An automated method to determine dynamic pump efficiencies in deep-level mines

M Modiba

orcid.org/0000-0001-6967-0588

Dissertation accepted in fulfilment of the requirements for the degree *Master of Engineering in Mechanical Engineering* at the North-West University

Supervisor: Dr JC Vosloo

Graduation: November 2022

Student number: 37202944

ACKNOWLEDGEMENTS

Firstly, I would like to thank God for this opportunity. This endeavour would not have been possible without Him.

I would also like to extend my deepest gratitude to my family and friends who kept me motivated throughout this journey.

I wish to acknowledge the guidance provided by Mr. S. Taljaard and Dr L. Watters in writing this dissertation. Thank you for your ever-present support.

Thank you to Dr J.C. Vosloo for supervising this study and always providing detailed feedback.

I would like to thank Prof. E.H. Mathews and CRCED Pretoria for giving me the opportunity to do my master’s. I would also like to thank ETA Operations for funding my research.

ABSTRACT

Title: An automated method for determination of pump efficiencies in deep-level mines

Author: M Modiba

Supervisor: Dr JC Vosloo

Degree: M.Eng. (Mechanical Engineering)

Keywords: dewatering, pump efficiency, continuous, automatic, data quality, DA

Deep-level mines use dewatering systems that are responsible for the removal of underground water. A dewatering system consists of a series of dewatering dams and pumps. Dewatering pumps are responsible for pumping water from lower-level dams to upper-level dams. The water is moved through pumps and pipes until it reaches the mine's surface.

Due to harsh conditions under which dewatering pumps operate, their efficiencies deteriorate rapidly. Research revealed a gap in literature pertaining to quantifying dewatering pump efficiencies in deep-level mines, considering the data challenges faced by the industry. The challenges relevant to this study were found to be missing data that occur due to limited instrumentation or communication issues, static data challenges due to faulty instrumentation issues or communication issues, and inaccurate data such as negative values.

This study followed a data analysis approach to develop a methodology through which the data challenges were addressed. The data analysis process revealed that pump discharge flow rate, discharge pressure and power consumption readings were critical to the accuracy of efficiency measurements. A unique and automated method was developed through which dewatering pump efficiencies can be quantified despite the challenges.

The developed method was tested on two case studies. The case studies are from a deep-level gold mining group in South Africa. Results from both case studies highlighted the data challenges and thus the importance of having a data quality monitoring system. From the results, it is also clear that, despite the data challenges, pump efficiency is quantifiable.

The solution was automated to require minimal human intervention and can also help mines reduce the cost associated with hiring external contractors who conduct efficiency audits. The solution can also help enhance labour productivity. The developed method can be applied to any deep-level mine that uses dewatering pumps.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BDA	Big Data Analytics
CRISP-DM	Cross Industry Standard Process for Data Mining
DA	Data Analysis
IEEE	Institute of Electrical and Electronics Engineers
KDD	Knowledge Discovery in Database
NRSA	Nominal Range Sensitivity Analysis
OPC	Open Platform Communications
PLC	Programmable Logic Controller
RTB	Reporting Toolbox
SA	Sensitivity Analysis
SCADA	Supervisory Control and Data Acquisition
SEMMA	Sample, Explore, Modify, Model and Assess
TOU	Time-of-Use

CHAPTER 1

INTRODUCTION AND LITERATURE STUDY

1.1 Background

1.1.1 Deep-level mine dewatering systems

Some of the deepest mines in the world are found in South Africa [1]. Deep-level mines can reach depths of about 4 km [2]. With such depths, a significant amount of water accumulates underground [3]. The water originates from fissure and mine service water that is used for mining activities such as drilling, rock cleaning and cooling of the working areas [4], [5]. The water needs to be extracted to prevent flooding of the mine. A mine dewatering system is responsible for removing the water from underground levels to the surface [5], [6], [7].

A dewatering system comprises an upward cascading dam and pump configuration [4], [6], [8]. The system consists of pumping stations that are distributed on individual levels according to the layout and activities on the mine [5], [9]. A simplified layout of a typical dewatering system is shown in Figure 1.

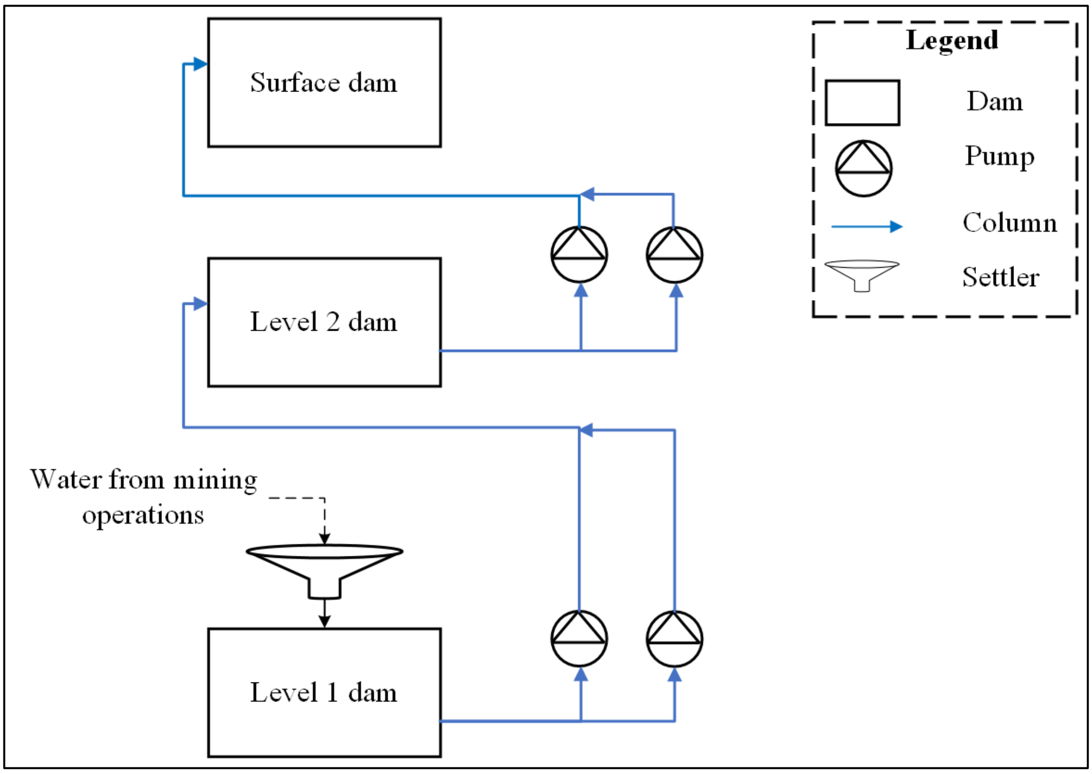


Figure 1: Simplified layout of a dewatering system

Lower mining levels usually have settlers that are responsible for removal of suspended particles in the water such as mud or debris. The particles accumulate at the bottom of the settler dam, and only the cleaner water flows into the water dam [4], [6], [10]. The water is pumped from a lower-level dam to an upper-level dam. The water is delivered through large pipes, also referred to as pumping columns. The cycle is repeated until the water reaches the surface. Some of the water might be used as service water on its way up [6]. The remaining water is pumped to surface and often cooled and recirculated for reuse by the aforementioned mining activities [6]. Typically, two or more pumps are found per pumping station, depending on the size of the mine and the amount of water that is required to be pumped [4], [11], [12].

1.1.2 Centrifugal pump theory

Multistage centrifugal pumps are mostly used in the deep-level mining sector [4], [13]. They are preferred due to their durability and tolerance to harsh environments [14], [15]. They are also easier to operate and are less prone to wear [16]. Figure 2 shows a simplified illustration of a typical single-stage centrifugal pump[1].

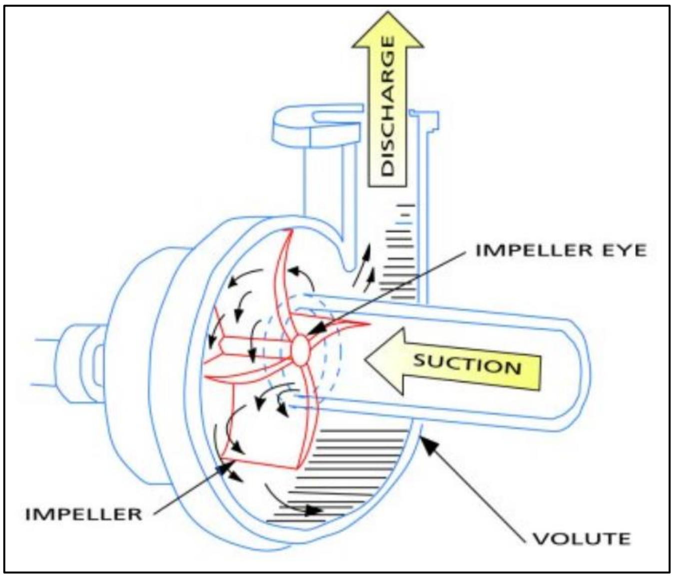


Figure 2: Simplified overview of a centrifugal pump

As illustrated in Figure 2, a single-stage centrifugal pump consists of an impeller attached to a shaft, which is driven by a motor and rotates in a volute casing [17]. The rotating impeller transfers kinetic energy to the fluid, which is later converted to a static head in the volute casing [17]. Static head in its simplest definition is the maximum height that a pump can move fluid against gravity [18]. Figure 3 shows a simplified depiction of the static head (adapted from [19]).

