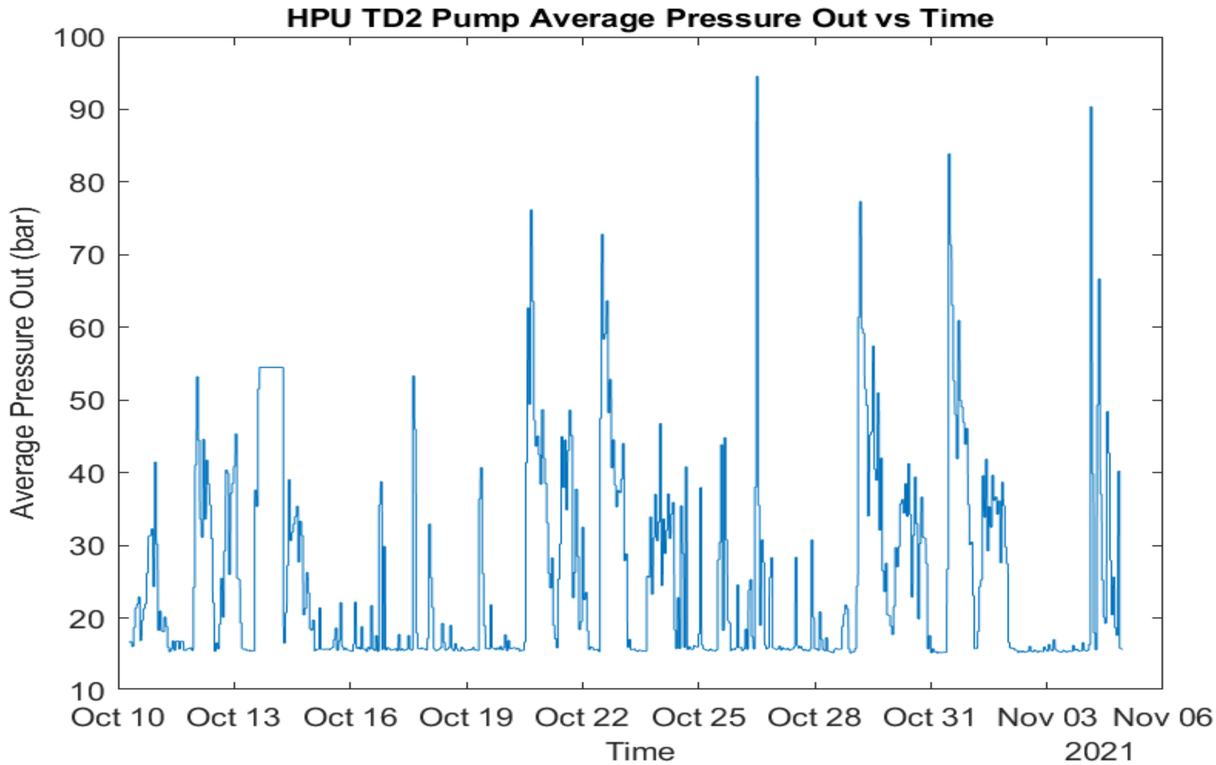

Figure 80: TD2 Average Pressure Out vs. Time (Cleaned Data)

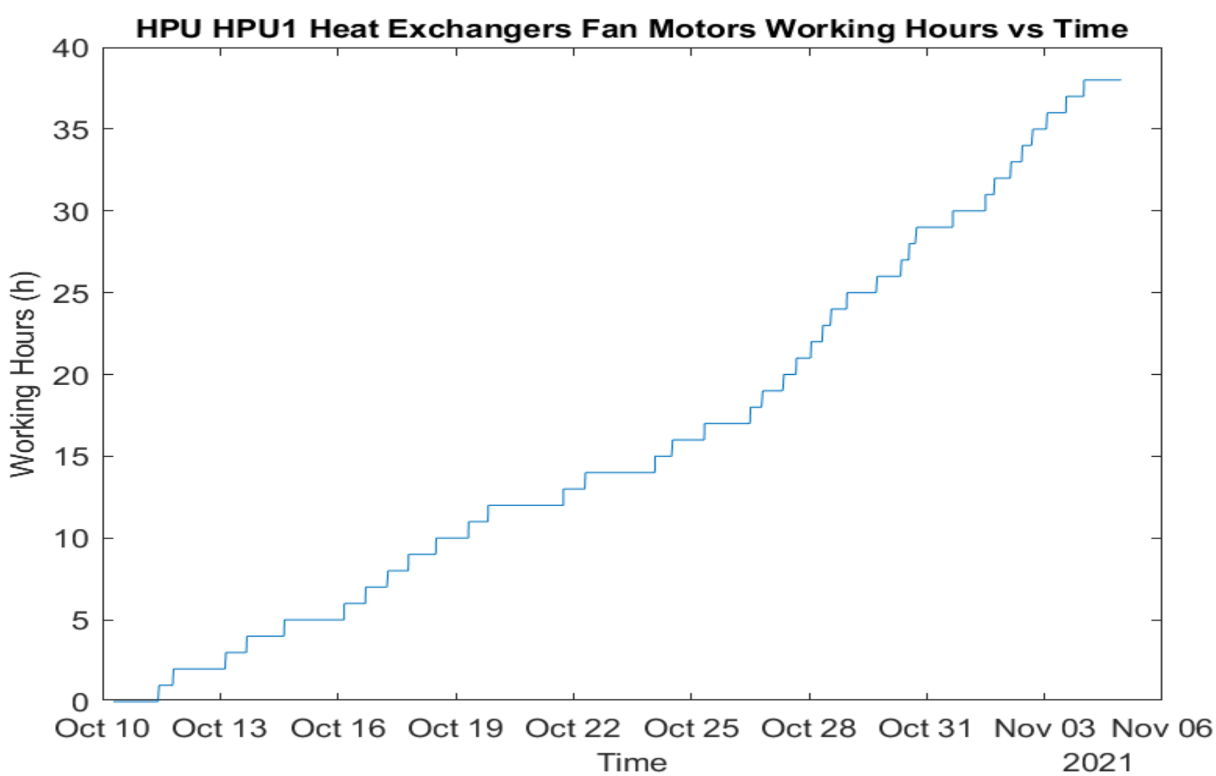

Figure 81: Heat Exchanger Working Hours vs. Time (Cleaned Data)

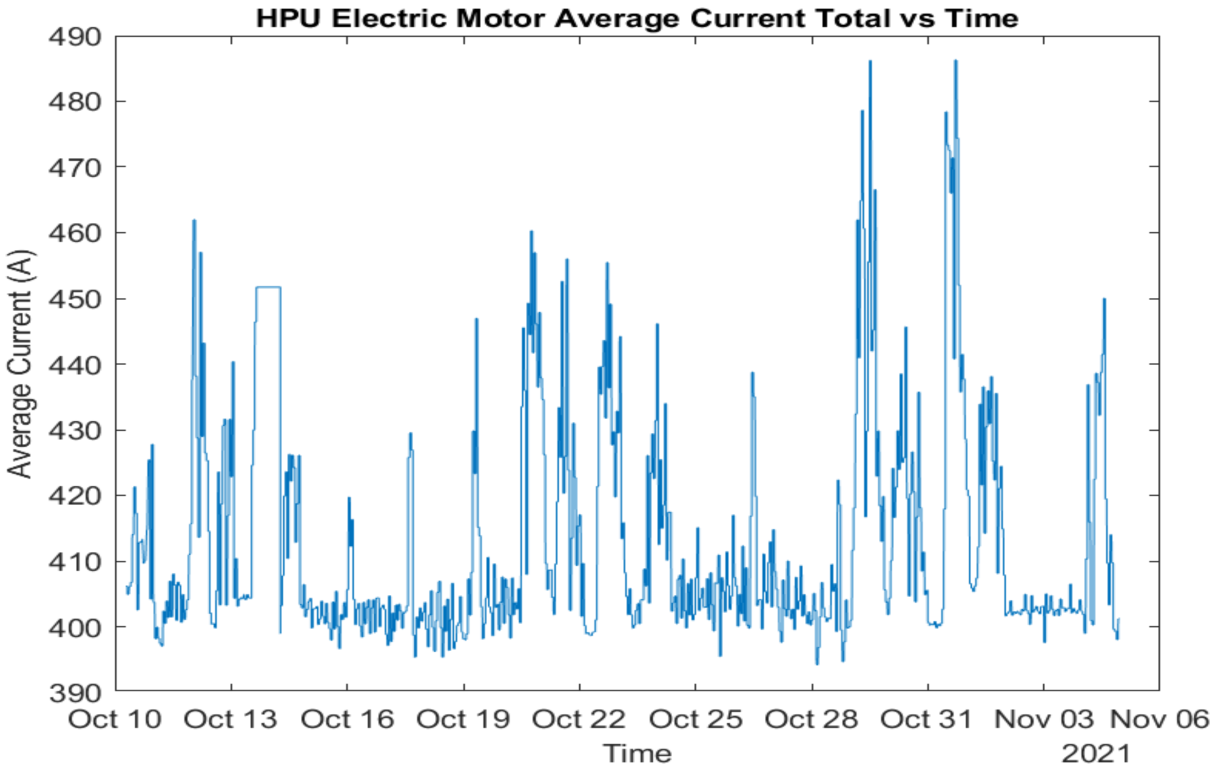

Figure 82: Total Average Current vs. Time (Cleaned Data)

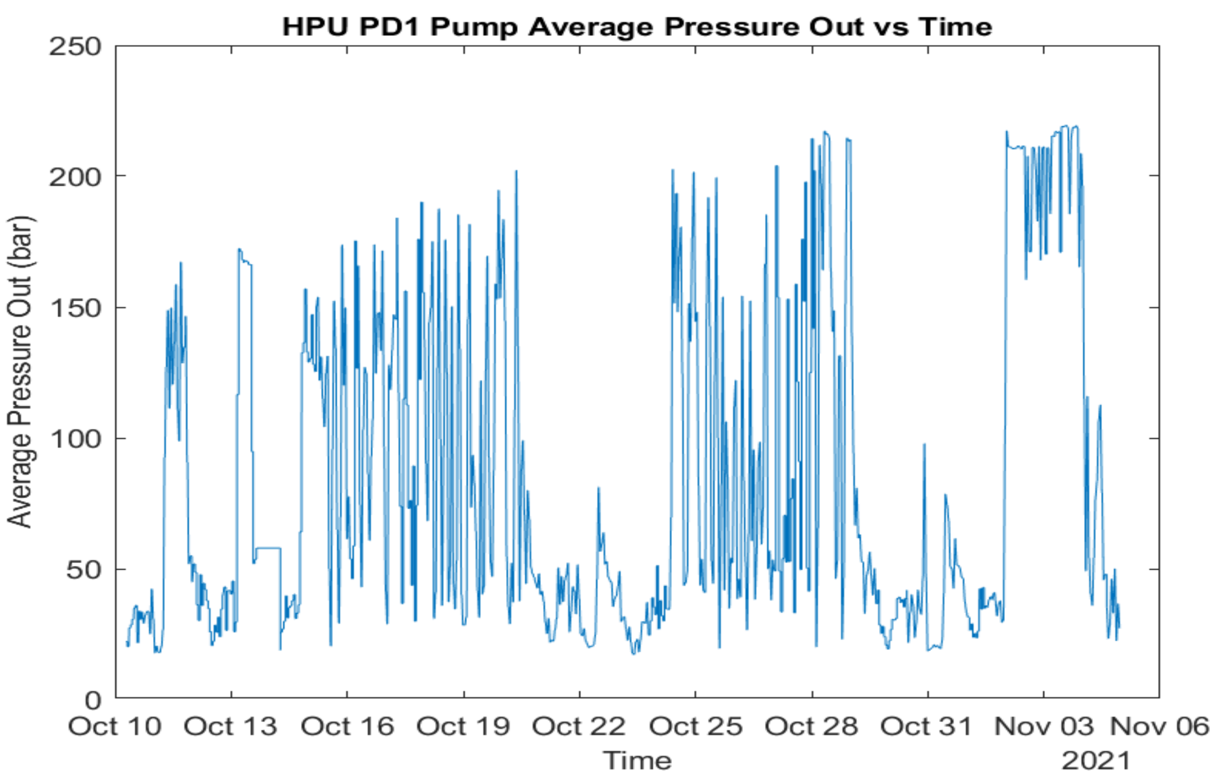

Figure 83: PD1 Average Pressure Out vs. Time (Cleaned Data)

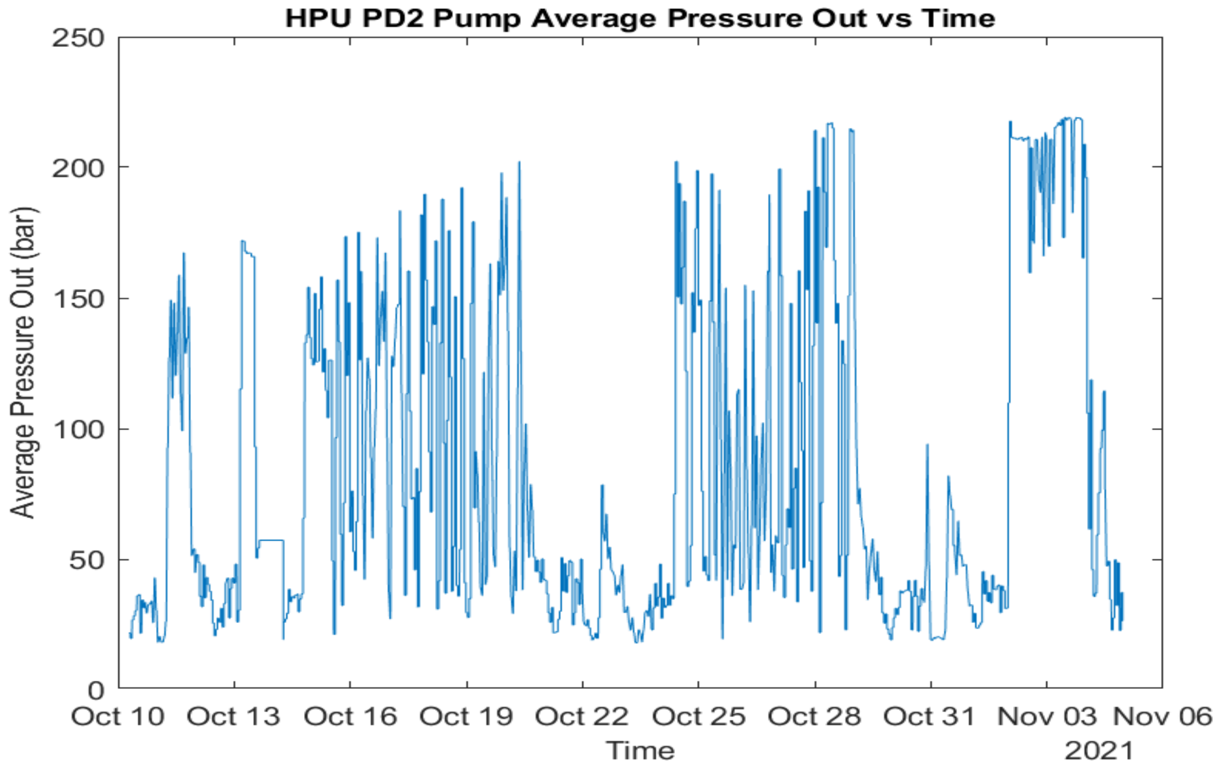

Figure 84: PD2 Average Pressure Out vs. Time (Cleaned Data)

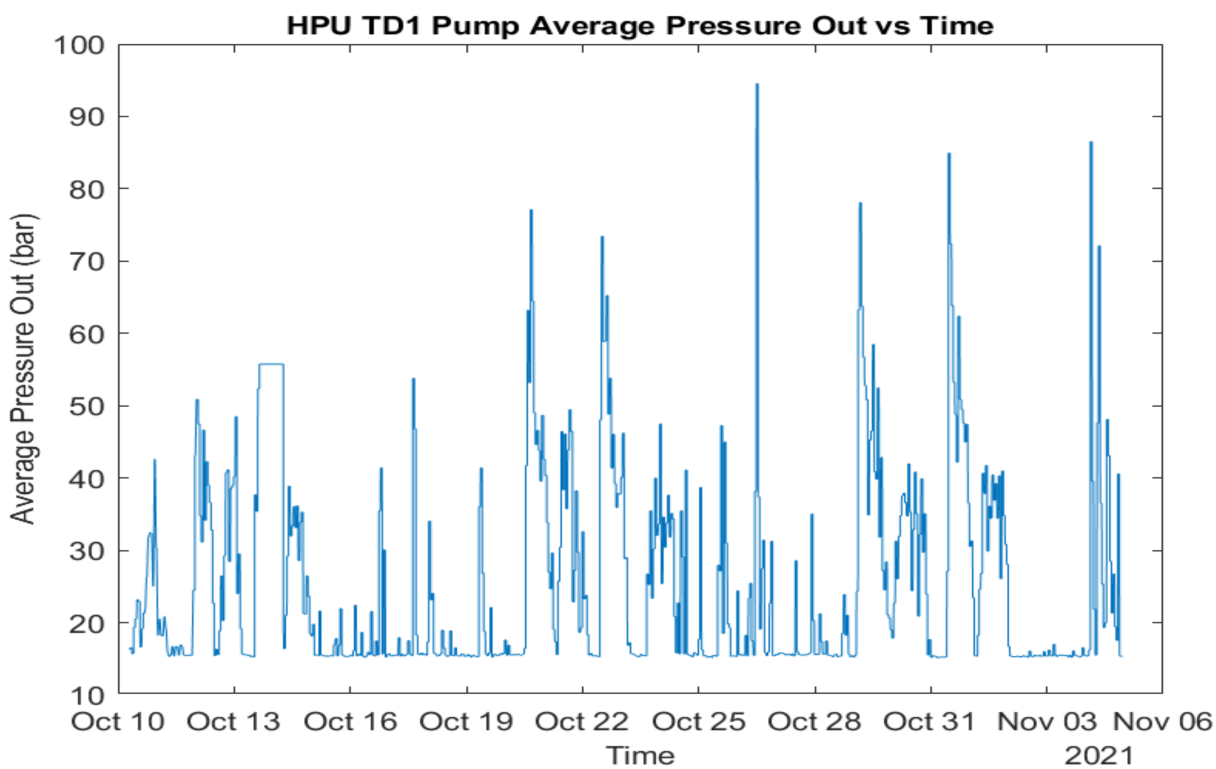

Figure 85: TD1 Average Pressure Out vs. Time (Cleaned Data)

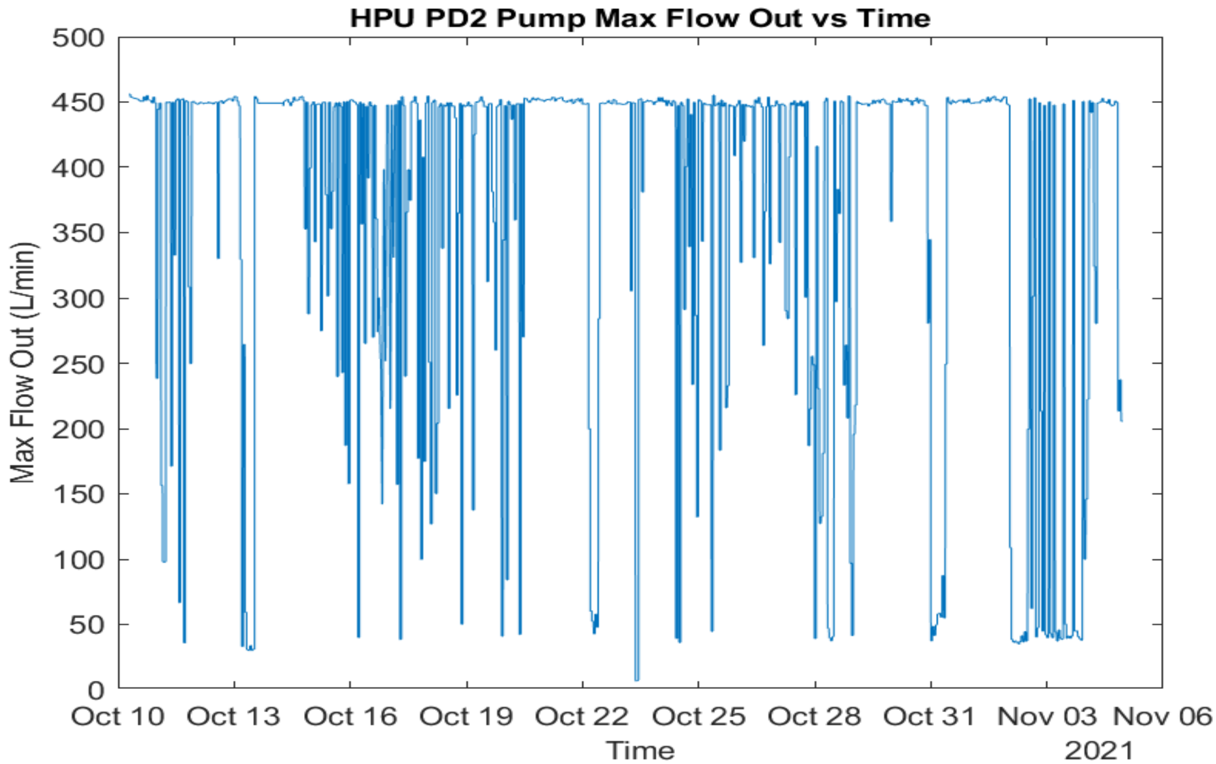

Figure 86: PD2 Max Flow Out vs. Time (Cleaned Data)

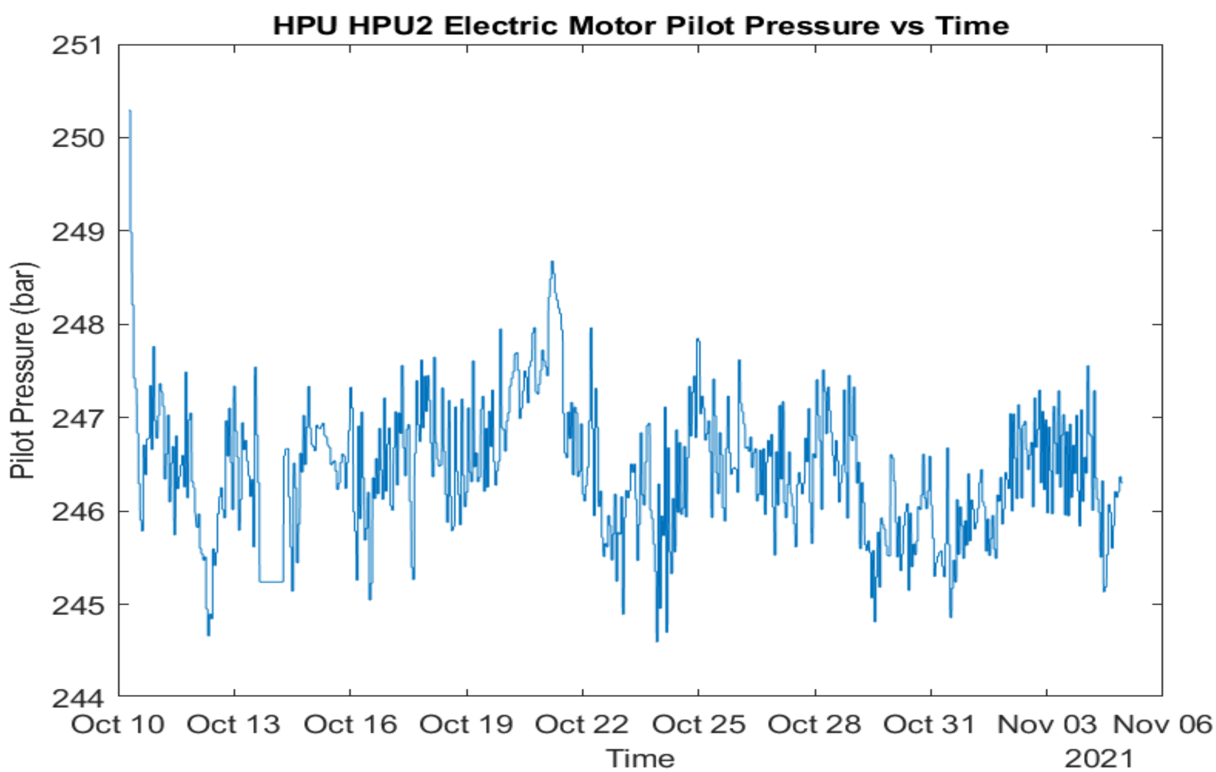

Figure 87: HPU2 Pilot Pressure vs. Time (Cleaned Data)

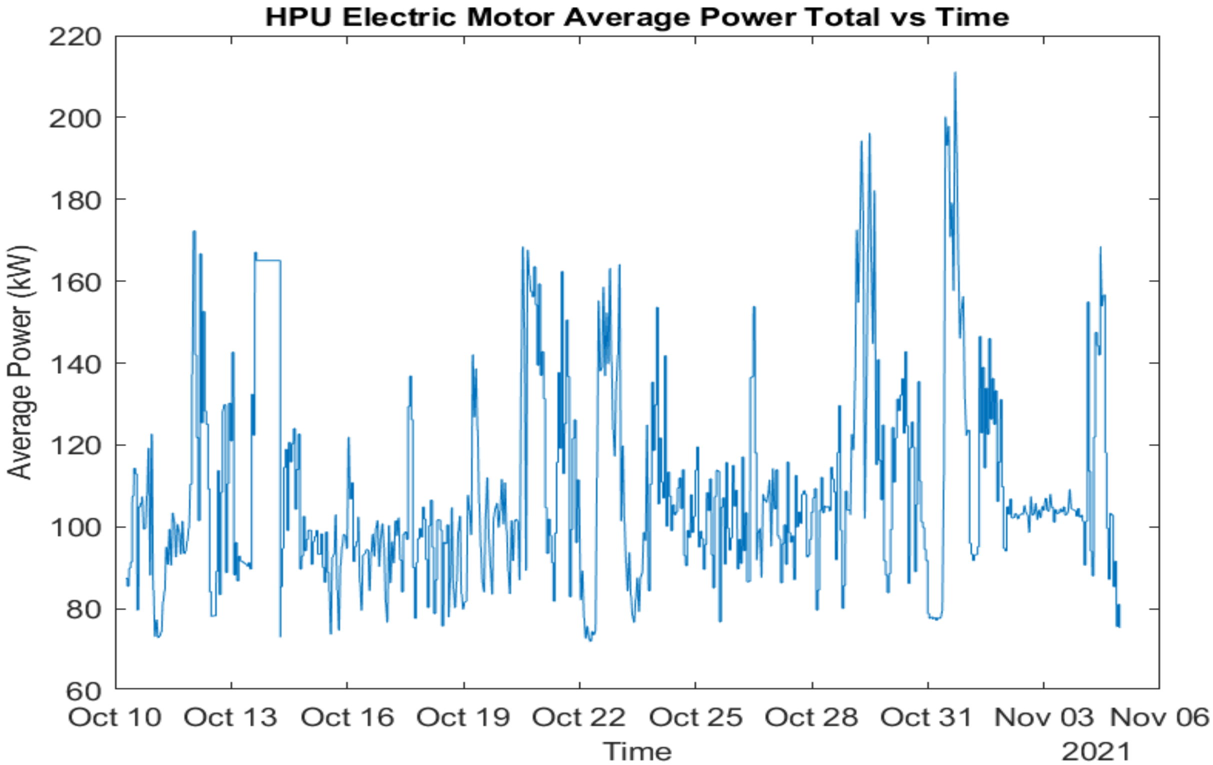

Figure 88: Total Average Power vs. Time (Cleaned Data)

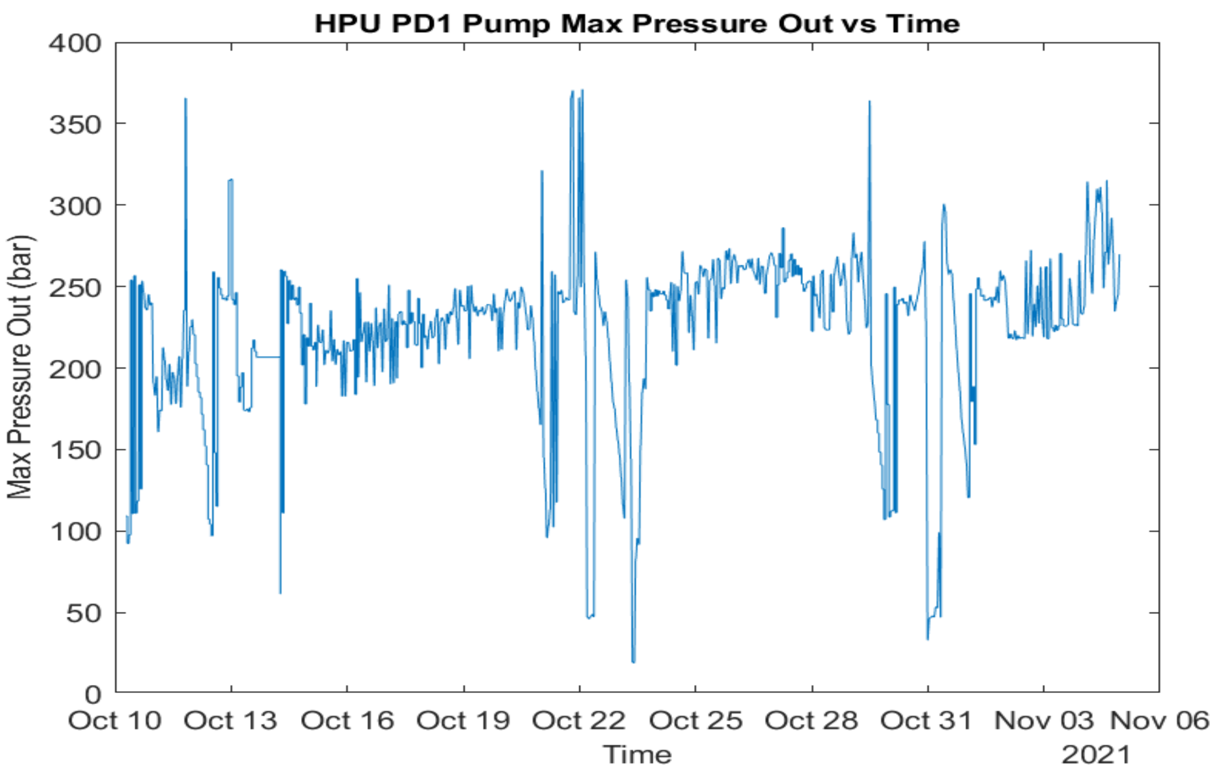

Figure 89: PD1 Max Pressure Out vs. Time (Cleaned Data)

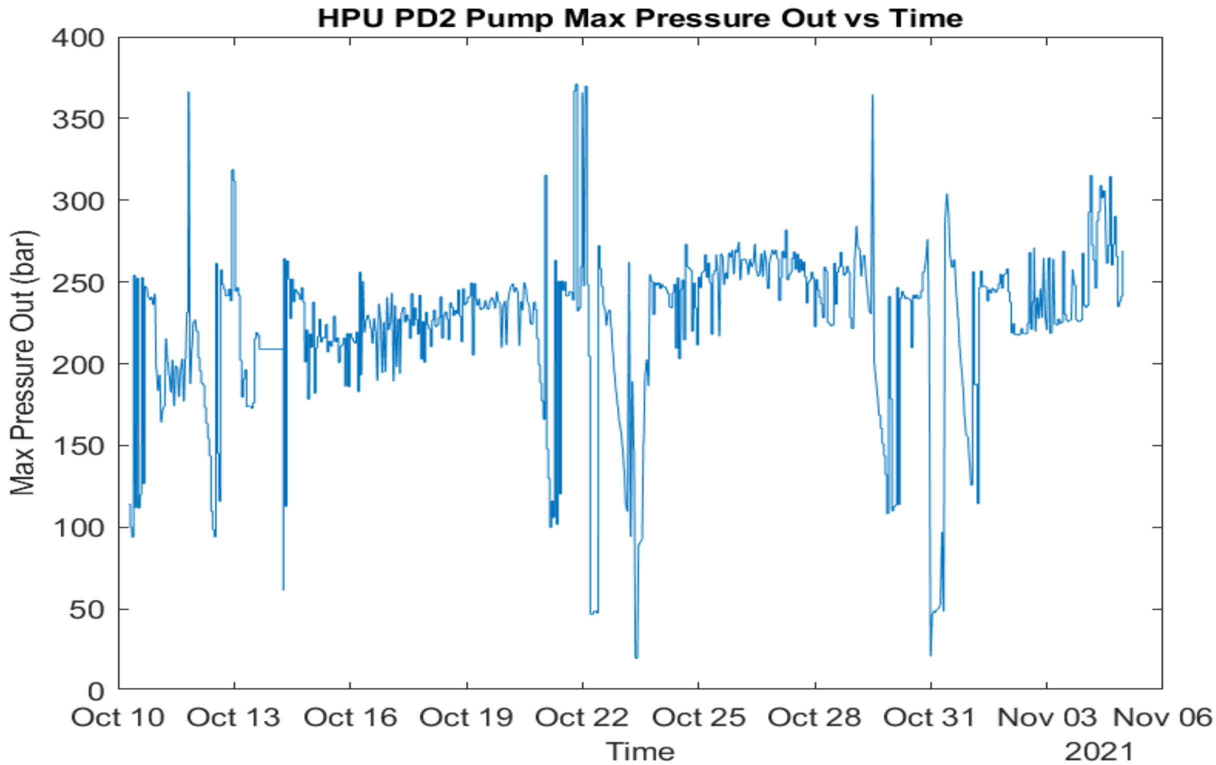

Figure 90: PD2 Max Pressure Out vs. Time (Cleaned Data)