[1] S. Tembhre. "Centrifugal Pump Design and Analysis Using Solidworks." https://skill-lync.com/student-projects/centrifugal-pump-design-and-analysis-102 (accessed July, 2021).

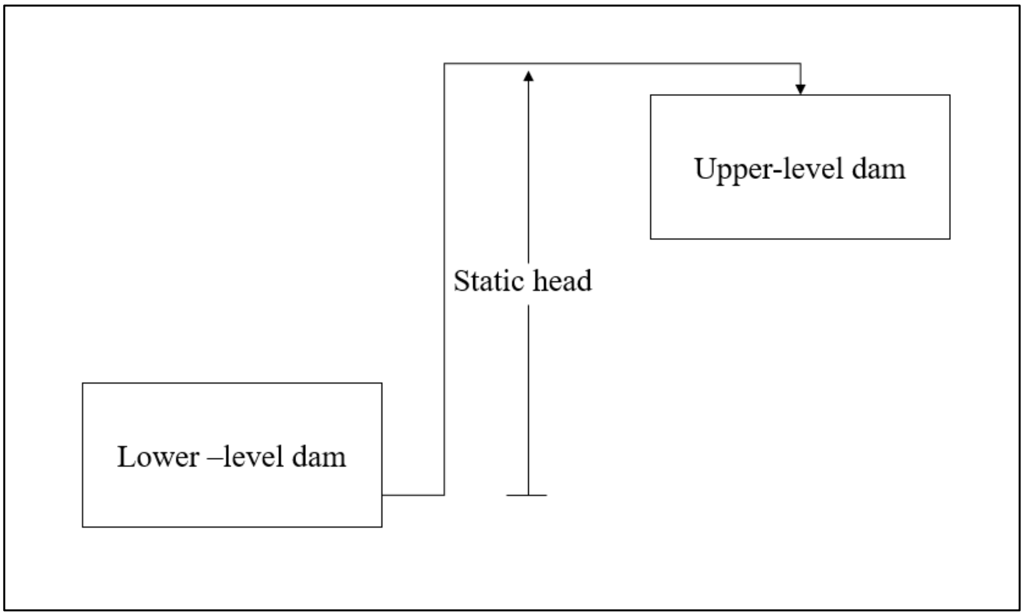


Figure 3: Simplified depiction of static head

A multistage centrifugal pump consists of multiple successive “single stages”, allowing the pump to impart more energy [15]. Centrifugal pumps can be classified into three categories, namely radial flow (liquid enters through the centre of the impeller and is pushed out along the impeller blades perpendicularly to the shaft), axial flow (fluid is pushed parallel to the shaft as in the case of a propeller), and mixed flow (combination of radial and axial flow) [15], [17].

1.1.3 Pumping energy consumption

One of the main disadvantages of pumps is their high energy consumption [4], [20]. Studies reveal that 15% of a mine’s total electricity is used for pumping [21], [22]. The cost of electricity in South Africa has been rising at a rate higher than inflation in recent years [23]. Due to high electricity demand, Eskom introduced time-of-use (TOU) tariff structures. Mines and other industries that have a notified maximum demand of greater than 1 MVA are billed through the Megaflex tariff structure [12]. The structure consists of electricity charges that are based on three TOU periods; namely off-peak, standard, and peak times. Figure 4 shows the 2021/22 Megaflex tariffs for a typical gold mine in South Africa (adapted from [24]).

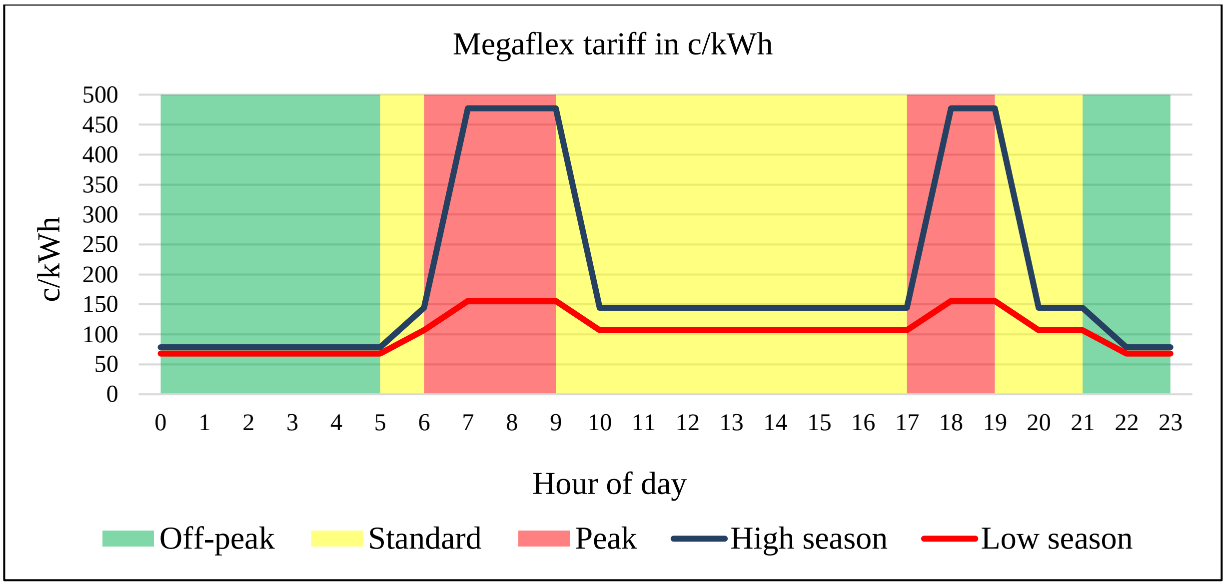


Figure 4: Megaflex tariff in c/kWh for 2021/22

Figure 4 shows that the tariffs vary according to seasons, with winter (referred to as the high season) having the highest tariffs. Mines must minimise their expenses on electricity to remain profitable. Mines usually implement load shifting strategies to reduce electricity costs. In the case of a dewatering system, the load shifting is typically implemented by pumping as much water outside peak TOU and avoiding pumping during peak TOU [25], [10], [13], [26]. Some energy management projects suggest considering pump performance when allocating pumps during the different TOU periods [4], [27]. The purpose of this intervention is to minimise financial cost during peak and standard TOU by using best-performing pumps during these periods [27]. Thus, it is important to know dewatering pump performances.

1.1.4 Quantifying pump performance

The performance of a centrifugal pump can be measured through various output parameters such as volumetric flow rate or discharge pressure. However, the most frequently-used parameter is efficiency [28]. Pump efficiency (η_{pump}) is defined as the ratio between useful energy delivered to the fluid as work and the total energy consumption of the motor, and can be expressed as follows [16]:

$$\boldsymbol{\eta_{pump}}\ (\%) = \frac{\boldsymbol{\rho * (DP - SP) * Q}}{\boldsymbol{P}} * \boldsymbol{\eta_{motor}} \qquad \text{Eq. 1}$$

where:

- ρ is the density of water and is considered as 1 g/cm^3 at 25˚C [29].

- DP is discharge pressure – this is the pressure at the pump's outlet and is proportional to the total static head required by the system. As mentioned in 1.1.2, static head is the total elevation of the discharge column and represents the pressure the pump must overcome to meet the pumping need [30]. DP is measured in kPa or bar and can be calculated as follows [31]:

$$\boldsymbol{Discharge\ pressure}\,[\boldsymbol{kPa}] = \boldsymbol{Static\ head}[\boldsymbol{m}] * \boldsymbol{g}[\boldsymbol{m^3/kg.s^2}] \qquad \text{Eq.2}$$

where: g is the gravitational constant (9.8 $m^3/kg.s^2$).