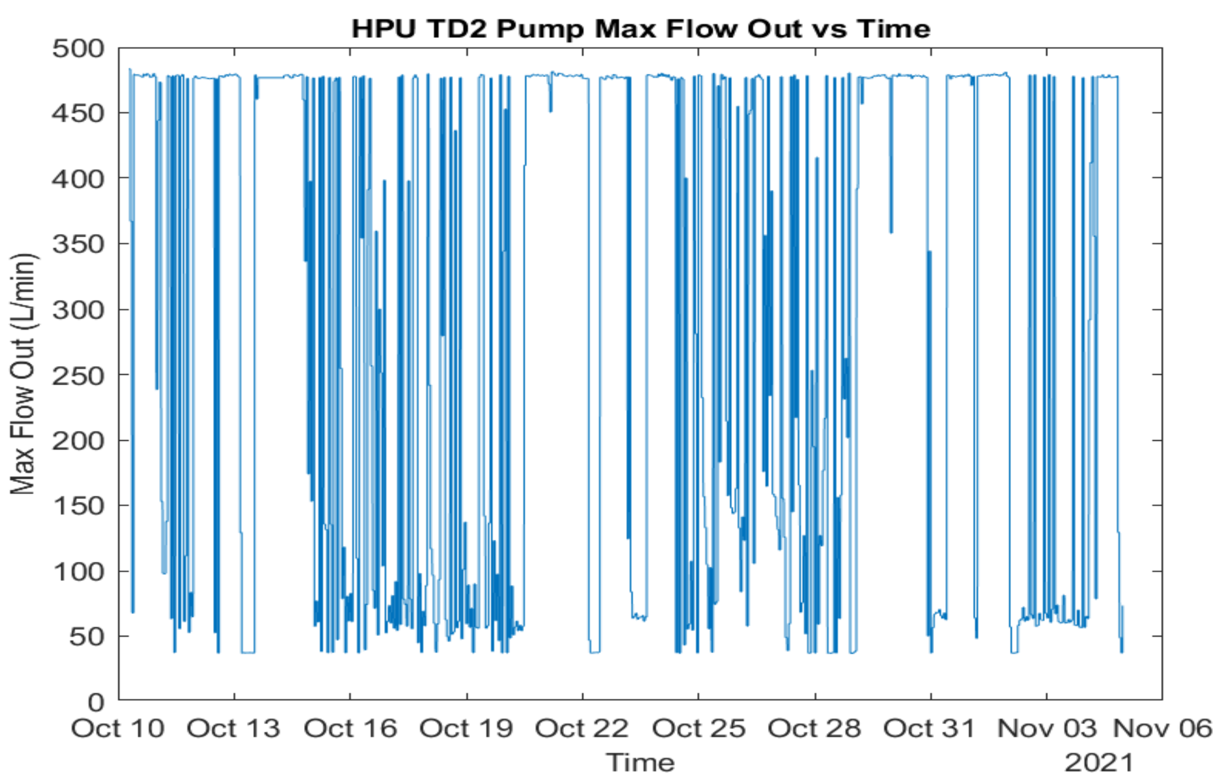

Figure 91: TD2 Max Flow Out vs. Time (Cleaned Data)

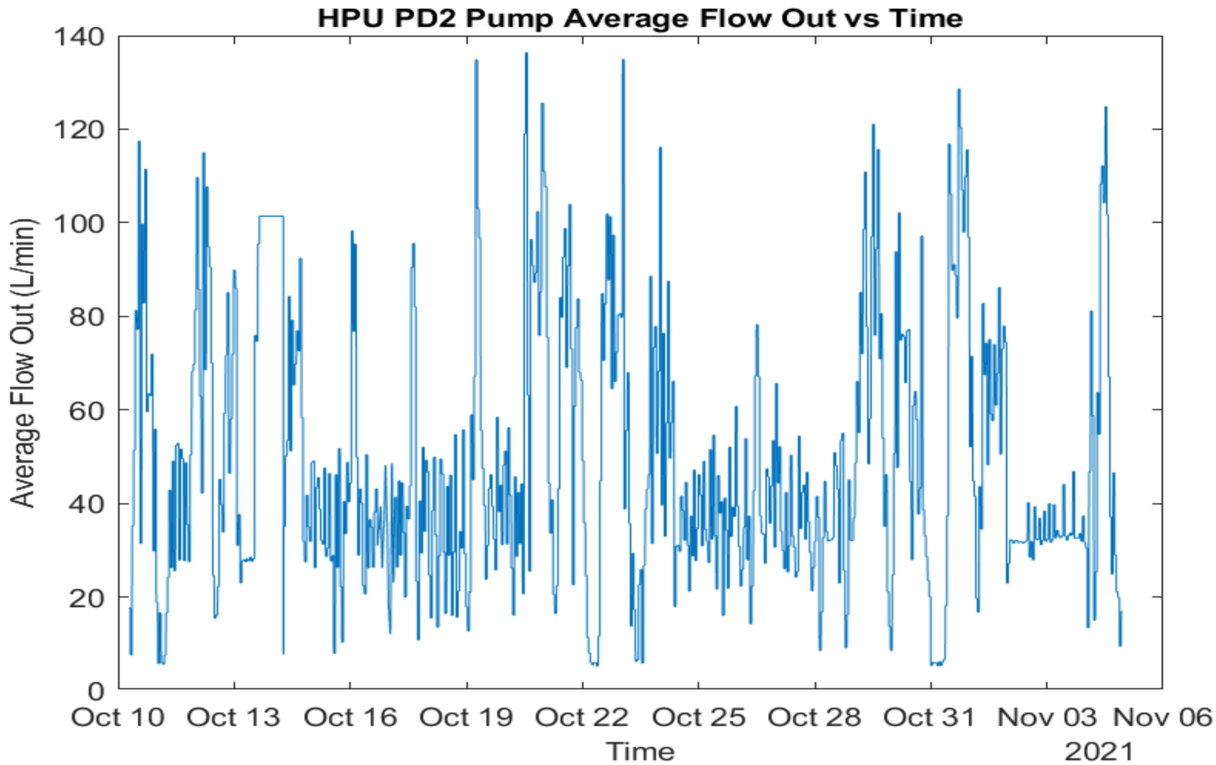

Figure 92: PD2 Average Flow Out vs. Time (Cleaned Data)

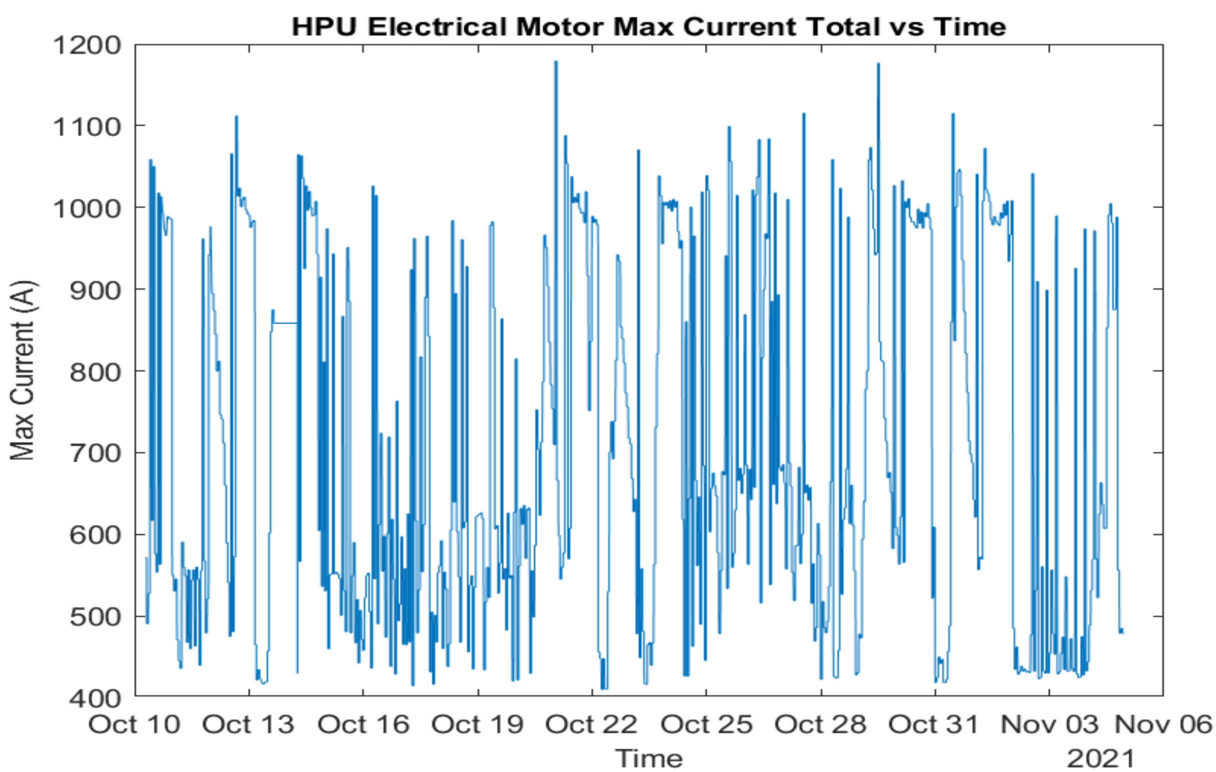

Figure 93: Total Max Current vs. Time (Cleaned Data)

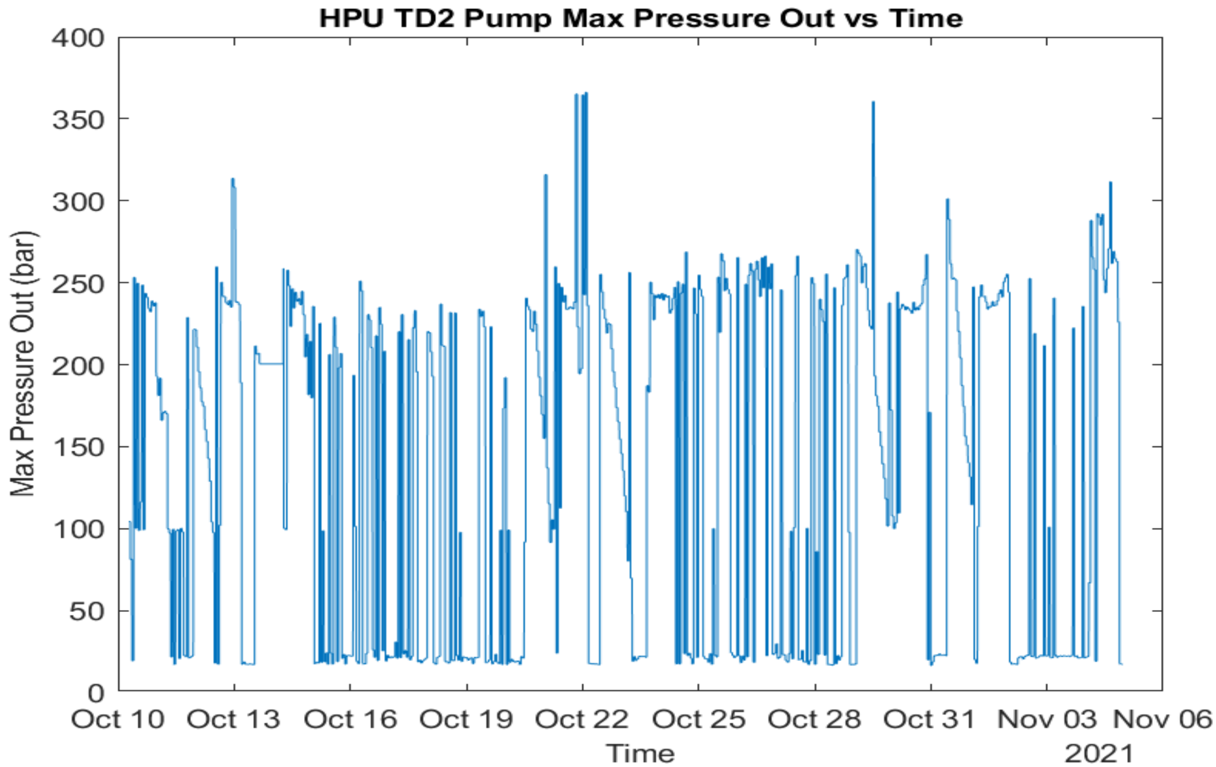

Figure 94: TD2 Max Pressure Out vs. Time (Cleaned Data)

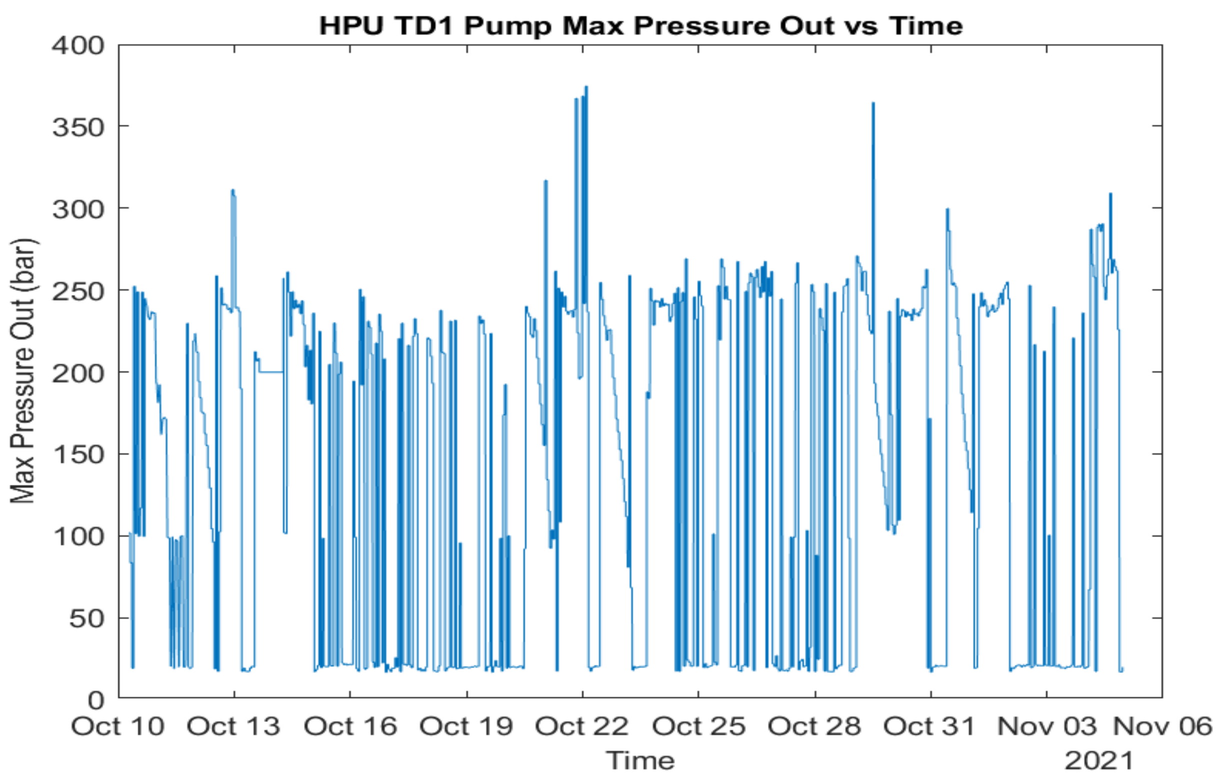

Figure 95: TD1 Max Pressure Out vs. Time (Cleaned Data)