- SP is suction pressure – this is the pressure generated by the weight of the water contained in the dam from which a pump is connected. Therefore, it is dependent on the water level in the dam and position of the pump [6], [30], [19]. SP is also measured in kPa or bar.
- Q is the volumetric flow rate – this is the total water volume delivered by the pump per unit of time [19]. The volumetric flow rate is inversely proportional to the static head [31]. An ultrasonic flow meter is typically used to measure the volumetric flow rate on the discharge column. The units are usually in litres per second (l/s) [23], [19].
- P is the motor power consumption – this is the total power consumption of the motor while in operation and is measured in kilowatts (kW). Each piece of electrical equipment has a rated motor power, which is often provided on the nameplate. Most electric motors are designed to run at 50% to 100% of the rated power [32]. Electric motors have a tendency to consume more than the rated power during start-up [32], [33]. The motor power consumption can be calculated from the motor electric current using the three-phase power formula as shown below:

$$\boldsymbol{Motor\ power\ consumption}[\boldsymbol{kW}] = \sqrt{\boldsymbol{3}} \times \boldsymbol{PF} \times \boldsymbol{V} \times \boldsymbol{I} \qquad \text{Eq.3}$$

where: PF is the power factor, the phase angle between the motor electric current and voltage and a value of 0.9 is typically assumed [34], V is the voltage at which the motor operates (typically 6.6 kV for mines [24]) and I is the electric current.

- η_{motor} is the motor efficiency – this is the ratio of the motor output power to the power input into the system[2]. The output power is what is used to do mechanical work. The motor efficiency is usually provided by the manufacturer, and ranges between 90 – 98%. However, a motor efficiency of 100% will be assumed as it is not under investigation.

[2] Ramzy. "The Simple Guide to Motor Efficiency: What It is and What to Do." https://www.linquip.com/blog/motor-efficiency/ (accessed: Nov., 2021).

As mentioned, pump performance can be measured through various output parameters such as volumetric flow rate or discharge pressure. However, from these parameters, pump efficiency is the most comprehensive metric as it compares the electrical input into the system versus the mechanical output, therefore revealing any losses [18], [28]. Therefore, this study will use pump efficiency as a performance metric.

1.1.5 Pump condition monitoring

Stols [35] found that it costs the mining industry approximately 20% to 50% of its total operational cost to maintain machines. It also costs approximately three to five times more to repair a failed component than to replace it before it experiences failure, as it might cause downtime in production [35]. According to Oberholzer [4], if correct maintenance is conducted, the regular life cycle (15 years) of a pump can be increased by up to 15%. It is thus imperative that the conditions of the pumps be monitored.

Pump condition monitoring tools in most South African mines only focus on temperatures and vibrations of the motor and pump as well as the motor power consumption [35]. While these parameters are key from a safety perspective, they do not explain whether a pump executes its mechanical duties effectively. Pump efficiency is a metric that provides a comprehensive overview of a pump's performance.

Due to poor water quality and inefficient operation, the efficiencies of dewatering pumps deteriorate rapidly [4], [36], [37] [38], [39], [40]. According to Anis [39], the efficiency of a pump deteriorates linearly to its total running hours. This means that the more a pump operates, the more maintenance it requires. Thus, it is necessary to monitor the efficiencies continuously and automatically to keep up with the dynamic nature, and act upon any suspicious pump behaviour.

As mentioned in Section 1.1, pumps can be selected based on performance to realise cost savings. The selection criteria can use pump efficiency as a performance metric [41]. Pump efficiency can also be used as an additional parameter to improve condition monitoring [16], [40].

Most mines outsource services relating to pump performance monitoring to external contractors. According to Statista [42], employee salaries and external contractors account for approximately 31% of the operating cost of a mine. In most cases, the work that is carried out by these contractors involve conducting scheduled periodical on-site audits. This process is time and resource consuming as the mine must allocate a foreman to provide supervision. With the manual work, there is also a possibility of human error that can affect the accuracy of results [27]. Therefore, automating the process can reduce human error and hence provide more accurate feedback.

Automation will also ensure that timeous feedback is provided given the dynamism of the efficiencies [27].

To carry out any kind of monitoring or reporting, data is needed. In the mining sector, various machinery is equipped with a wide variety of sensors [43]. However, most of the data produced by these sensors are not used [43].

1.2 Data challenges in deep-level mines

1.2.1 Data acquisition

Mines have been investing in means of monitoring their systems. Most South African mines use a common data acquisition process [22]. Figure 5 shows a diagrammatical representation of the process used at most mines [22].

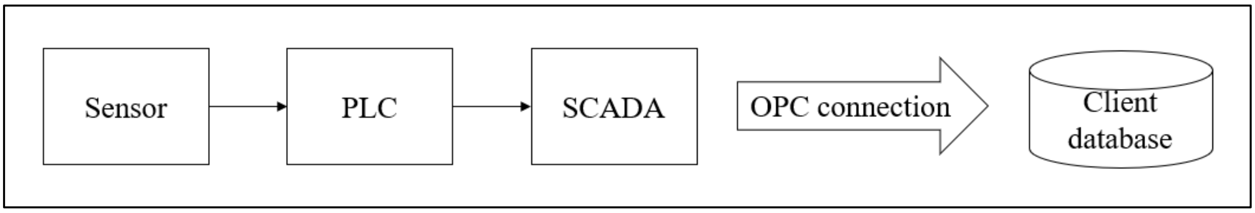


Figure 5: Overview of data transmission process

As shown in Figure 5, the process consists of a sensor(s), programmable logic controller (PLC), supervisory control and data acquisition (SCADA), open platform communications (OPC) and a database. The above steps/tools are described below:

Sensor

These are all devices that are physically placed on components that are of interest, for example loggers attached on pumps. Multiple sensors can be mounted on the same component [22]. In this study, the measurements of interest are discharge flow, discharge pressure, suction pressure, and power consumption.

Programmable logic controller (PLC)

A PLC is an electronic device that consists of a programming memory and can execute certain functions such as logical instructions and timing. PLCs can control analog or digital modules [44], [45], [46]. The PLC gets control signals from the SCADA, which are composed of setpoint values, start/stop signals, and on/off signals.

Supervisory control and data acquisition (SCADA) system

The SCADA is responsible for monitoring, gathering, and processing real-time data from the PLC, and has a user interface with the sensors being monitored [44], [46], [47]. The SCADA is usually configured to log data at a set frequency [47].

Open Platform Communications (OPC)

The OPC is a system or platform that enables other systems and software to securely access data from the SCADA [22] [35], [48].

Database

The last phase includes storing the data in a database, which receives data from SCADA through an OPC connection [47]. Databases are designed according to the kind of data they are required to handle [22]. Data can be exported through a data exporter.

1.2.2 Data challenges

In many applications, raw data is summarised or aggregated before being stored in a database [49]. This is usually done for various reasons such as to reduce data volume and noise in the data [49]. Real-time data consists of a higher volume (more data points per time interval) and requires large storage space. For example, some mines log data at two-minute intervals on to SCADA. In some instances, the raw data can be aggregated to a lower resolution that will reduce the data volume before storing in a database.

The data acquisition process is highly dependent on electricity and a reliable internet or network connection for effective communication of the involved subsystems. Due to the nature of mines, power cables are sometimes damaged by rock falls, flooding of water in lower levels or other mining activities [50]. Other mining characteristics such as ore type, which may have electromagnetic properties interfering with signals, can also affect radio signals and therefore affecting communication between the systems [50]. This can compromise data availability and or quality. Measurements related to pumps are also affected by these challenges as the pumps are often located on the lower levels of the mine where the aforementioned activities take place.

There are other predominant data-related challenges in the mines, such as faulty sensor readings [51]. De Meyer [52] mentioned that faulty sensor readings can be a result of heavy equipment vibrating, which may cause sensors attached to it to read inaccurate data. These factors can compromise data quality. Some common signs of low-quality data are static data, negative values, outliers, etc.

Due to high volumes of water accumulating in dewatering dams, pumps are typically operated in pairs to meet the pumping demand. As mentioned in Section 1.1, pumps that located at the same pumping station usually share columns. In most cases, only a single flow meter is placed on a column. This implies that such a flow meter will only measure a combined flow, depending on the number of pumps running at a time. With only combined flow data available, it is not possible to determine each pump's (separate) efficiency if there are multiple pumps running at once. Different pump combinations can also alter the running efficiency of a pump [53].

As mentioned in the above discussion, these factors can negatively affect data quality and usability. According to Lettner *et al.* [54] and Hellerstein [49], poor data quality can cause errors in analytical processes, which may result in wrong decisions. Different researchers defined data quality as multidimensional and further proposed various sets of data quality dimensions such as accuracy, accessibility, completeness, etc. [55], [56], [57], [58], [59], [60], [61], [62]. A data quality dimension provides a set of techniques and methods that must be followed to address data quality with respect to that dimension [63]. After analysing these dimensions in the mining context, Goosen [51] revealed that missing data, static data, data elements that exceed specified limits, negative values, and outliers are found in various measurements on different mining equipment. Figure 6 shows the proportions of each erroneous data characteristic for pumps in the mining sector (adapted from [51]).