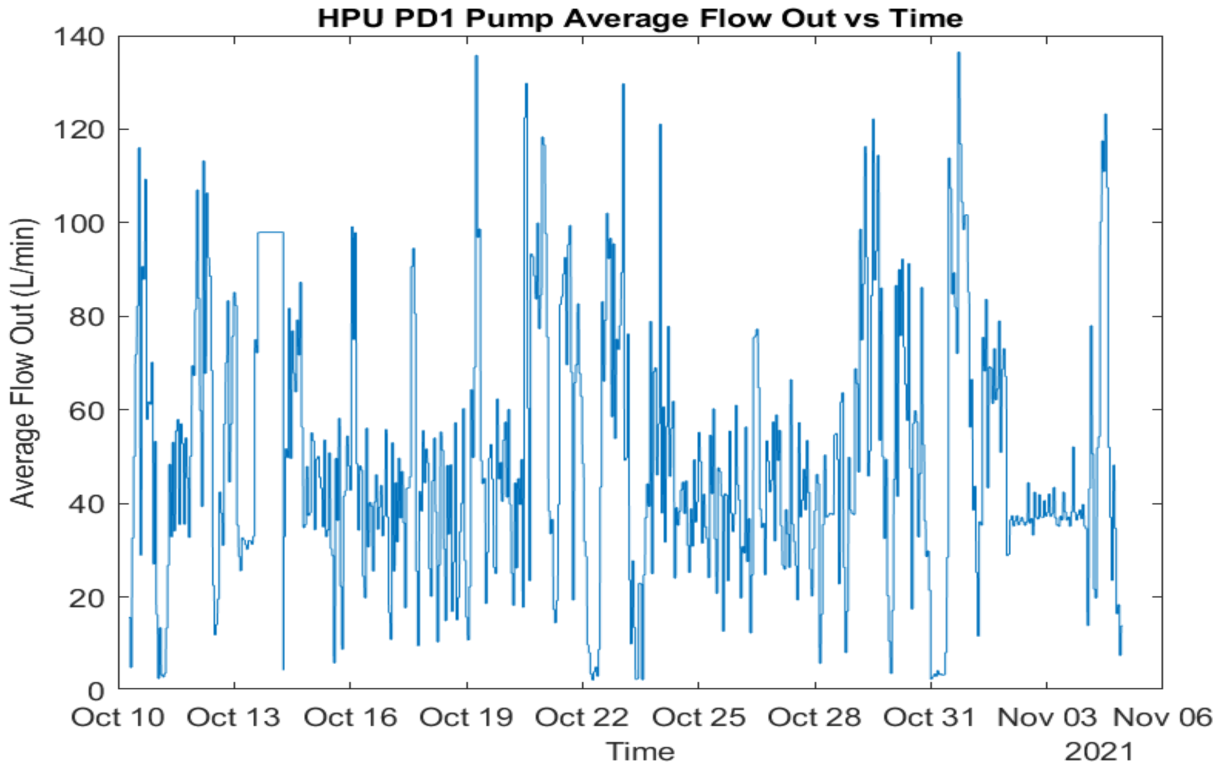

Figure 96: PD1 Average Flow Out vs. Time (Cleaned Data)

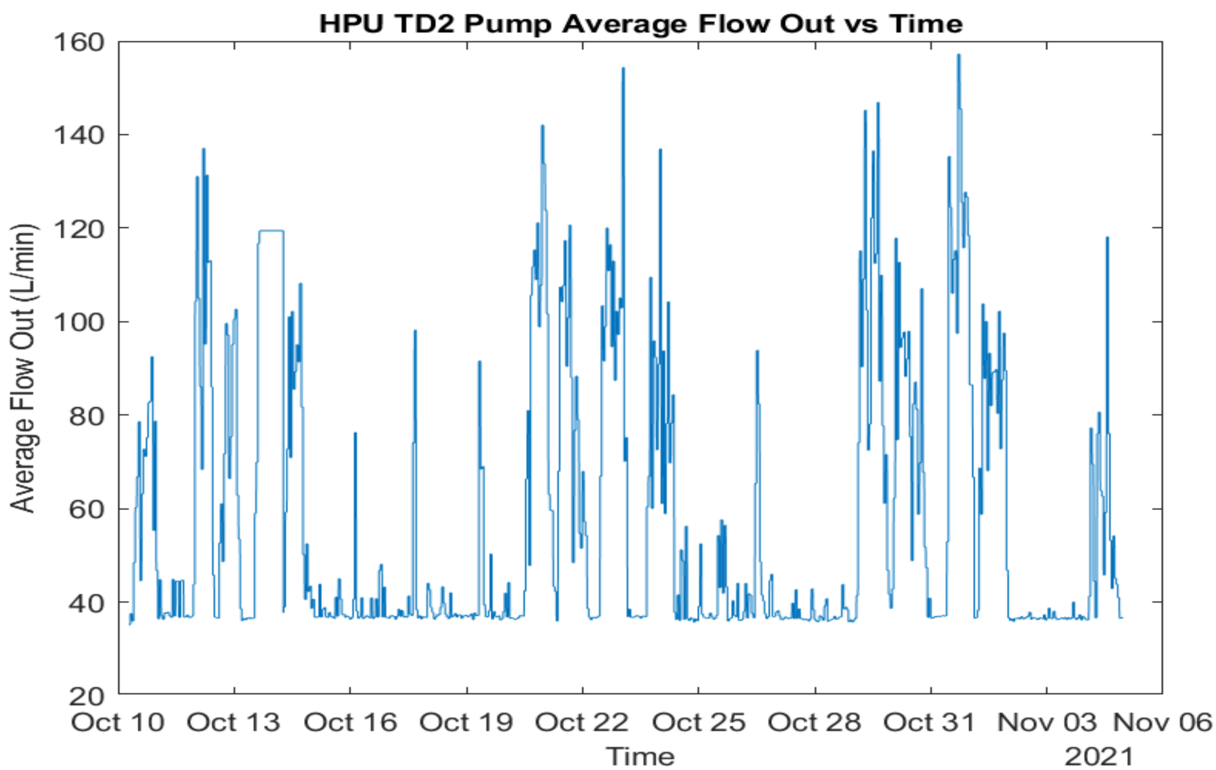

Figure 97: TD2 Average Flow Out vs. Time (Cleaned Data)

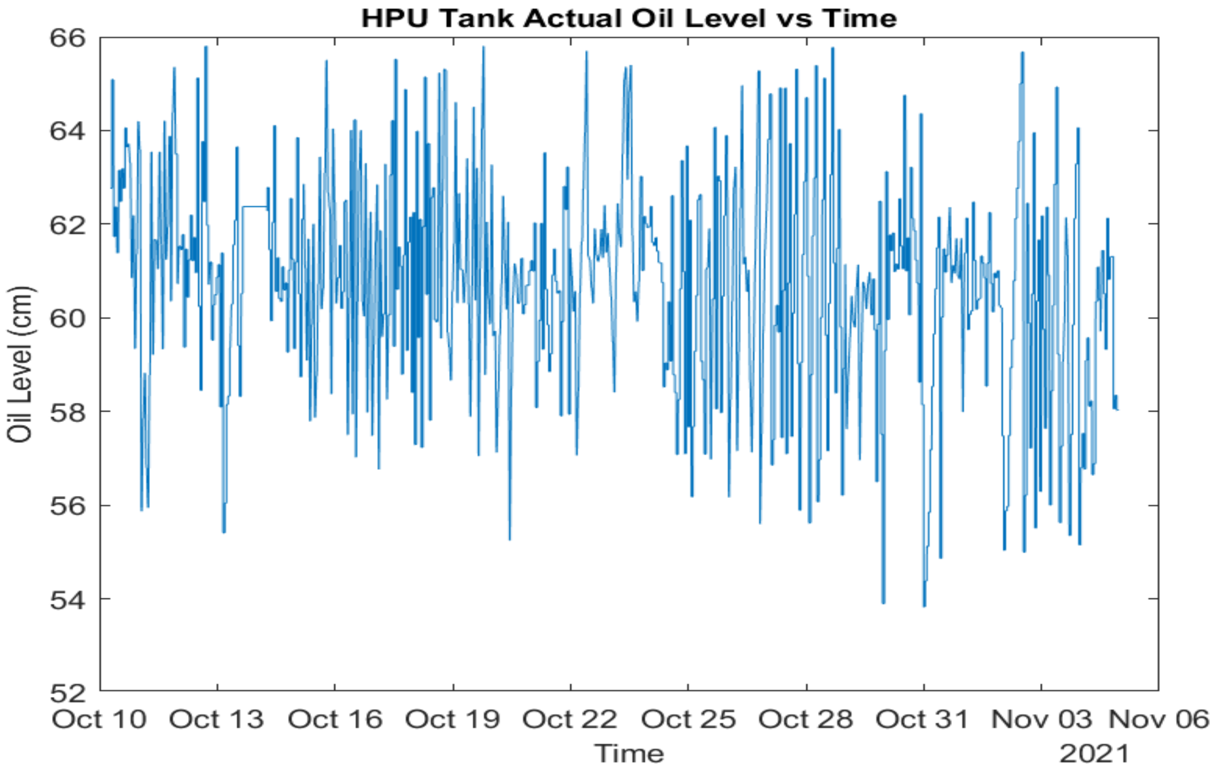

Figure 98: Actual Oil Level vs. Time (Cleaned Data)

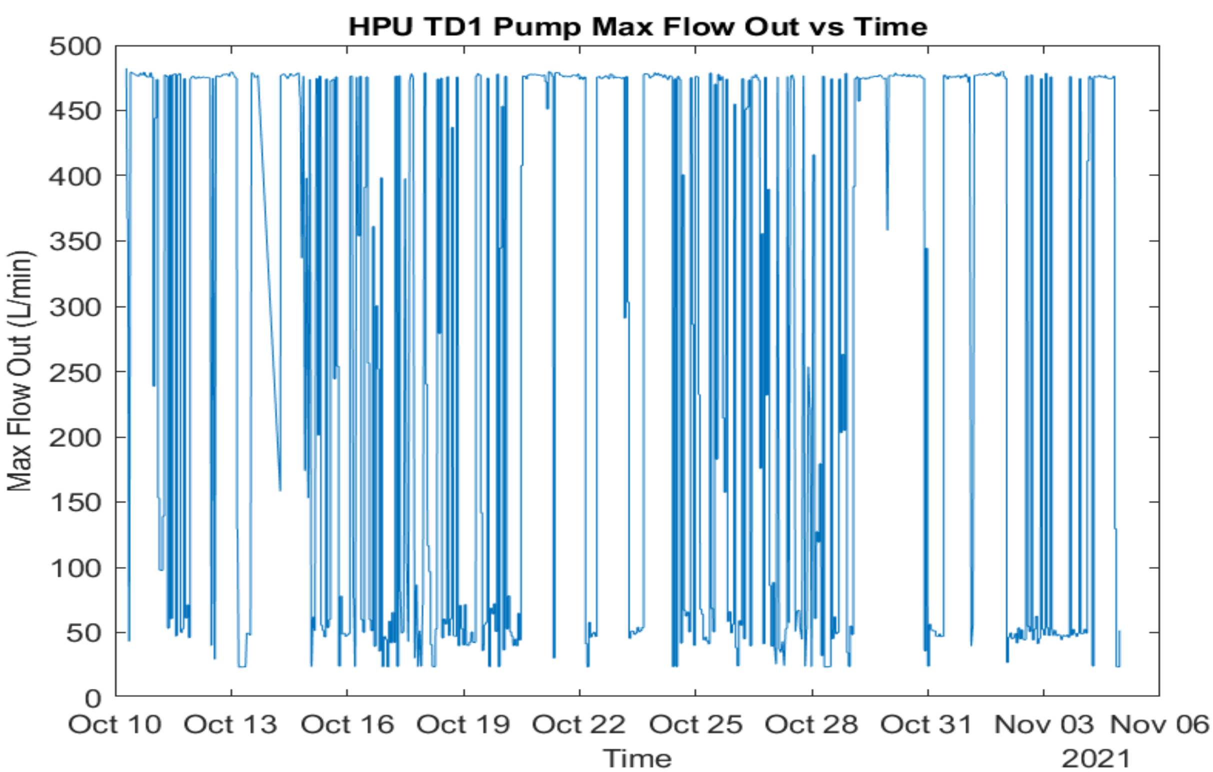

Figure 99: TD1 Max Flow Out vs. Time (Cleaned Data)

8.3.3 TDS Data Cleaning

Similarly to the two previous datasets, the first seven days of TDS data were discarded for all the same reasons, and plus here after this first week the number of low-pressure alarms of the blower, its working hours as well as the working hours of the electric motor and the hydraulic brake were reset. Hence it was preferred to use the values after the reset. Two other common steps were performed on all the parameters which are the removal of redundant rows and the filling of missing single rows by using average values or linear interpolation.
In the coming bullet points, some further data cleaning that was recommended for some parameters will be discussed:

- For the number of low-pressure alarms where it was clear in the plot of Figure 55 that there are some missing records, these empty rows were filled with the constant value of 8. This value corresponds to the last value of the alarms before the discontinuity appeared. That is because after it the first value that was recorded was 9 which means that there were no additional alarms in that time interval (Figure 105).

- For the average torque, the two high peaks very different from all other measurements, as seen in Figure 56, were removed and replaced by average values. This was done after making sure they were not connected to a stuck pipe by looking at the plots of the SPMs mud pumps (Figures 66, 67, and 68), the RPM (Figure 65), and the hook load (Figure 71) where three of them had no behavior linked whatsoever to a stuck pipe and that meant that they were only related to sensor's misreading (Figure 106).

- For the max torque, a similar procedure to the one performed for the average torque was followed, to remove the peaks (Figure 102)

All other cleaned data plots relating to the remaining parameters are shown in Figures 101, 103, 104, 107, 108, 109, 110, 111, and 112.

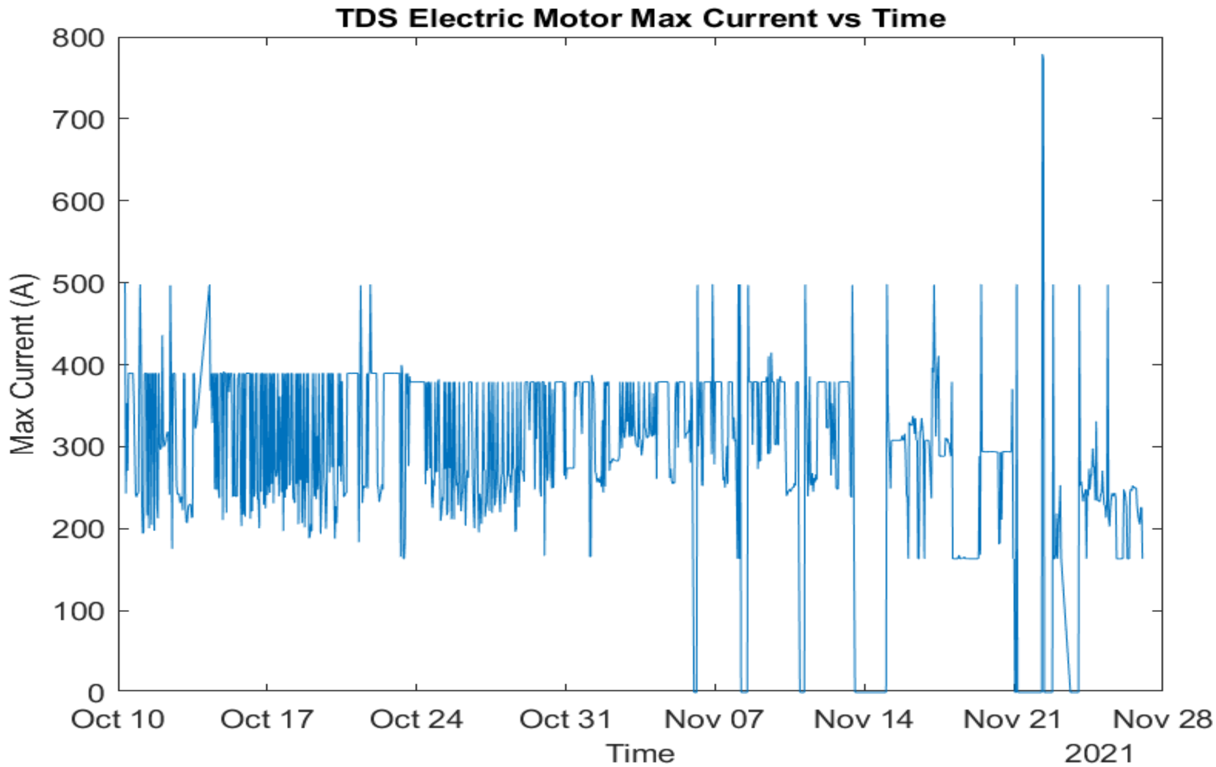

Figure 100: Max Current vs. Time (Cleaned Data)