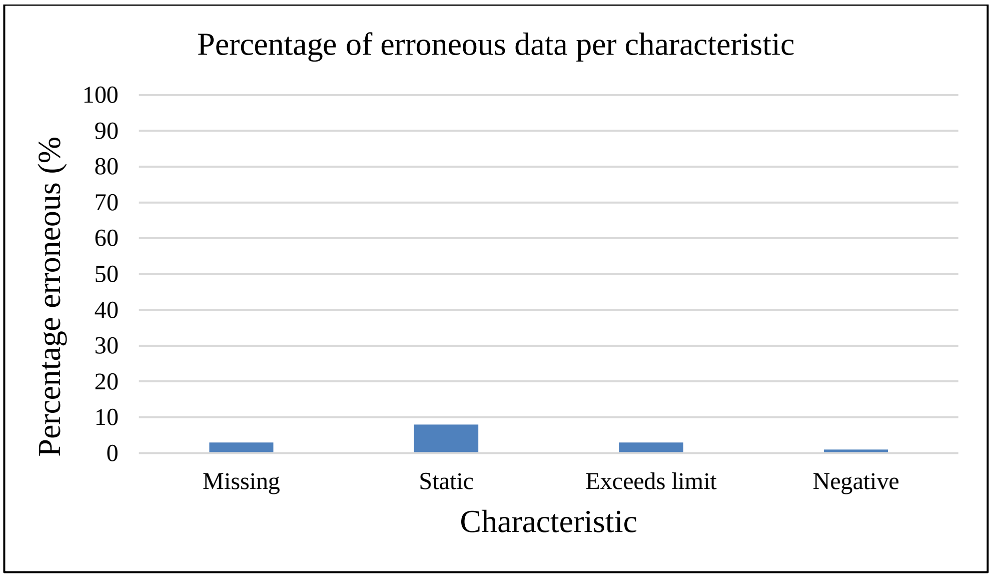


Figure 6: Percentage of erroneous data per characteristic

As can be observed from Figure 6, four data quality characteristics (missing, static, exceeds limits and negative) were found on pumps. Henceforth, the last two characteristics (exceeds limit and negative) will be referred to as outliers.

Goosen [51] further investigated the quality of data for different parameters on various equipment. The parameters were position, flow, vibration, running status, pressure, temperature, and power. Figure 7 presents erroneous data scores that were obtained for the parameters that are applicable to this study (adapted from [51]).

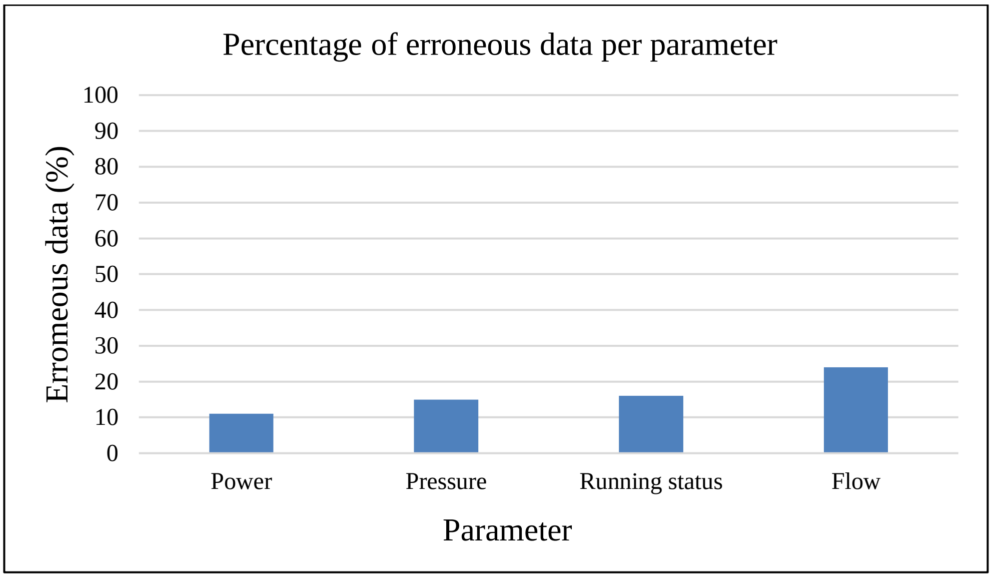


Figure 7: Percentage of erroneous data per parameter

From Figure 7 it can clearly be seen that at least 10% of data for each parameter involved in the efficiency calculation can be expected to be erroneous. Flow was found to have the most erroneous data of these parameters.

The findings made by Goosen are relevant to this study. Erroneous data can distort results and lead to inaccurate decisions being drawn. Therefore, these data quality issues must be addressed to ensure accurate reporting.

The above discussions about data challenges can be summarised to the following key points:

- There are data challenges in mines with regard to missing data that occur due to limited instrumentation or communication issues [50].
- There are static data challenges due to faulty instrumentation issues or communication issues [50].
- There are inaccurate data points (exceeds limit, outlier and negative).

All the above factors can negatively affect the quality of results and thus need to be considered when quantifying dewatering pump efficiencies. Data analysis (DA) can be used to study data and address data quality problems.

The next section will discuss existing relevant literature regarding the process of quantifying centrifugal pump efficiency.

1.3 Literature study

This section presents current existing literature applicable to this study. The studies will be thoroughly scrutinised. The purpose of this is to identify studies that implemented similar solutions and identify gaps which will be addressed by this study. The studies were evaluated according to the criteria detailed in Table 1.

Table 1: Description of criteria used to assess related studies

Criteria	Description
Pump efficiency	Study used pump efficiency as an indication of pump performance
Deep-level mining industry	Study conducted in a deep level mine
Data-related challenges	Study considered data challenges
DA	DA approach followed
Continuous monitoring	Efficiencies evaluated continuously
Automated	Efficiencies quantified automatically

Each study was evaluated on whether it based the pump performance investigation on efficiency. This is relevant since pump performance can be monitored through other parameters such as water volumetric flow rate, discharge pressure, etc. However, as mentioned in Section 1.1.4, pump efficiency provides a comprehensive overview of a pump's performance and is the preferred method for determining pump performance.

In addition, previous studies were evaluated on whether their investigations were carried out in the deep-level mining industry. This is relevant due to the uniqueness of deep-level mines, as discussed in preceding sections (conditions, water quality, instrumentation, and data quality).

As mentioned before, mines are faced with several data-related challenges which are seldom addressed. Hence, studies were also assessed on whether the challenges were considered when evaluating efficiencies.

Studies are also assessed on whether they followed a data analysis (DA) approach in dealing with data quality problems. A DA approach entails any data exploration and cleaning procedure that takes place before implementation.

In Section 1.1.4 it was mentioned that efficiencies of dewatering pumps are continuously deteriorating due to factors such as poor water quality and improper maintenance. It was subsequently established that efficiencies must be monitored continuously. Thus, studies that monitored pump performances on a continuous basis will be identified.

The importance of implementing an automated method for quantifying efficiencies was discussed in Section 1.1.4. Therefore, studies that quantified efficiencies continuously will be further assessed on whether they implemented an automated system.

Studies were searched for across various academic databases such as the Institute of Electrical and Electronics Engineers (IEEE), ScienceDirect, Scopus, Access Engineering, Elsevier eBooks, ResearchGate, etc. Initially, the studies were searched using the phrase "centrifugal pump performance". This was to identify studies that investigated pump performance regardless of the industry in which they were applied. To identify studies that were conducted in the deep-level mining industry, the search was narrowed to studies containing the phrases "centrifugal pump performance" and "deep-level mine". Nineteen studies that were closely related to this study were identified. Twelve studies were published journal articles and seven were academic conference papers. The studies varied from 2012 to 2022. Table 2 provides a summary of the criteria that each identified study met.

Table 2: Critical assessment of existing literature

Study	Pump efficiency	Deep-level mining industry	Data-related challenges	DA	Continuous monitoring	Automated
Smith *et al.* [27]	✓	✓	✗	✗	✗	✗
Shojaeefard *et al.* [29]	✓	✗	✗	✗	✗	✗
Vodovozov *et al.* [40]	✓	✗	✗	✗	✗	✗
Augustyn [41]	✓	✗	✗	✗	✗	✗
Bohn *et al.* [64]	✓	✗	✗	✗	✗	✗
Qazizada *et al.* [65]	✓	✓	✗	✗	✗	✗
Qazizada *et al.* [66]	✓	✓	✗	✗	✗	✗
Huang *et al.* [67]	✓	✗	✗	✗	✗	✗
Wang *et al.* [68]	✓	✗	✗	✗	✗	✗
Kara Omar *et al.* [69]	✓	✗	✗	✗	✗	✗
Martinus *et al.* [70]	✗	✓	✗	✓	✓	✗
Tan *et al.* [71]	✓	✗	✗	✗	✗	✗
Aiamsamang *et al.* [72]	✓	✗	✗	✗	✗	✗
Ping *et al.* [73]	✓	✗	✗	✗	✗	✗
Li *et al.* [74]	✓	✗	✗	✗	✗	✗
El-Naggar [75]	✓	✗	✗	✗	✗	✗
Matlakala *et al.*[76]	✓	✗	✗	✗	✗	✗
Jacobs *et al.* [77]	✗	✓	✓	✓	✗	✗
Al-Obaidi *et al.* [78]	✓	✓	✗	✗	✗	✗

Smith [27] and Augustyn [41] investigated energy efficiency optimisation of pumping systems holistically. Both studies explained pump efficiency and its importance to the overall efficiency of a pumping system. However, none of them quantified pump efficiency. Smith [27] discussed the formula used to quantify pump efficiency (similar to the one that was presented by Eq. 1). However, the study did not indicate whether efficiencies were practically quantified nor how the formula would be applied in the deep-level mining industry. Augustyn [41] also suggested that a pump must be selected for use based on its efficiency as published by its manufacturer.