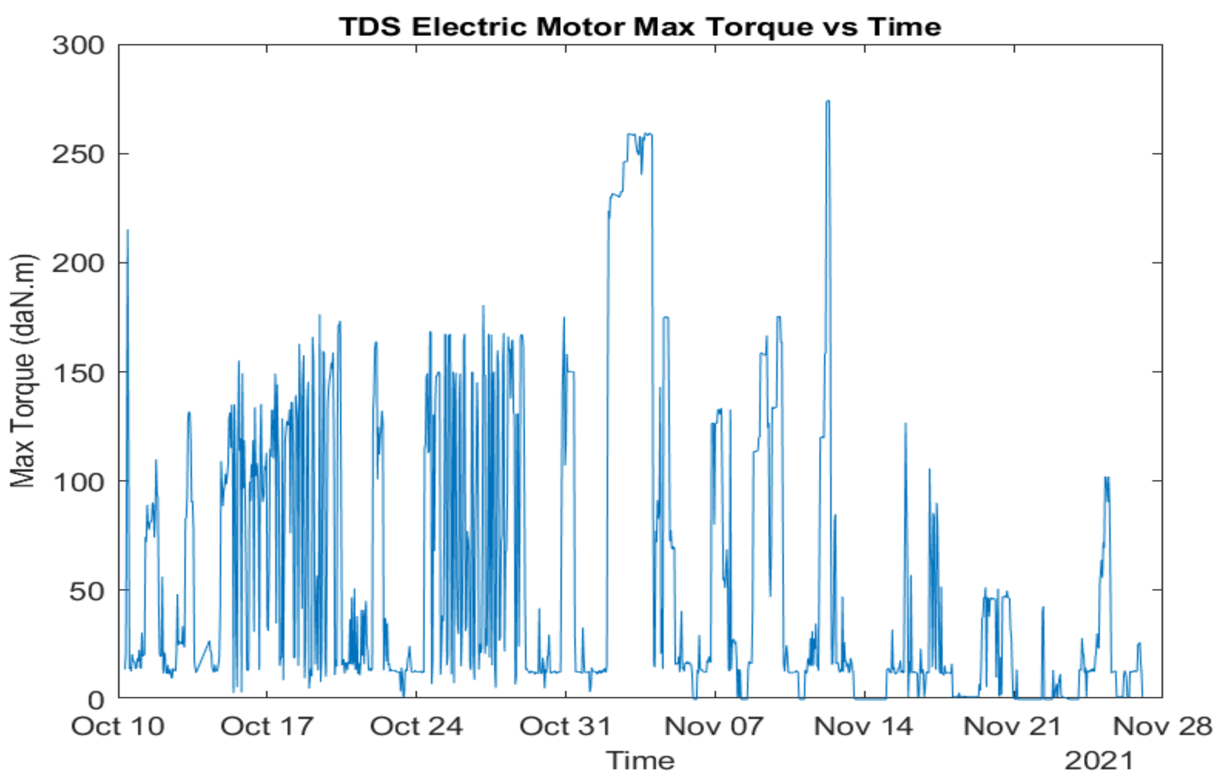

Figure 101: Max Torque vs. Time (Cleaned Data)

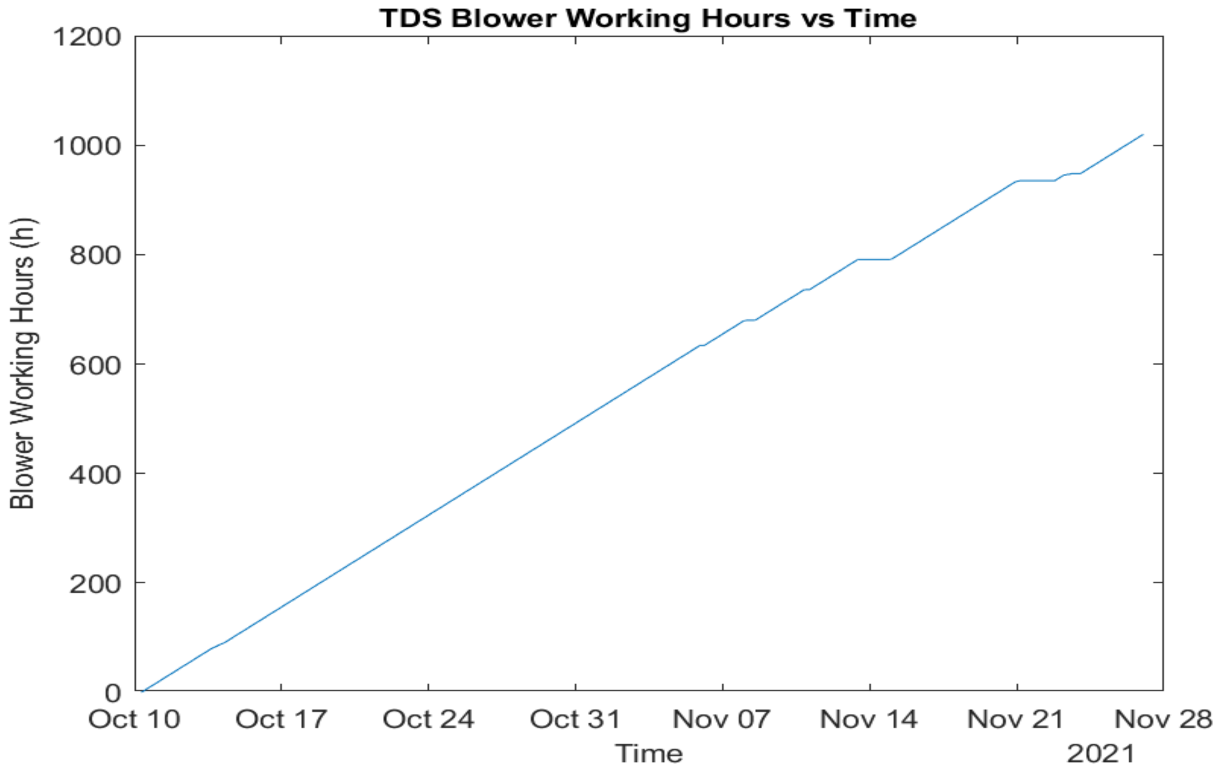

Figure 102: Blower Working Hours vs. Time (Cleaned Data)

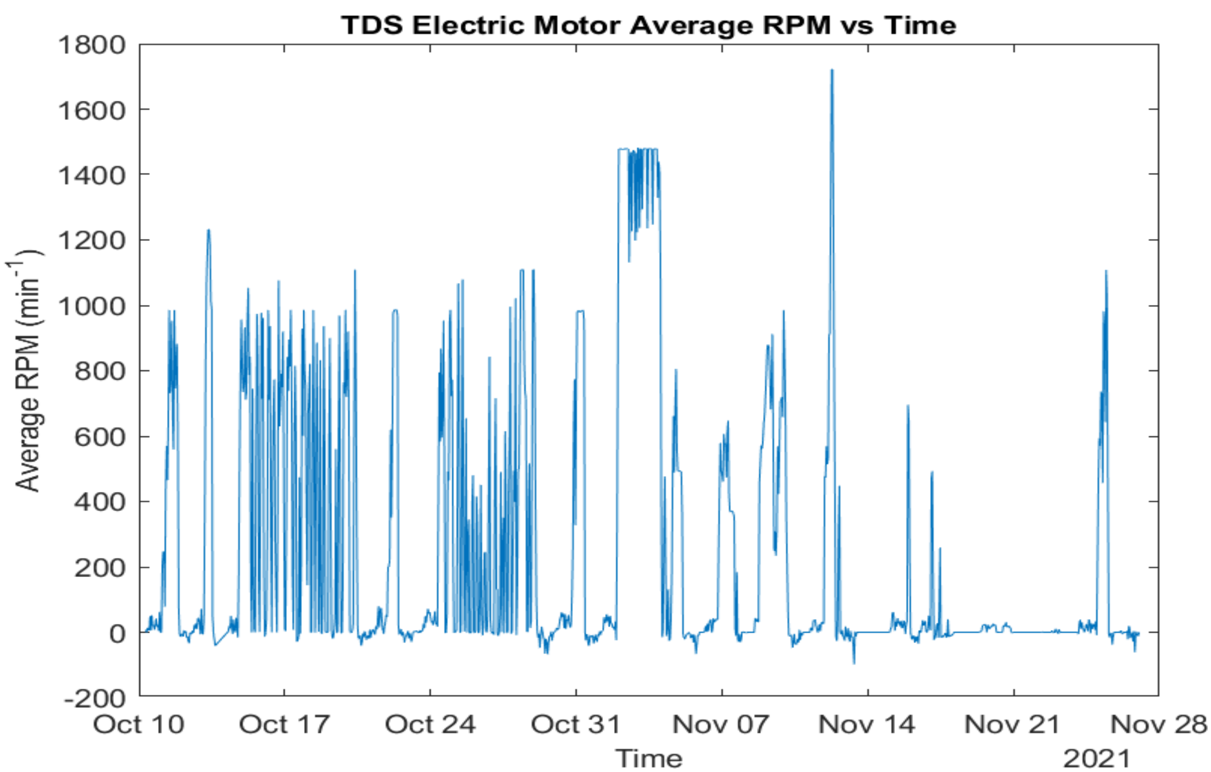

Figure 103: Average RPM vs. Time (Cleaned Data)

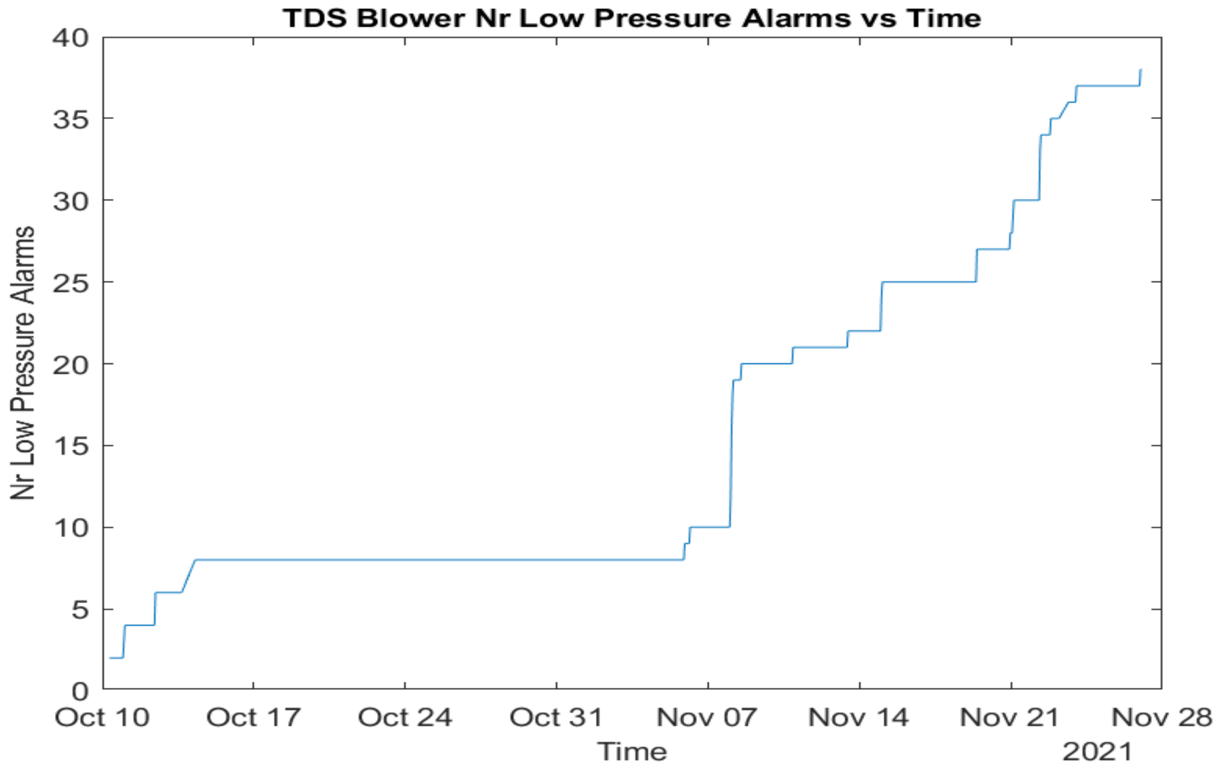

Figure 104: Blower Number of Low-Pressure Alarms vs. Time (Cleaned Data)

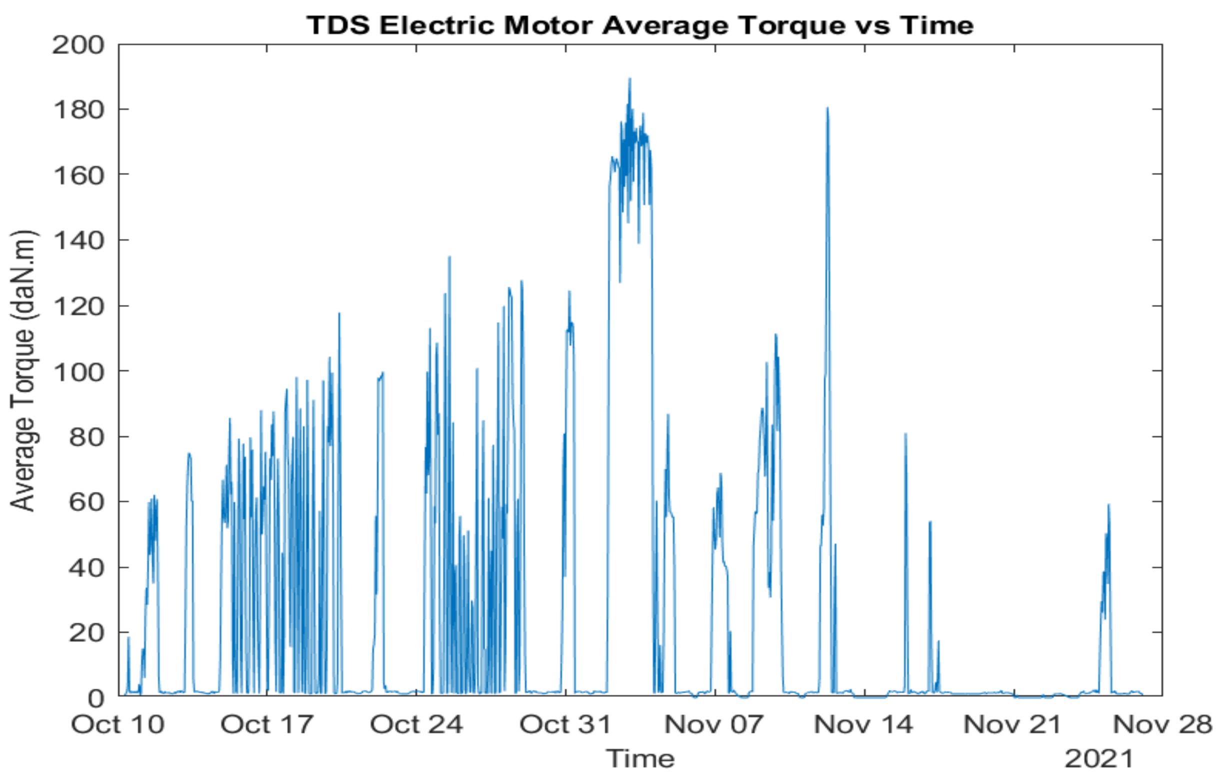

Figure 105: Average Torque vs. Time (Cleaned Data)

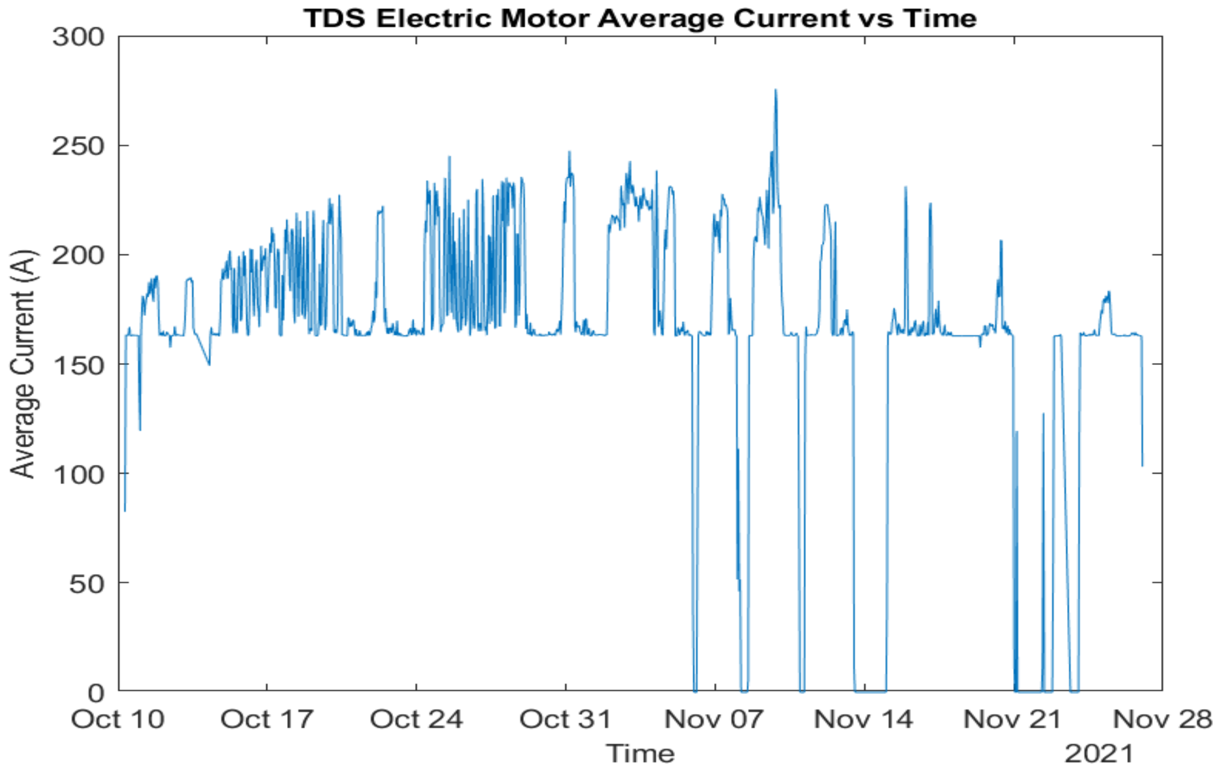

Figure 106: Average Current vs. Time (Cleaned Data)

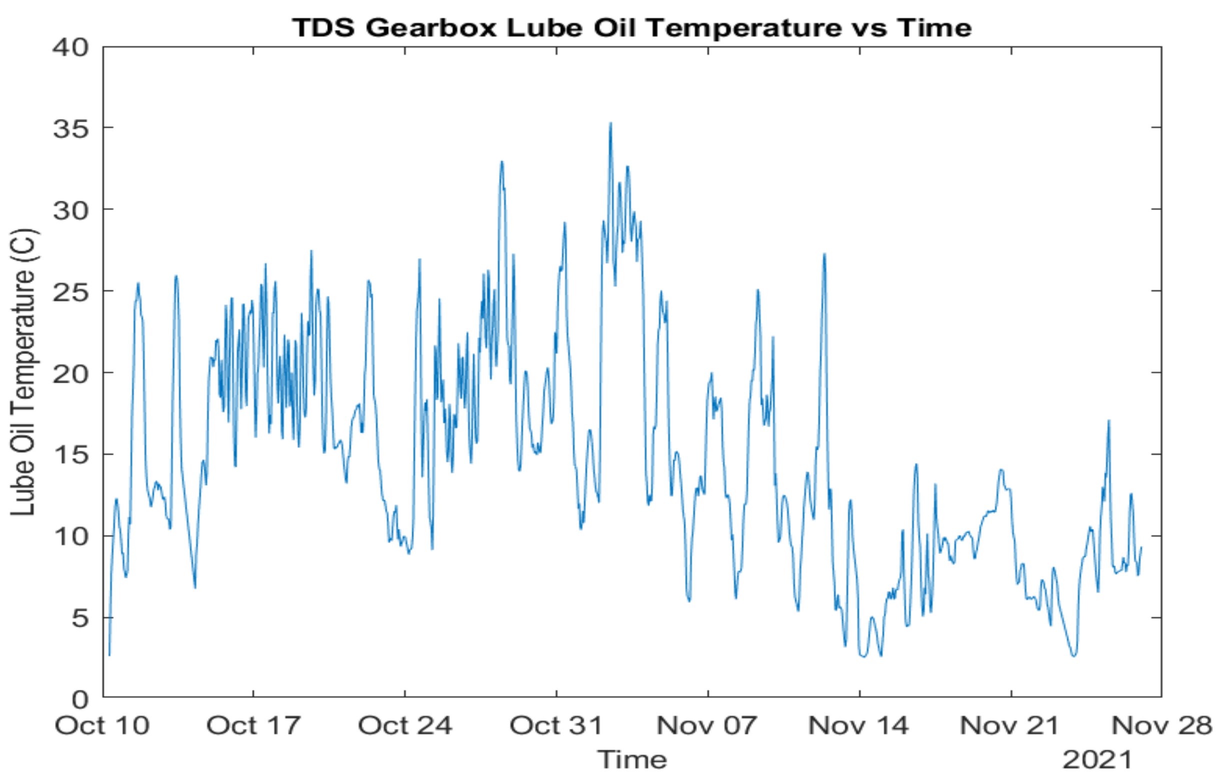

Figure 107: Lube Oil Temperature vs. Time (Cleaned Data)

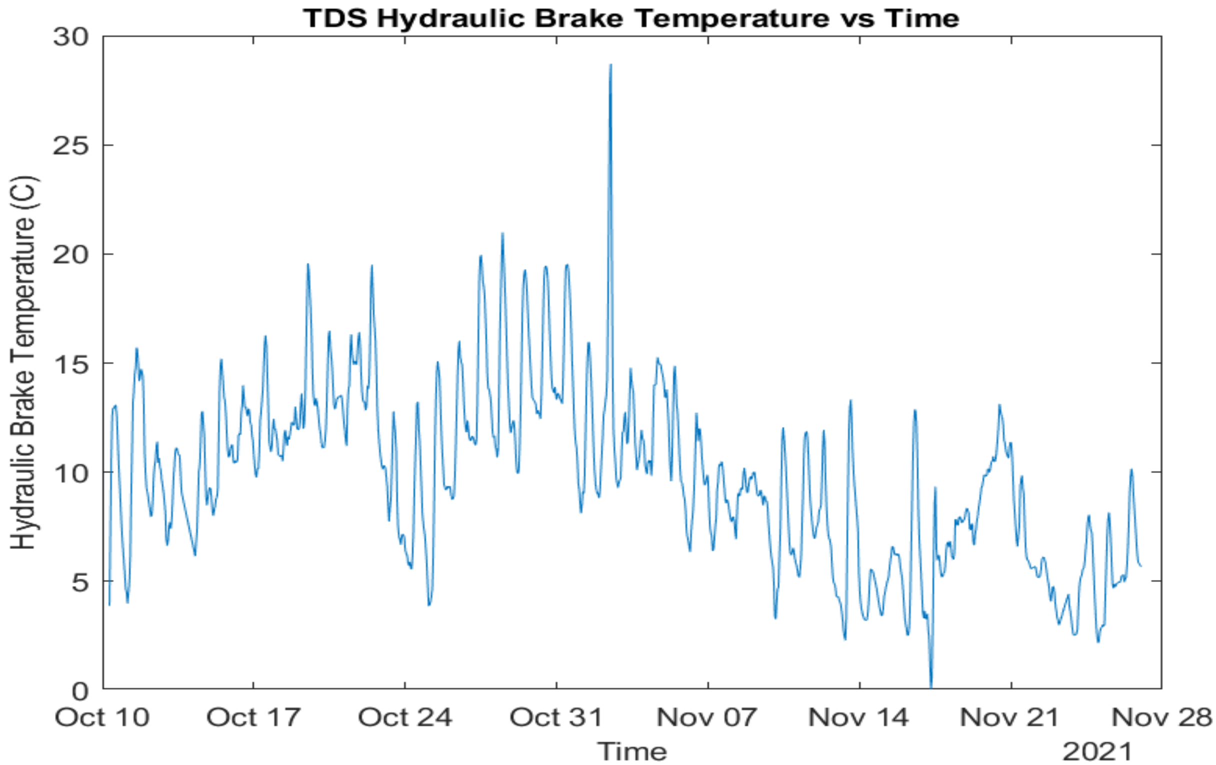

Figure 108: Hydraulic Brake Temperature vs. Time (Cleaned Data)

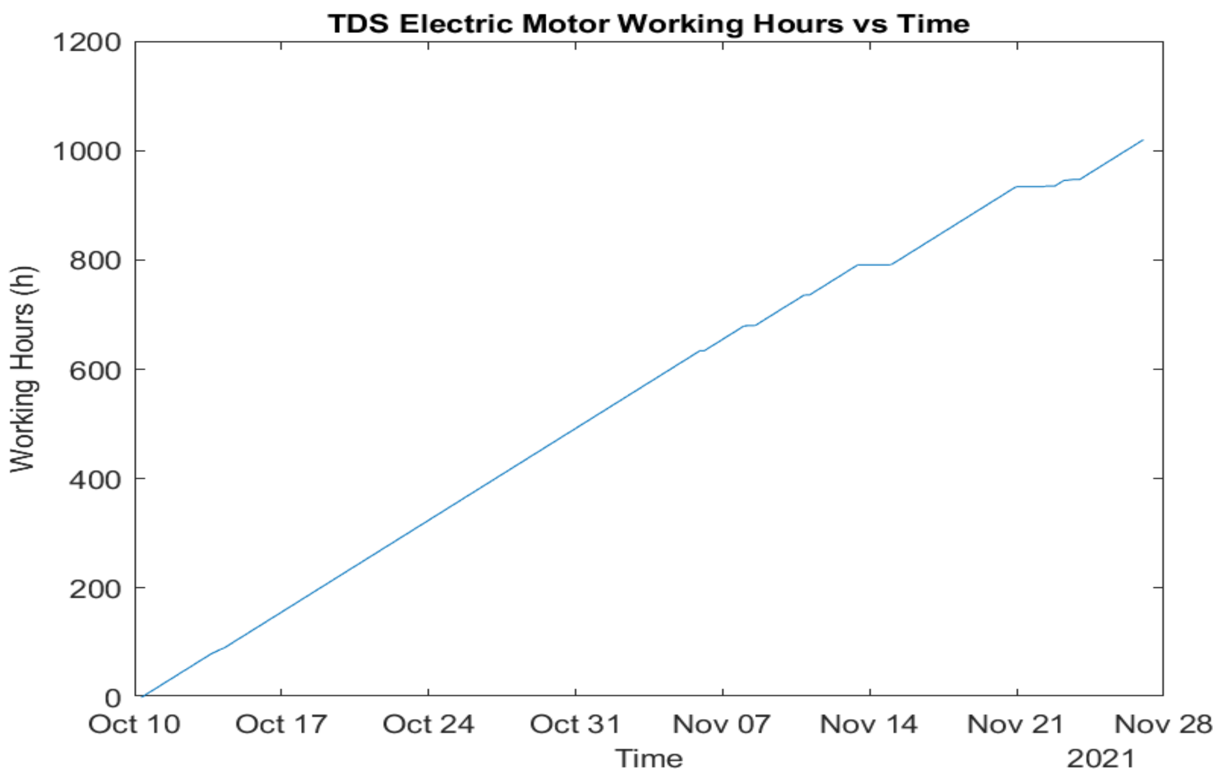

Figure 109: Electric Motor Working Hours vs. Time (Cleaned Data)

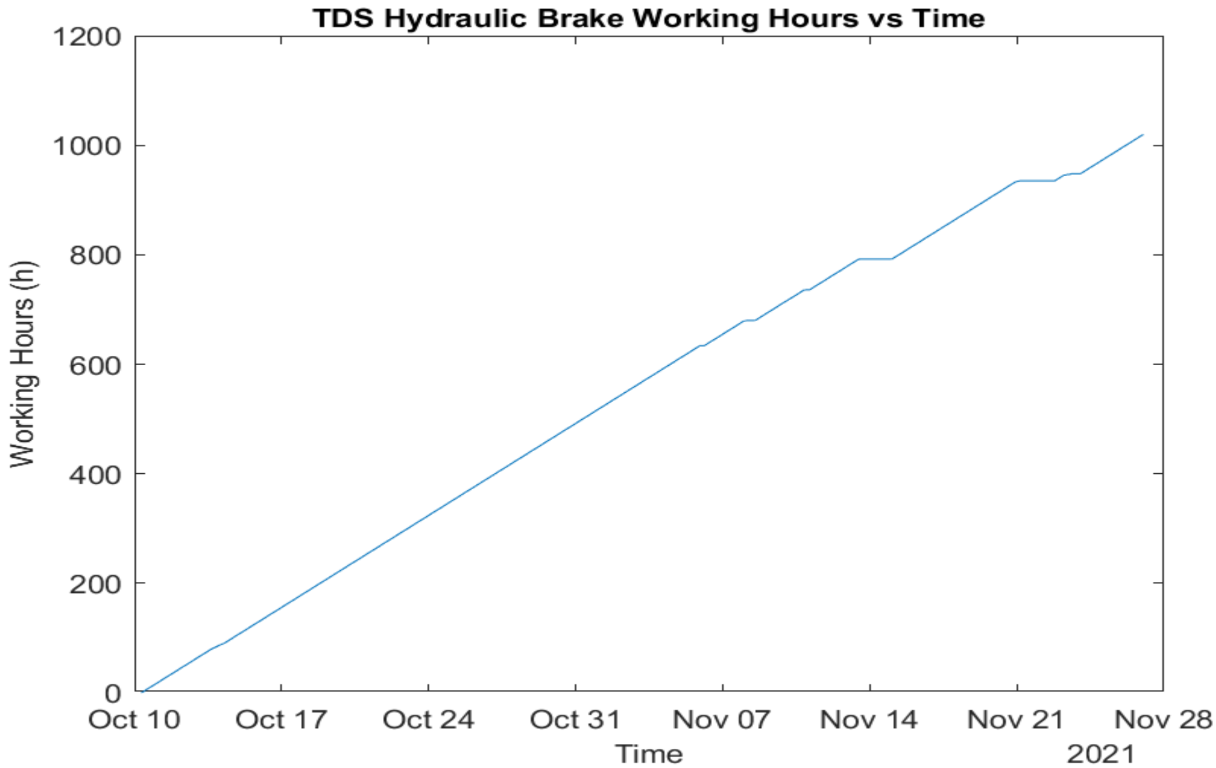

Figure 110: Hydraulic Brake Working Hours vs. Time (Cleaned Data)

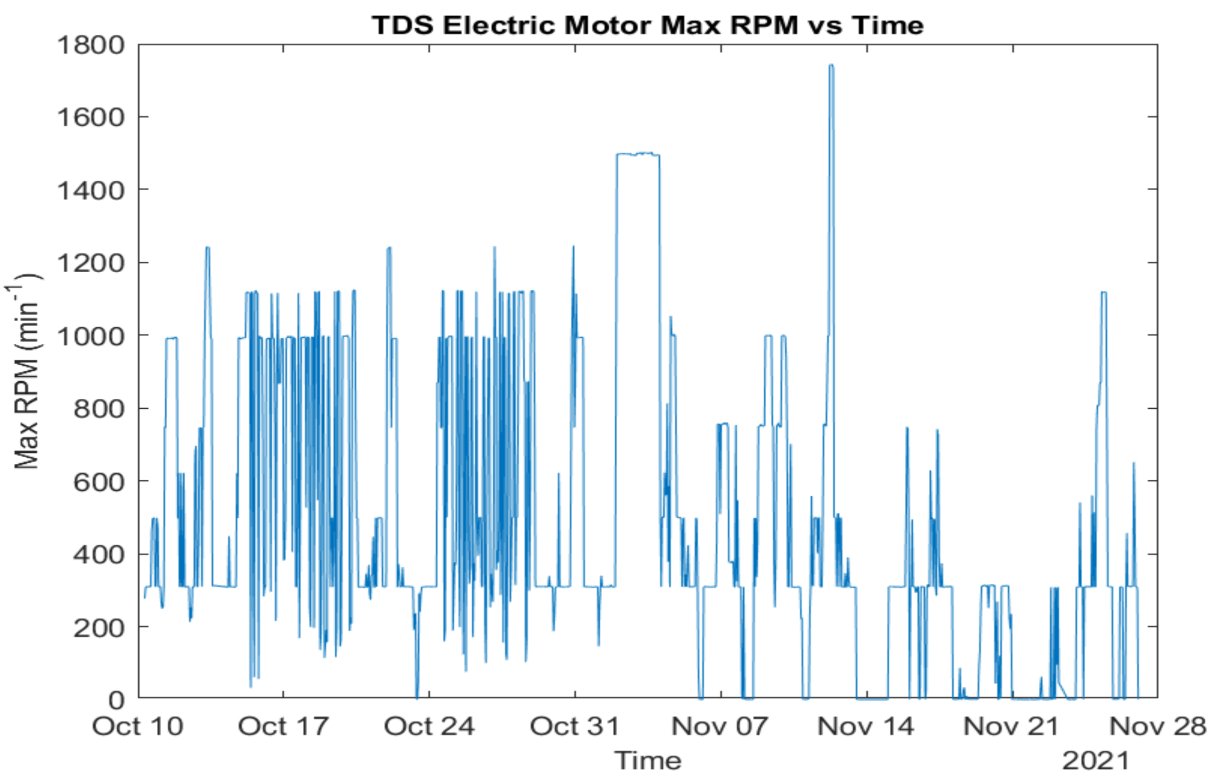

Figure 111: Max RPM vs. Time (Cleaned Data)

9 Results and Discussions

In the following subparagraphs, the model implementation and the results obtained for the prediction of the number of low-pressure alarms of the blower are showcased and discussed. In addition to the KPIs, the following equipment is described: the hydraulic pumps of the HPU, the filter clogging of the HPU, the heat exchanger of the HPU, and the lube oil quality of the TDS.

9.1 Blower Failure Prediction

After the collection and cleaning of the top drive system's data, it was used as the input for the time series predictive model in ThingWorx analytics. This model used the artificial neural network, decision tree, gradient boost, linear regression, and ensemble machine learning techniques to give the output. The goal of this model or the output is to predict the number of low-pressure alarms of the blower in the top drive system. Knowing the probable failure frequencies ahead of time allows the proper maintenance of the components before arriving at a failure. The trial was set to predict the upcoming 64 values equivalent to 64 hours, which is also the maximum look ahead that can be done by the system. The advantage of using the ThingWorx platform is its ability to perform advanced analytics and building a web-based interactive application to test and deploy a model so that it could be used by customers. Another strong point of this platform is the generation of profiles which are subpopulations within the data. These subpopulations give an Intel about the parameters that are the most correlated to the goal and that give a better accuracy when taken into consideration together.

The model results are shown in different ways depending on the goal variable's type. For the model used in this study, since the number of low-pressure alarms is a continuous variable, the results will be represented by a bubble plot. In addition to that, the root average square error and the normalized one, the R-squared, and the Pearson correlation are also calculated by the model.