Shojaeefard *et al.* [29] developed a method to quantify centrifugal pump performance parameters, namely head, input power and efficiency. The method was validated through an experimental procedure conducted on a test pump connected to a water tank. The data collection process followed in the experiment did not present the data quality compromising sources as is the case with data in the deep-level mining industry, hence the author did not mention any data quality assessment procedure. Furthermore, the author did not indicate whether the pump performance was continuously monitored nor whether the method was automated following the experimental process.

Vodovozov *et al.* [40] implemented a method to improve the efficiency of a pumping system by improving the efficiency of each pump in the system through optimal pump control. The author did not specify the industry in which the method was tested but considering the sizes of the pumps as specified by the author (power consumptions of approximately of 1kW, and they typically range between 800-2800 kW [79] in mine dewatering pumps), it can be concluded that it was not in the mining sector. Furthermore, the experimental tests were only conducted once to validate the study.

The study by Bohn *et al.* [64] proposed a system to quantify the performance of centrifugal pumps through efficiency, vibration and fluid pressure analysis. This study did not specify the industry in which it was based. The study also did not implement the method in practice as it was based on theoretically developing a novel method to quantify efficiency without using flow sensors.

Qazizada *et al.* [65] analysed the performance of centrifugal pumps based on head, power, flow and efficiency. The study was validated through tests conducted on test benches that included four pumps.

Qazizada *et al.* [66] investigated the reliability of centrifugal pumps in parallel and serial configuration in mines. The study evaluated head, power, efficiency, and net positive suction pressure of the pumps. Although the study was conducted in a mine, data challenges were not addressed by the study.

Huang *et al.* [67] and Kara Omar *et al.* [69] developed centrifugal pump performance prediction methods using an energy loss analysis method. The performance metrics used by both studies were head, power and efficiency. Both studies validated their results by conducting experiments on test rigs.

Like Huang *et al.* [67] and Kara Omar *et al.* [69], Wang *et al.* [68] developed a method to optimise the design and predict the performance of centrifugal pumps. However, Wang *et al.* [68] also included net positive suction head as an additional metric. The study was also tested and validated on a test bench.

Martinus *et al.* [70] developed a continuous method for evaluating pump performance in a mine. However, the study based the pump performance evaluation on flow rate and not efficiency. The author gathered the first set of condition monitoring data through questionnaires. This was to try to establish the type of condition monitoring that is practiced at the mine (predictive, reactive, etc.). The author then gathered the second set of pump-related data by visual inspection. These data were used for analytical purposes.

Tan *et al.* [71] investigated the performance of centrifugal pumps based on head, efficiency, vibration, pressure and flow. The study was validated through an experiment conducted on a test rig.

Aiamsamang *et al.* [72] implemented a system through which pump manufacturers can determine the efficiency of pumps on the production line (for quality control).

Ping *et al.* [73] developed an experimental centrifugal pump efficiency prediction model. In this study, experimental tests were conducted on a test bench.

The study by Li *et al.* [74] was similar to the one by Shojaeefard *et al.* [29] in terms of the metrics used to evaluate pump performance (head, input power and efficiency), as well as the experimental procedure carried out.

El-Naggar [75] implemented a theoretical model of determining centrifugal pump performance characteristics for design purposes.

Matlakala *et al.* [76] investigated the efficiency losses of centrifugal pumps at a bulk water utility. Like the other studies mentioned above, the study did not include any data analysis as it was not conducted in the mining industry where data challenges are prevalent.

Jacobs *et al.* [77] developed a method to predict failure of mine dewatering pumps. Although the study was conducted in a deep-level mine, the author did not address data challenges. The study included a data visualisation process to assess the usability of the data for the model that was developed. However, the data visualisation process used in the study is not applicable to this study as it was not addressing data quality issues. Furthermore, the method was not implemented to make continuous predictions nor was it automated.

Al-Obaidi *et al.* [78] developed a simulation-based performance evaluation for centrifugal pumps in a mine. The study also conducted experiments where pump characteristic curves, including head, power and efficiency were investigated. The study was conducted to serve as guidance for the design of high-power pumps.

From Table 2, as well as the subsequent breakdown and discussions for each study, it is clear that numerous published studies are available focusing on monitoring the performance of centrifugal pumps. It can also be observed from the table that most of the studies based their pump performance evaluation methods on pump efficiency. However, most studies were conducted in laboratory setups with state-of-the-art equipment. It is also apparent that none of these studies addressed the data-related challenges faced in deep-level mines. One study was identified which used DA and was conducted on a deep-level mine. However, the performance evaluation used by that study was not based on pump efficiency nor was the method of evaluating pump performance automated.

Thorough examination of available literature did not reveal any continuous pump efficiency quantification study that was implemented in the deep-level mining industry. It was found that most existing studies mainly focused their work on experimental outcomes, but did not implement methods to continuously quantify efficiencies automatically.

Moreover, there were very few studies that focused on quantifying pump efficiencies in deep-level mines. These studies also did not consider the data challenges encountered in deep-level mines.

1.4 Problem statement

It is essential to monitor the performance of pumps in deep-level mines as they account for a large portion of a mine's operational cost. Pump efficiency is the most comprehensive metric for evaluating pump performance. Knowing a pump's efficiency can help with condition monitoring and cost savings. Due to the nature of dewatering systems, these efficiencies are continuously deteriorating, and thus should be continuously and automatically monitored.

Various existing literature considered pump efficiency evaluation. However, two main shortcomings were identified. Firstly, there was no system to monitor the efficiencies continuously and automatically. Secondly, studies that were conducted in the deep-level mining industry did not address the prevalent data challenges faced by the industry.

Thus, a need exists to develop a method through which dewatering pump efficiencies in deep-level mines can be quantified continuously and automatically. The method should address data challenges encountered in deep-level mines.

1.5 Study objective

The main objective is to investigate and develop a method by means of which dewatering pump efficiencies in deep-level mines can be quantified. The method must address shortcomings identified in existing literature by:

1. Addressing the data-related challenges that are faced in deep-level mines through DA steps.
2. Building a system to automatically evaluate efficiencies on a continuous basis.
3. Making the system applicable to any dewatering pump from a deep-level mine.

1.6 Document overview

Chapter 1: Introduction and literature study

This chapter investigates the need for an automated method to quantify efficiencies of dewatering pumps in deep-level mines. Various data-related challenges that should be addressed are discussed. Shortcomings of previous relevant studies are considered. Finally, the problem statement and study objectives are defined.

Chapter 2: Development of solution

In this chapter, various generic steps adopted from the existing DA approaches are outlined. All steps followed in developing the methodology are also discussed. The procedure that will be followed when implementing the solution is described. The chapter is concluded with a description of the method through which the obtained results will be evaluated.

Chapter 3: Implementation and results

Chapter 3 presents and discusses the results obtained by following the methods described in Chapter 2. The results are presented for each case study on which the method was implemented. The results are validated through comparison with impartial third-party pump efficiency measurements.

Chapter 4: Conclusion and recommendations

This chapter provides a conclusion on the accuracy of the developed method, as well as an indication of whether the study objectives have been met. Limitations and recommendations for areas of further research are mentioned.

CHAPTER 2

METHODOLOGY

2.1 Preamble

The objective of this study is to investigate and develop a method by means of which the efficiency of dewatering pumps in deep-level mines can be automatically determined on a continuous basis. This study considers a method through which the objectives can be reached by addressing the data-related challenges that are present in the mines. This chapter provides the methodology followed to achieve the above-mentioned objectives.

2.2 Data analysis (DA) methods

The purpose of this study is to identify a method through which dewatering pump efficiencies can be quantified despite the data-related challenges found in deep-level mines. For this exercise to take place, data must be analysed to ensure that their quality is fit for use. DA refers to the process of sorting, organising, storing, processing, analysing and studying data with the aim of discovering trends and new knowledge from the data [80]. DA can be used to develop strategies to address data quality problems [49]. The following sections explore some of the quantitative DA models that were found in existing literature. Although most of these methods are focused on data mining problems, they are relevant to this study as they include generic DA steps that can be adopted. These methods define a set of key steps to be followed when carrying out DA. These DA methods are then combined to form the DA methodology developed for this study.