The bubble plot is a graph where the x-axis is the actual results, and the y-axis is the predicted ones. It gives an approximate location of the records with bubbles. Each bubble represents one or more records with predicted and actual values near the bubble's location since these records are binned for charting. The size of the bubble is related to the number of records in it while its color is linked to how good the prediction was. Dark blue is used for the most accurately predicted records, which means predicted results are within 25% of the actual results. Light blue is for the under-predicted records and that means that the predicted results were 25 to 75% below the actual ones. Red is equivalent to highly under-predicted outputs thus predicted results are more than 75% below the real ones. Orange is for over-predicted where predicted records are 25 to 75% above the actual results. Finally green means highly overpredicted and these are the predicted values that are more than

75% above the actual ones.

The root mean square error (RMSE) measures the difference between the model's predicted values and the actual values. The lower the RMSE is, the better the model is. The Pearson correlation measures how much a linear correlation exists between the predicted and the actual results, it ranges from -1 to 1 where -1 means that they are inversely correlated, 0 means that no linear correlation exists and 1 means a high linear correlation is present. The R-squared measures the proportion of the variance for a dependent variable that is explained by independent variables in a model.

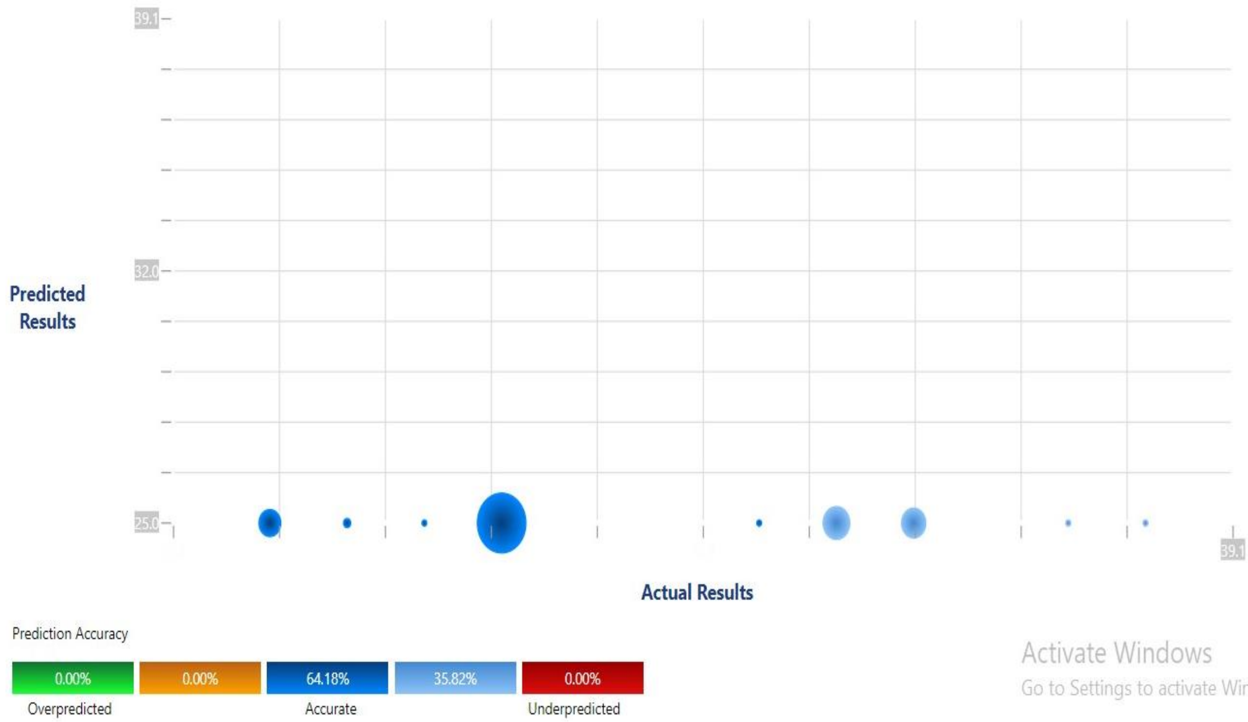

Figure 112: Number of Low-Pressure Alarms Model Results

As shown in Figure 113, the created model that takes 80% of the data to learn from and 20% to validate was able to predict 64.18% of the unseen values accurately (dark blue bubbles) and 35.82% were under-predicted (light blue bubbles). There were no over nor highly over-predicted records nor highly under-predicted values. The RMSE of the model came out to be 6.9551 and the normalized RMSE was 0.3661. The Pearson correlation value is 0 and the R-squared is -4.4415.

Now by looking at each parameter alone, no good interpretation could be withdrawn. This is why all of them should be considered to know how good this model is. The accuracy of the model was taken to be around 65%, which was considered a good result for the first implementation of the model but not enough to rely on. In addition to that, the value 0 of

the Pearson correlation meant that no linear correlation whatsoever between the predicted records and the actual ones existed. This cannot be a good model indicator because as the real numbers of low-pressure alarms increase, the values of the predicted number of low-pressure alarms should also increase and that is not happening here. For the RMSE and especially the normalized root mean square error, the value by itself even if it seems a bit low and a positive parameter, will not be meaningful. When other models in the future are created or even the same model but with a bit of fine-tuning, the new RMSE values should be compared to this one to see if there is any improvement towards a more efficient and reliable solution. Finally, the R-squared which gave a negative value means that the model is performing worse than taking the mean value of the number of low-pressure alarms as a prediction. These results could be due to the quantity of data used to train and evaluate the model. As mentioned earlier, the model takes 80% of the data to train on and uses 20% for validation thus if originally there was not enough data for failures or even for normal working conditions, this will lead to weaker training and a not so meaningful validation.
Only one profile was generated by the model containing only two parameters, the blower's working hours and the lube oil temperature. This means that these were the two most influential parameters of the model's output. The blower's working hours are linked to the goal because the more the blower will work, the more the probability of having failures will increase. On the other hand, the lube oil temperature is linked to the goal by the fact that by cooling down the top drive system, the lube oil's temperature will also be affected and reduced.
Moreover, a predictive scoring test was performed on the model that was able in the end to predict 25 alarms out of 38. Before reaching the value of 25, the predicted values were always lower than the real ones. This proves that the model is still not ready to be used and can still be improved.
As a next step to improve the accuracy of the model, several steps or suggestions are to be done:

- Collecting more and more data for both failures and normal operating conditions. This could be done in the same way as it was done for this study, through the ProRig system, helping the model to have more data to train on and to validate to have more meaningful and better results.

- Adding new and different sensors to the existing ones of the top drive system, since only one profile was generated containing only two parameters. By doing that, more parameters correlated with the goal will be available leading to better results and predictions. This happens because the model will be able to find more profiles with higher complexity and a number of parameters will be helping the model in having higher accuracy. Some of these sensors could be multiple vibration sensors on the blower that would help in seeing if any abnormal vibration could be linked to its failures. Another one is a temperature sensor on the motor of the top drive because low pressure from the blower means less cooling hence more heating of the motor. These additional sensors especially the vibration ones have a high chance of improving the results because [41] and [42] used vibration sensors on an industrial air blower's bearings to capture the vibration in different directions. By doing that they were able to find a correlation between the blower's failures and the vibration's anomaly and predictive maintenance was performed by condition-monitoring.

These new data, if collected in future works, should be like the one cleaned and prepared for the model to have reliable outputs. Along with trying and improving this model, new models built from different combinations of machine learning techniques than those used here can also be trained. After that, a comparison between all the models should be done to see which has the better accuracy in predicting the number of low-pressure alarms. To know if any improvement is happening with more data and new models, the value of the accuracy should increase, one the Pearson correlation should be getting closer to one to have a linear correlation between the model's output and the real outputs meaning that when one is increasing, the other is also increasing. Concerning the RMSE values, the better model is the one having the lowest value. Finally, for the R-squared, the value that is closer to one is the best because it means that the model is performing much better than simply taking a mean value of the goal as the prediction. When the final best model is chosen it should be tested before relying on it for a predictive maintenance program and always updated because the external conditions where the rig is working differ a lot from one place to the other and that could affect the model's performance.

9.2 HPU's Heat Exchanger Failure Prediction

The heat exchanger of the hydraulic power unit is responsible for the cooling of the hydraulic oil with fans. For this component, the only two directly correlated parameters to its performance that were recorded are its working hours and the hydraulic oil temperature in the tank which is equivalent to the outlet temperature of the heat exchanger. Even though the correlation between these two was clear and obvious as shown in the following Figure 114, it was not possible to get any significant key performance indicator (KPI) from it. Plus no failure occurred in this equipment and that made it impossible to see how these parameters would behave in that case.

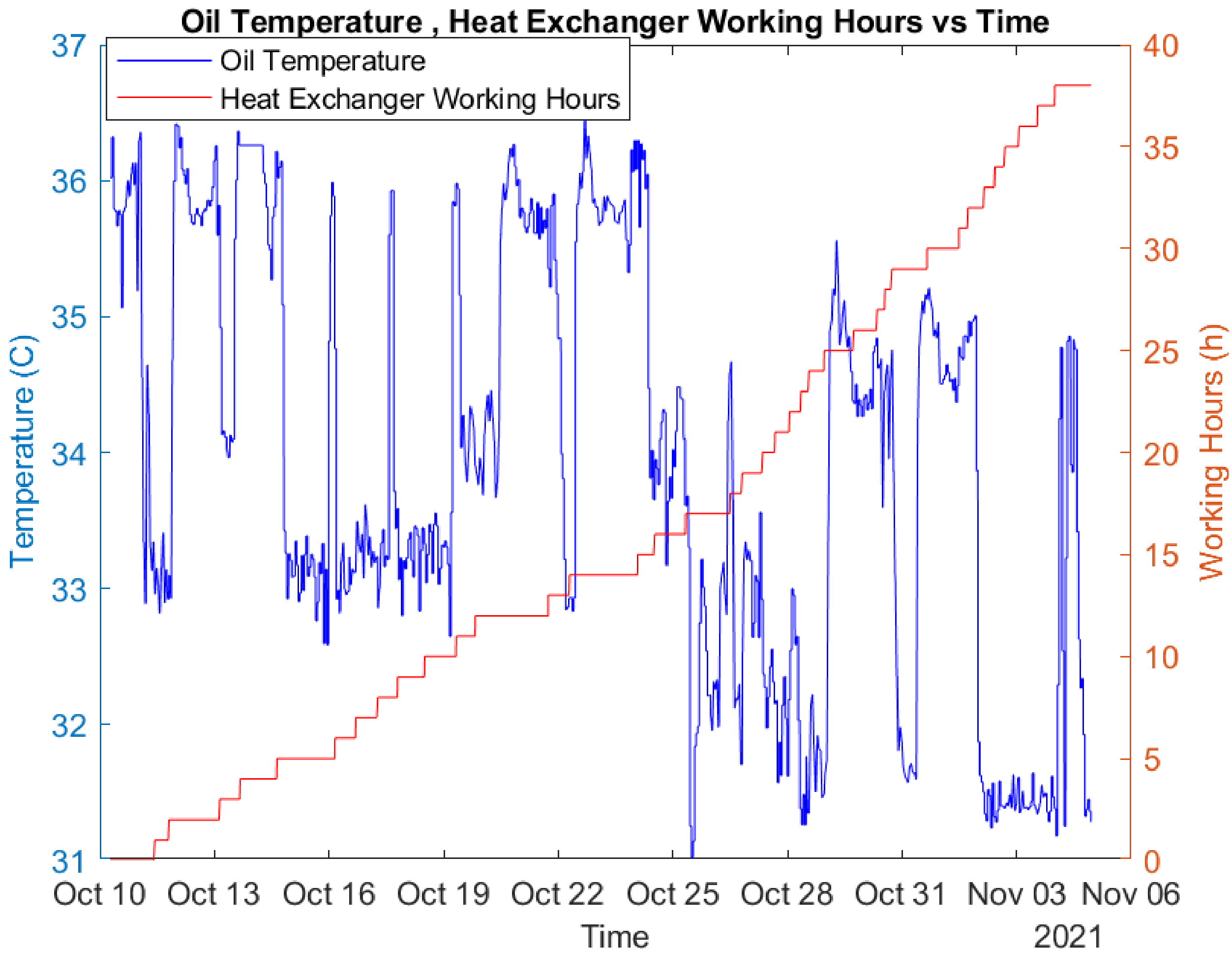

Figure 113: Hydraulic Oil Temperature vs. Time (blue) and Heat Exchanger Working Hours vs. Time (Red)

It is clear to see in above Figure 114 that whenever the heat exchanger working hours increased, the hydraulic oil temperature decreased. Moreover when the heat exchanger was not working (working hours remained constant), the temperature increased but there was never any alarming temperature that could mean that there was a problem that needed to be solved.

For future work, many refinements can be done to find a KPI and build a predictive model for the heat exchanger's failure. The most important would be adding several sensors such as the inlet temperature of the hydraulic oil, and the inlet temperature of the air blown to cool the oil. By doing that it would be possible to calculate the efficiency of the heat exchanger and get an important KPI to monitor and evaluate its performance. Other good sensors to add can be vibration sensors on the fans of the heat exchanger to capture whenever a malfunction happens. An inlet and outlet flow rate sensor that measures the flow of hydraulic oil going inside and outside to capture if any leakage is happening inside the heat exchanger's pipes. All these parameters could be beneficial for the model because it helps it in finding more correlations with the failure of this equipment and thus have higher accuracy and efficiency. Now adding sensors alone won't be enough, failures should occur during the recording time to let the model learn how these parameters behave in both normal and abnormal conditions because it cannot predict something it hasn't learned. If this happens all the same steps should be repeated from collecting the data up until the final model is generated.

9.3 HPU's Filter Clogging Prediction

The hydraulic power unit's filter has the role of cleaning the hydraulic oil by removing all the contaminants inside it. Even though here a filter clogging alarm was present, unfortunately no failures were recorded. In addition to that, the collected and cleaned data did not contain any parameters to conclude from them how far the filter is from getting clogged. For this, to build a predictive model in the future to capture the filter clogging before it occurs, many adjustments can be done, such as:

- Keep recording until several filter clogging alarms are triggered meaning a failure has occurred.

- Adding sensors to capture more meaningful parameters correlated with this goal. These sensors are pressure and flow rate sensors at the entrance and exit of the filter. By using these, a good KPI could be extracted and that is the difference between the pressure out and in, as well as the difference between the flow rate in and flow rate out. The flow rate coming out from the filter decreases with time due to the obstruction of the cross-section which means that the difference between the flows in and out would increase. On the other hand, the pressure drop will increase when getting closer to the filter clogging. As in [43] and [44] by using pressure drop, they were able to predict when the filter will clog, this means that these sensors if added have high chances of success and helping the predictive model.

After adding the sensors and collecting new data containing the filter clogging event, which means that the alarm of filter clogging is triggered, the same steps should be taken. The data must be cleaned, and multiple models with different mixtures of machine learning techniques must be created by using ThingWorx analytics. All these models should have the same goal variable and that is the filter clogging alarm. The next step would be comparing the generated models' accuracy, RMSE, R-squared, and Pearson correlation parameters to see which one is better. After the best model is chosen, testing on real cases should be performed before implementing it for predictive maintenance

.

9.4 HPU's Hydraulic Pumps Failure Prediction

The four hydraulic pumps of the hydraulic power unit PD1, PD2, TD1 and TD2 are responsible of pumping the hydraulic oil to different equipment. For them also there were no recorded failures and even though the outlet pressure and flow rate of each one of them was recorded, no intuitive key performance indicator could be deduced. Additionally these parameters don't tell a lot about the health of the pump.
For future work many sensors could be added to collect more meaningful data in correlation with the pumps' health such as:

- Multiple vibration sensors could be mounted on different parts of the pump. After that, the data recording starts in normal operating conditions and continues until a failure occurs. Then the data are taken, plotted and cleaned. Finally, using ThingWorx "signal" property, it is possible to know which vibration sensors are the most useful and linked to the prediction of the goal. Moreover, the "profile" generation property creates, as explained before, subpopulations of parameters that together help in predicting the output's value. By doing this, only the most important and correlated vibration sensors should remain and should be implemented to create the predictive model, since that helps in reducing the costs by reducing the number of sensors and by that reducing the amount of recorded data.

- The motor coil temperature values could also be in great correlation with the hydraulic pumps failures.

These sensors, in addition to the existing ones, could be enough to predict hydraulic pumps failures because in [45] they measured vibrations coming from houses of bearings because these are the places where the rotation parts (shaft) come into contact with stationary parts (casing). The horizontal, vertical and axial directions to the bearing axis were recorded. In [45], they mentioned that they were not only able to condition-monitor the pump, but they were also able to know the location from where the problem was coming. In addition to that, by measuring the values of the flow rate, the bearings vibration, the axial displacement and the motor coil temperature with a frequency of one record per hour, the pumps failures were predicted with an accuracy of 98.2% [46].
This means that by collecting the enough amount of data with failures and by cleaning and choosing the right model, predicting the failure of the hydraulic pumps is possible. Thus a predictive maintenance program can be put to application to avoid in the future sudden breakdowns and interruption of work.