2.2.1 CRISP-DM

CRISP-DM stands for cross industry standard process for data mining. This is a standard and industry-independent model that was released in the year 2000 [81]. The method is still being used today in different applications. It consists of six major steps as detailed below [81].

1. Business understanding – this step is focused on formulating a preliminary plan to achieve the project's objectives and requirements.

2. Data understanding – this step involves initial data gathering as well as assessing the data to discover initial insights.

3. Data preparation – this step includes all activities involved in constructing a new dataset from the initial data.

4. Modelling – this step involves the application of various data modelling techniques.

5. Evaluation – this step assesses whether the developed data models meet the specified objectives.

6. Deployment – in this step, results are conveyed to an end user.

2.2.2 SEMMA

The SEMMA process consists of the following steps: sample, explore, modify, model and assess [82].

1. Sample – in this step, a subset is extracted from a large dataset.

2. Explore – this step entails the exploration of data to identify trends and anomalies.

3. Modify – in this step, data is modified by transforming the variables.

4. Model – in this step, software is used to find a combination of data that predicts a desired outcome reliably.

5. Assess – during this step, the validity and reliability of the findings are assessed.

2.2.3 Methodology followed by Goosen

Goosen [51] investigated a method to quantify data quality in industries. Various DA methods were investigated. From these methods, Goosen developed a method that contains the most common DA steps:

1. Problem formulation – in this step the problem is defined.

2. Obtain data – this step entails the collection, cleaning, and transformation of data for analysis purposes.

3. Explore the data – this step involves visualising the data to identify anomalies or patterns.

4. Model data – in this step, a model is built and validated.

5. Data product – in this step, insight gained from the analysis is conveyed to end users.

2.2.4 Big data analytics (BDA) process

A BDA process enables organisations to uncover knowledge patterns and optimise their business processes accordingly [83]. It involves the following six major steps.

1. Data collection – this step entails the collection of big data from multiple sources.
2. Prepare data – this step consists of data pre-processing and integration.
3. Model – this entails the development of data models through statistical analysis and machine learning tools.
4. Evaluate – the models developed are evaluated and validated using test data.
5. Deploy – this consists of the deployment of data models to a real-world scenario.

6. Monitor – the performance of the models is monitored through the accuracy with which they make predictions.

2.2.5 Methodology followed by De Meyer

De Meyer developed a system to evaluate the integrity of industrial condition monitoring data [52]. The methodology was as follows:

1. Obtain data – in this step data is obtained and archived.
2. Models – in this step models are built.
3. System – in this step a system to assess data integrity is built.
4. Verification – in this step the system is verified by feeding it with clean data.

2.2.6 A summary of the DA methods

This section summarises the steps followed by these DA methods. Some of the steps were found to be common amongst the different methods. This enables generating a methodology that combines the DA methods. Table 3 presents the steps grouped based on how they were defined in each method.

Table 3: Summary of DA methods identified

Step	CRISP-DM	Goosen	SEMMA	BDA	De Meyer
Business understanding/problem formulation	✓	✓	×	×	×
Data understanding/sampling/ selection/collection	✓	✓	✓	✓	✓
Data preparation/pre-processing/exploring	✓	✓	✓	✓	✓
Data transformation/modification	×	×	✓	×	×
Data modelling/mining	✓	✓	✓	✓	✓
Evaluation/assessment/verification	✓	×	✓	✓	✓
Deployment/data product/implementation	✓	✓	×	✓	✓
Monitoring	×	×	×	✓	✓

As can be observed, CRISP-DM and Goosen's method are the two most similar methods. De Meyer's method also has all the steps found in both methods, except for the problem formulation step. Additionally, De Meyer's method included a monitoring step which entails continuously monitoring data integrity post implementation. Both Goosen's and De Meyer's methods were applied in the deep-level mining industry, whereas the rest of the methods only provide generic guidelines for any DA process. However, even both Goosen's and De Meyer's methods were mainly focused on building methods of detecting erroneous data. However, in this study, in addition to analysing and addressing data quality challenges, the clean data is used to quantify pump efficiencies. Therefore, the end product of this study is different from these other two studies. However, all the above methods provided a basis from which the methodology to be discussed in the next section was developed. The methodology was composed of a combination of the steps listed by the methods.

2.3 Methodology overview

The developed methodology is as detailed below. This methodology combines steps from the common popular data analysis techniques described in Section 2.2.

Step 1: Case study identification

The process of quantifying efficiencies relies on data availability and quality, as discussed in Chapter 1. Some mines have very limited instrumentation and thus might not have the necessary data required to calculate efficiencies. The purpose of this step is to identify mines with adequate and reliable data, from which case studies can be selected to validate the methodology.

Step 2: Data understanding and preparation

In this step the required data are gathered and visualised. Visualisation refers to the graphical representation of data through visual elements such as tables, graphs and charts. Data visualisation provides a way to see erroneous data and trends and understand patterns in data. This step was included in CRISP-DM, KDD, Goosen's method and BDA process. Data will also be cleaned and structured into a form that can be readily used. This step was included in all the identified DA methods.

Step 3: Development and automation

This step consists of implementation, whereby efficiencies are quantified using the clean data as well as validation of results. CRISP-DM, KDD and BDA included this step.

Step 4: Reporting

Reporting entails developing a method through which the efficiencies can be relayed to relevant end users. This step was found in CRISP-DM, Goosen's approach and the BDA process.

Figure 8 shows a visual representation of the methodology as discussed above.

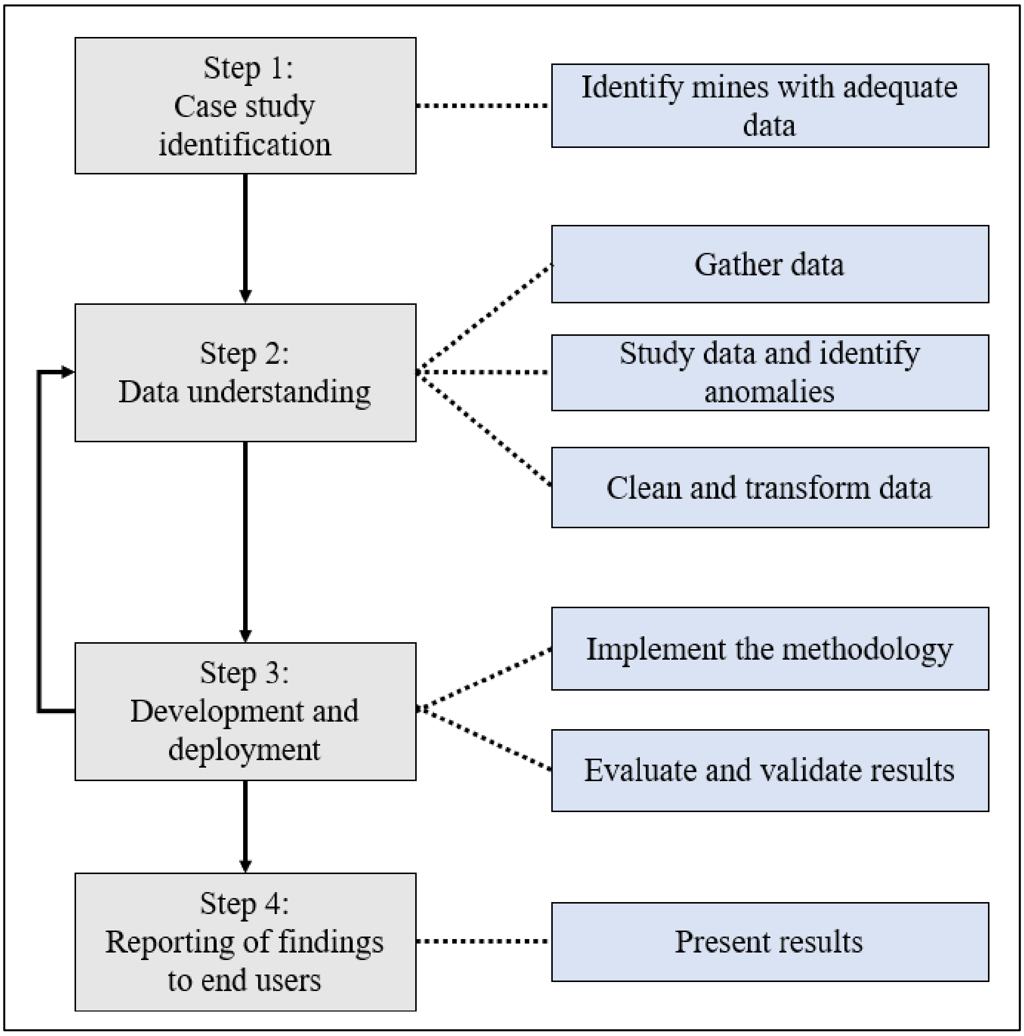


Figure 8: Methodology flow diagram

2.4 Step 1: Case study identification

The first step in the outlined methodology is to identify pumping systems (case studies) on which the methodology can be applied. To identify potential candidate systems, the following requirements or criteria are applied:

- The pumping system should have at least one pump, pumping water from one dewatering dam to another.
- Historic data for relevant parameters must be present or there must be a means of obtaining new data.

After potential case studies have been successfully identified, data are collected. The data should be able to explain the following:

- How many pumping stations are present?
- How many functional pumps are located on each pumping station?
- Are there any pumps that share a water discharge column?
- For pumps that share a common column, are their running statuses available?
- For pumps without power consumption readings, are there electric current readings from which power can be determined through Eq.3?

A case study that fulfils all the above requirements is selected and investigated further.

2.5 Step 2: Data understanding

The first objective of this study (Section 1.5) is to address the data challenges found in measurements required to quantify efficiency. In this step, the quality of data is assessed through data visualisation techniques. The data visualisation is done using Reporting Toolbox (RTB). RTB is a cloud-based reporting software that is available to the author. RTB was chosen as it makes it easier to set up reports and visuals. RTB has easy access to the mines' SCADA data. This simplifies the process of collecting and gathering data as that is done automatically. However, the same tasks can be completed using any other data analysis tool such as Excel.

The data are investigated with regards to the following:

- Data quality - this is supported by Goosen's study, which revealed that the following characteristics of erroneous data are prevalent in the mining industry [51]:
    - Missing data.
    - Static data.
    - Outliers.
- Start-up and shutdown conditions - equipment start-up and shutdown periods can often be characterised by inconsistent measurements [32], [33].
- Data resolution – aggregated/summarised data can lead to problems which will be discussed later in this section.

Data can contain one or any combination of these characteristics. Data visualisation is used to identify these errors with respect to each characteristic. A sample dataset is extracted from each case study from which the DA process is conducted.

2.5.1 Data quality: missing data

According to Kristin [84], failure to address missing data in the early stages of any DA process can lead to errors. According to Watson [85], improper handling of missing data can cause biases which can lead to inaccurate conclusions. Therefore, missing data is the first characteristic to be investigated.

Missing data can be a result of faulty instrumentation or an error in the data collection or transmission process [85]. Missing data can either be ignorable or non-ignorable [86]. Ignorable missing data involves variables that are not central to the analysis problem [87]. As the name suggests, ignorable missing data do not significantly influence results and therefore can be neglected if they cannot be acquired. However, non-ignorable missing data involve variables that are central to the analysis problem and can negatively affect results if not properly handled [87].

To determine if missing data in variables should be considered ignorable or non-ignorable, a sensitivity analysis (SA) is conducted based on the pump efficiency equation. SA explains how results (in this study: pump efficiency) can be affected by a change in input variables [88], [89], [90]. Frey [91] conducted a study where he reviewed SA methods. Amongst the methods that were explored in the study was a graphical and mathematical method: nominal range sensitivity analysis (NRSA), which was found to be applicable to this study [91]. NRSA involves adjusting one variable at a time while the rest are kept at the base-case [91], [92]. The difference between the base output and the output due to the variable input is the sensitivity of the model [91], [92].

To conduct a sensitivity analysis, an example base case was defined based on typical values that can be expected. The base case in this study had a resulting efficiency of 81% as shown in Table 4.

Table 4: Values at the base case defined for this study

Suction pressure (kPa)	Discharge pressure (kPa)	Discharge flow (l/s)	Power (kW)	Efficiency (%) (calculated using Eq. 1)
50	7000	130	1120	81

To investigate the sensitivity for each parameter the following procedure is followed in this example:

1. The parameter was varied in steps of 10% from the base parameter value (for example 50, 45, 40, etc.) while the rest were kept constant. The variations were done from -100% to +100% variation from the initial value. -100% variation means that the parameter value is zero and can thus be removed from the equation to avoid causing mathematical errors, for example power is a denominator in the equation and therefore should not be zero. +100% means that the parameter is twice the base value.
2. Output responses were calculated for each case.
3. The partial difference was calculated for each case.

The following figures present the results obtained for all step changes for each parameter.

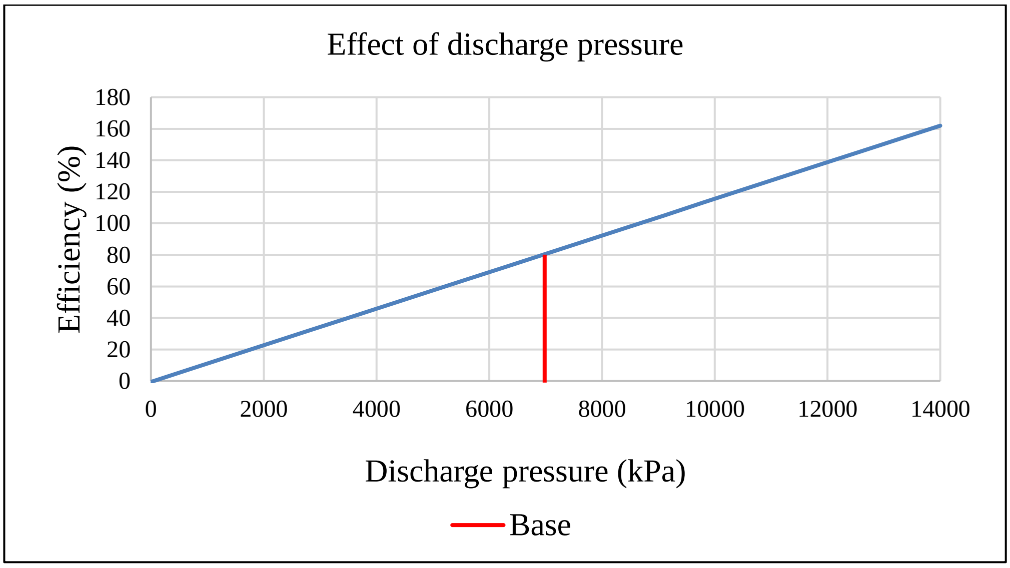


Figure 9: Efficiency differences caused by change in discharge pressure

As can be observed from Figure 9, changes in discharge pressure cause significant differences in efficiency. A pressure difference of -100% (0 kPa) from the base value yielded an efficiency of about 0%. This implies that, in the absence of discharge pressure, it will not be possible to calculate efficiency. It can also be seen in Figure 9 that a pressure difference of +100% yielded almost double the base efficiency. This relationship implies that efficiency is highly sensitive to discharge pressure. Therefore, discharge pressure should be treated as a non-ignorable parameter in this study.

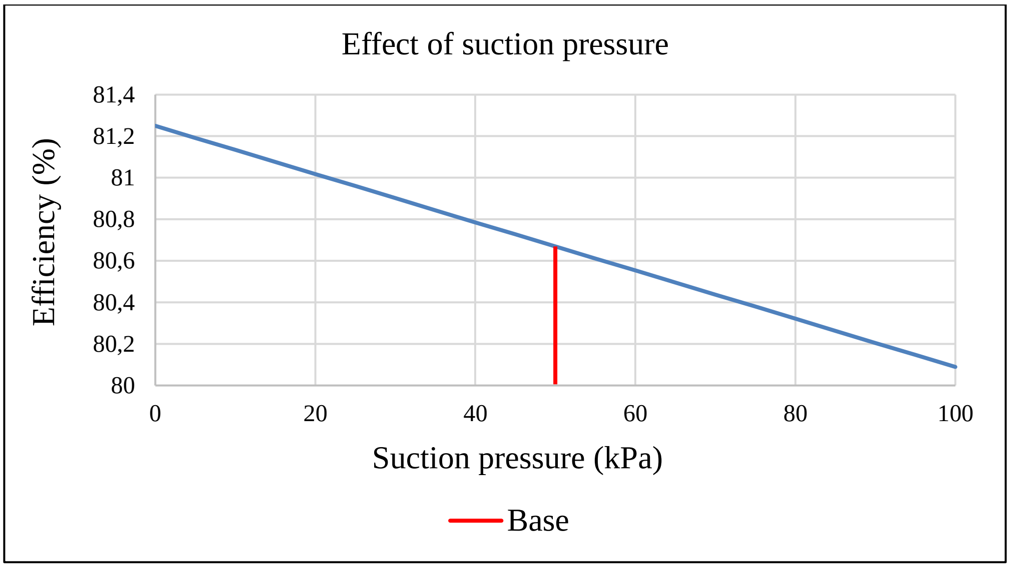


Figure 10: Efficiency differences caused by change in suction pressure

Figure 10 shows the impact caused by suction pressure. It can be clearly seen that the differences in efficiencies for both -100% and +100% variations were within 1% of the base efficiency. This is mainly because the base suction pressure is much smaller in comparison to discharge pressure, and thus its step changes do not cause a significant impact. As mentioned, the suction pressure value at the base case is typical for dewatering pumps. Therefore, these results are valid. These results showed that suction pressure carries negligible sensitivity and thus can be treated as an ignorable parameter in this study.