9.5 TDS's Lube Oil Quality Prediction

The lube oil, as its name says , has the role of lubricating the top drive system thus reducing friction between the different components, as well as to dissipate some heat from the TDS. Over time, the lubrication oil's quality gradually degrades because of oxidation and after a certain time the oil has to be changed due to its inability of performing its role anymore. The goal was to try to predict when this lube oil should be changed. A very clear correlation between the motor's average RPM and the lube oil's temperature can be seen in Figure 115. When the RPM increases, the load on the oil increases, which leads to an increase in temperature. Oppositely, when the RPM is close to 0, the load on the oil decreases, leading to a temperature reduction.

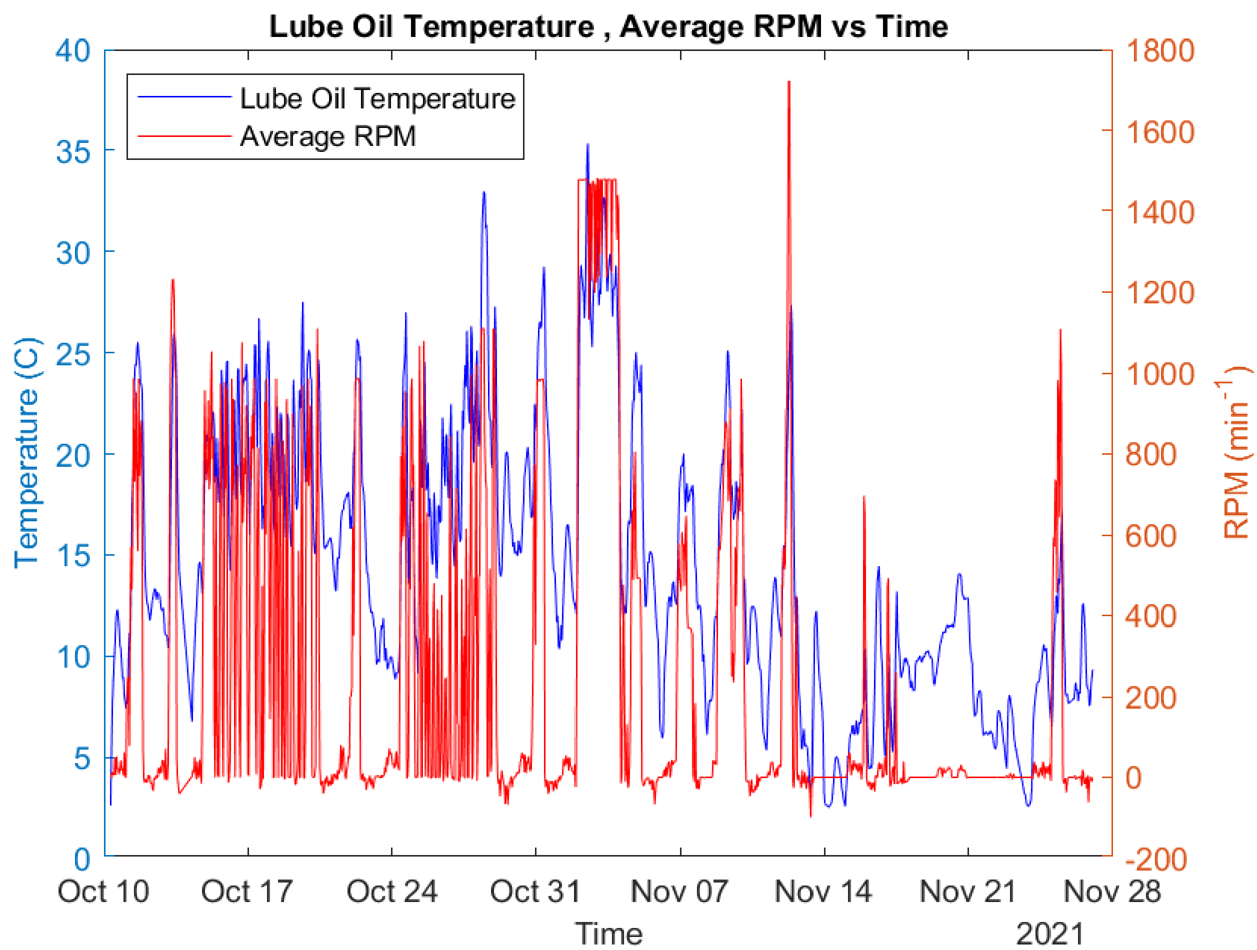

Figure 114: Lube Oil Temperature vs. Time (Blue) and Average RPM vs. Time (Red).

Although these parameters are important and could help in future works for predicting the oil's quality degradation, they were not enough to extract any KPI to measure oil performance.

A very important sensor could be added to help in achieving this goal, that is a kinematic viscosity sensor. Viscosity is one of the most critical properties to be considered when selecting the lube oil for a certain job. This property is very dependent on both the composition and the temperature of the oil. The temperature and viscosity are two inversely correlated parameters, meaning that when temperature increases, viscosity decrease and vice versa. As for the composition or quality of the oil, the oxidation of the oil increases over time, leading to a greater change in the composition of the oil, which involves a thicker oil, likely to contain also some contaminants, leading to an increase in viscosity. Moreover, a dielectric constant measuring sensor can also be used to help in monitoring the quality of the oil. The dielectric constant is a measure of the ability of a certain substance to store electrical energy. This constant increases with the degradation of the lube oil. As described in [47], kinematic viscosity and dielectric constant were used not only to monitor the condition of the oil, but also to predict its residual useful life; it was also specified that the same thing could be done with each parameter alone or both combined. Moreover, even using the kinematic viscosity, the residual useful life of the lubricating oil was predicted and it was possible to extend it by 26 months [48].

This means that, by using these two sensors in addition to the existing ones, it is possible to remotely monitor the quality of the lube. In this way it is also possible to know its residual useful life and, thanks to that, the oil change would only take place when it is necessary.

10 Conclusion

The goal of this project was to generate a model able to predict when the blower in the top drive system could fail, so that a predictive maintenance would be performed. This had to be done due to the numerous low-pressure alarms triggered by this component in a short period of time. In addition to generating this model, key performance indicators for the monitoring and the prediction of the performance of other components had to be found. These components were, on one hand, the heat exchanger, the hydraulic pumps, and the filter of the hydraulic power unit; on the other hand, the lube oil quality of the top drive system.

With the aim of achieving this goal,, many steps have been taken. First, using the Internet of Things data management system developed by Drillmec, the ProRig, it was possible to collect the raw data from the TDS, the HPU and the DAS. Secondly, the records coming from the different components of the rig were plotted. Then, a series of data cleaning was performed to prepare the following step. Finally, once the data were ready, by using ThingWorx analytics, a model for the prediction of the number of low pressure alarms of the blower was created. This model was generated with an ensemble of machine learning techniques: artificial neural network, decision tree, gradient boost and linear regression.

The results turned out to be with good accuracy, but not enough to consider the prediction model working. Furthermore, the RMSE, the R-squared and the Pearson correlation's values also meant that it was still possible to improve the model to reach the desired goal.

Regarding the heat exchanger, the hydraulic pumps, the filter and the lube oil, none of them were found to have any KPI for two main reasons: the first is that the parameters collected were not enough to do that; the second is that no failures were registered to verify if the recorded data would have been different from normal operating conditions.

For future work and to improve the results towards reaching the goal, a lot of suggestions were given for different equipment. Some of them were common such as collecting more data of both normally operating conditions and failures in addition to the generation of multiple models and comparing the outputs. The remaining propositions differed between them. For each apparatus some additional sensors were needed to improve the already existing model or to come out with important KPIs. That would help in performing the desired predictive maintenance, thus reducing the costs of maintenance as well as the risk of sudden breakdowns and would improve safety. All these recommendations have been supported by documents reporting real cases of monitoring and forecasting the performance of these components with the use of the suggested sensors. This means that in the future this project has a high probability of becoming a working predictive maintenance program for several elements of the drilling rig.

References

[1] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. d. P. Francisco, J. P. Basto and S. G. S. Alcala, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering,* vol. 137, 2019.

[2] S. Brdbury, B. Carpizo, M. Gentzel, D. Horah and J. Thibert, "Digitally enabled reliability: Beyond predictive maintenance," *McKenzie & Company,* 4 October 2018.

[3] D. T. Pham and A. A. Afify, "Machine-learning techniques and their applications in manufacturing," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture.,* vol. 219, no. 5, pp. 395-412, 2005.

[4] J. Thangaraj and R. L. Narayanan, "INDUSTRY 1.0 TO 4.0: THE EVOLUTION OF SMART FACTORIES," *Reaserch Gate,* October 2018.

[5] A. S. Gillis, "What is the internet of things (IoT)?," *TechTarget,* March 2022.

[6] B. Matt, "What is the Internet of Things? WIRED explains," *Wired,* 02 February 2022.

[7] C. McClelland, "IoT Explained - How Does an IoT System Actually Work?," *Leverage,* 26 October 2016.

[8] D. G. Micheal , "Chapter 5 - Maintenance Management," in *Facility Integrity Management*, Gulf Professional Publishing, 2016, pp. 87-139.

[9] W. Wu, "What is Reactive Maintenance? Types, Benefits, Cost, and Examples," *Coast,* 19 August 2020.

[10] B. Christiansen, "The role of reactive maintenance in your overall maintenance strategy," *MRO,* 25 May 2018.

[11] W. Wu, "What is Preventive Maintenance? Types, Benefits, Cost, and Examples," *Coast,* 18 August 2020.

[12] C. Eisner, "What Is Preventive Maintenance? The Beginner's Guide to Running PMs," *MaintainX,* 06 June 2022.

[13] D. D. Granberg, "Major Equipment Life-cycle Cost Analysis," Office of Transportation System Management, Minnesota, 2016.

[14] R. Bharati, "Assessment of Hire Rate Analysis of Heavy Equipment in Nepal," *Journal of Advanced Research in Business Law and Technology Management,* vol. 2, no. 1, pp. 1-13, 2019.

[15] W. Zhang, J. Guo, F. Gu and X. Gu, "Coupling life cycle assessment and life cycle costing as an evaluation tool for developing product service system of high energy-consuming equipment," *Journal of Cleaner Production,* vol. 183, pp. 1043-1053, 2018.

[16] C. Carnero, "An evaluation system of the setting up of predictive maintenance programmes," *Reliability Engineering & System Safety,* vol. 91, no. 8, pp. 945-963, 2006.

[17] M. Paolanti, L. Romeo, A. Felicetti, A. Mancini, E. Frontoni and J. Loncarski, "Machine Learning approach for Predictive Maintenance in Industry 4.0," IEEE, Finland, 2018.

[18] Z. M. Cinar, A. A. Nuhu, Q. Zeeshan, O. Korhan, M. Asmael and B. Safaei, "Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0," *Sustainability,* vol. 12, no. 19, 2020.

[19] S.-j. Wu, N. Gebraeel, M. A. Lawley and Y. Yih, "A Neural Network Integrated Decision Support System for Condition-Based Optimal Predictive Maintenance Policy," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* vol. 37, no.

2, pp. 226-236, 2007.

[20] K. Wang and Y. Wang, Advanced Manufacturing and Automation VII, Singapore: Springer, Singapore, 2018.

[21] M. Cakir, M. A. Guvenc and S. Mistikoglu, "The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system," *Computers & Industrial Engineering,* vol. 151, 2021.

[22] Y. Kosarenko, "How to Create Decision Trees for Business Rules Analysis," *Why Change Consulting,* 13 November 2021.

[23] A. J. Myles , R. N. Feudale, Y. Liu, N. A. Woody and S. D. Brown , "An introduction to decision tree modeling," *Journal of Chemometrics,* vol. 18, no. 6, pp. 275-285, 2004.

[24] A. Priyam , Abhijeet, R. Gupta, A. Rathee and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms," *International Journal of Current Engineering and Technology,* vol. 3, 2013.

[25] S. Suthaharan, "Decision Tree Learning," in *Machine Learning Models and Algorithms for Big Data Classification*, Boston, Springer, Boston, MA, 2016, pp. 237-269.

[26] S. I. Rufaida, J.-S. Leu, K.-W. Su and A. Haniz, "Construction of an indoor radio environment map using gradient boosting decision tree," Springer, 2020.

[27] J. Wang, P. Li, R. Ran, Y. Che and Y. Zhou, "A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree," *Applied Sciences,* vol. 8, no. 5, 2018.

[28] N. Aziz, E. A. P. Akhir, I. Abdul Aziz, J. Jaafra, M. H. Hasan and A. N. Che Abas, "A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems," in *2020 International Conference on Computational Intelligence (ICCI)*, Bandar Seri Iskandar, Malaysia, 2020.

[29] A. S. Ponraj, "Daily evapotranspiration prediction using gradient boost regression model for irrigation planning," *The Journal of Supercomputing,* vol. 76, 2020.

[30] H. Cao, M. N. Nguyen, C. Phua, S. Krishnaswamy and X.-L. Li, "An integrated framework for human activity recognition," in *UbiComp '12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, Association for Computing Machinery, New York, NY, United States, 2012, p. 621–622.

[31] D. H. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine," *Journal of Applied Science and Technology Trends,* vol. 1, pp. 140-147, 2020.

[32] H. Tran, "A SURVEY OF MACHINE LEARNING AND DATA MININGTECHNIQUES USED IN MULTIMEDIA SYSTEM," Department of Computer ScienceThe University of Texas, Dallas Richardson, 2019.

[33] J. Zenisek, F. Holzinger and M. Affenzeler, "Machine learning based concept drift detection for predictive maintenance," *Computers & Industrial Engineering,* vol. 137, 2019.

[34] K. Kumari and S. Yadav, "Linear regression analysis study," *Journal of the Practice of Cardiovascular Sciences,* vol. 4, no. 1, pp. 33-36, 2018.

[35] T. Le, M. Luo, J. Zhou and H. L. Chan, "Predictive maintenance decision using statistical linear regression and kernel methods," in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, Barcelona, Spain, 2014.

[36] M. Beretta, A. Julian, J. Sepulveda, J. Cusido and O. Porro, "An Ensemble Learning Solution for Predictive Maintenance of Wind Turbines Main Bearing," *Sensors,* vol. 21, no. 4, 2021.

[37] J. Xie, J. Huang, C. Zeng, S.-H. Jiang and N. Podlich, "Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering," *Geosciences,* vol. 10, no. 11, 2020.

[38] Z. Wang, Y. Wang and R. S. Srinivasan, "A novel ensemble learning approach to support building energy use prediction," *Energy and Buildings,* vol. 159, pp. 109-122, 2018.

[39] I. K. Nti, A. F. Adekoya and A. B. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data,* vol. 7, no. 20, 2020.

[40] I. H. Laradji, M. Alshayeb and L. Ghouti, "Software defect prediction using ensemble learning on selected features," *Information and Software Technology,* vol. 58, pp. 388-402, 2015.

[41] H. El-Rammal, S. El-Gayyar and S. H. El-Emam, "Vibration and Faults Prediction for Air Blowers - Case Study," *MANSOURA ENGINEERING JOURNAL, (MEJ),* vol. 41, no. 4, 2016.

[42] C. R. Kumar and S. Pujari, "CONDITION MONITORING AND DYNAMIC BALANCING OF A HOT AIR CIRCULATION BLOWER BY VIBRATION TOOL," *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY,* vol. 5, no. 3, pp. 40-49, 2016.

[43] O. F. Eker, F. Camci and I. K. Jennions, "Physics-based prognostic modelling of filter clogging phenomena," *Mechanical Systems and Signal Processing,* vol. 75, pp. 395-412, 2016.

[44] S. Bourrous, L. Bouilloux, F.-X. Ouf, P. Lemaitre, P. Nerisson, D. Thomas and J. Appert-Collin, "Measurement and modeling of pressure drop of HEPA filters clogged with ultrafine particles," *Powder Technology,* vol. 289, pp. 109-117, 2016.

[45] R. O. Saied, M. S. Mostafa and H. A. Hussein, "Predictive Maintenance Program Based on Vibration Monitoring," in *Design and Modeling of Mechanical Systems - II*, Hammamet, Tunisia, 2015.

[46] P. F. Orru, A. Zocchedu, L. Sassu, C. Mattia, R. Cozza and S. Arena, "Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry," *Sustainability,* vol. 12, no. 11, 2020.

[47] J. Zhu, J. M. Yoon, D. He and E. Bechhoefer, "Online particle-contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines," *Wind Energy,* vol. 18, pp. 1131-1149, 2015.

[48] N. M. Nazari, M. Muhammad and A. A. Mokhtar, "Prediction of Lubrication Oil Parameter Degradation to Extend the Oil Change Interval Based on Gaussian Process Regression (GPR)," *Tribology Online,* vol. 17, no. 3, pp. 135-143, 2022.