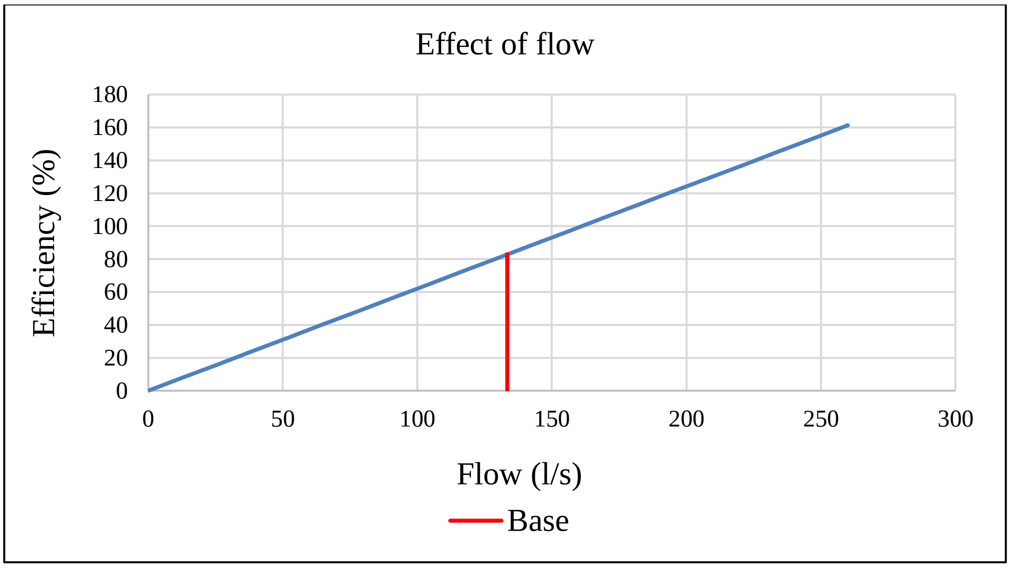


Figure 11: Efficiency differences caused by change in flow

Figure 11 resembles the pattern observed in Figure 9. This implies that flow has the same impact as discharge pressure. Therefore, flow is significant and can be treated as a non-ignorable parameter in this study.

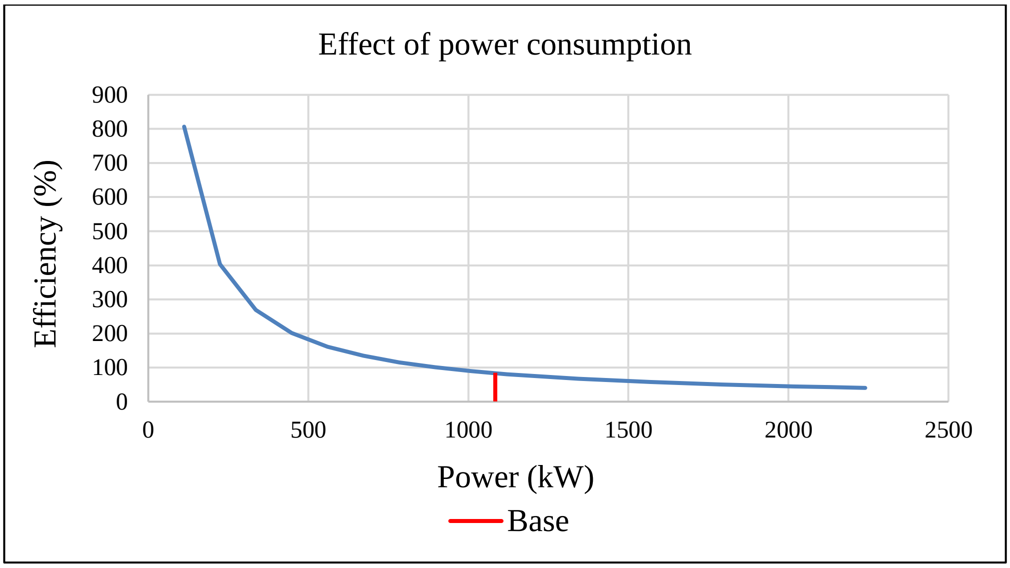


Figure 12: Efficiency differences caused by change in power consumption

From Figure 12 it can be observed that power consumption resembles a hyperbolic relationship with efficiency. At -100% difference (0kW), the solution to the equation is undefined/infinite as power is the denominator. At +100% difference, the efficiency is approaching 0%. This implies that power consumption is significant and thus should be treated as a non-ignorable parameter in this study.

From the above, the non-ignorable parameters were found to be discharge pressure, flow, and power consumption. Suction pressure was found to be the only ignorable parameter. This implies that in cases where suction pressure is missing, it can be removed from the data without causing any significant error.

Possible remedies for non-ignorable missing data include deleting all data records at affected time stamps or predicting missing values through various techniques [87]. SA is used to categorise parameters into ignorable or non-ignorable. Ignorable parameters are ignored in cases where they are missing. Non-ignorable parameters lead to efficiencies not being calculated at affected timestamps.

Data can have either one of the following missing data patterns [93], [94]:

- Partially missing - one or more parameters missing periodically or at random time stamps. This can be a result of a fault in the data transmission process, power failure or temporary instrumental fault.

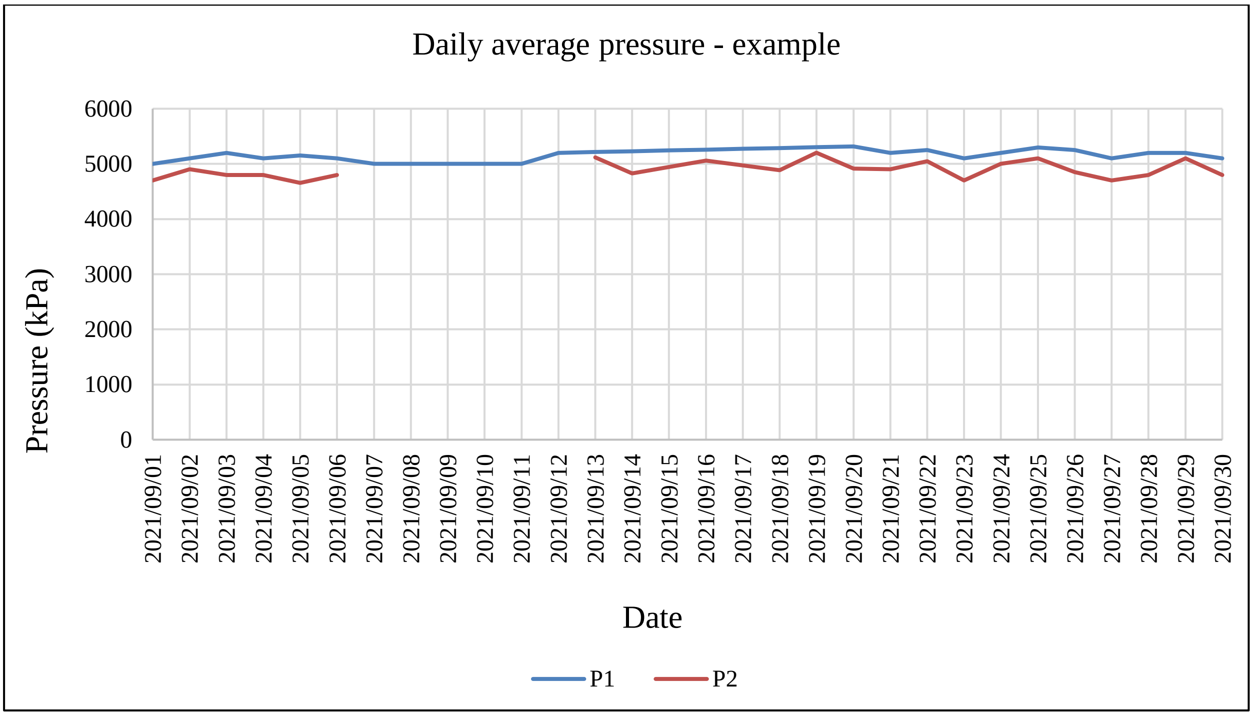


Figure 13: Example of partially missing data

As can be observed in the example on Figure 13, the pressure values for pump P2 were missing between the 7th and the 12th. This is regarded as partially missing data if the reporting period is for the entire month. This type of missing data is managed through a data monitoring report that is implemented using RTB. The purpose of the report is to detect missing data as it occurs and issue alerts to relevant parties, so that action can be taken to resolve the issue.

- Completely missing - one or more parameters missing at all time stamps. This can be a result of damaged instrumentation or a lack of instrumentation. Figure 14 presents an example of such a case.

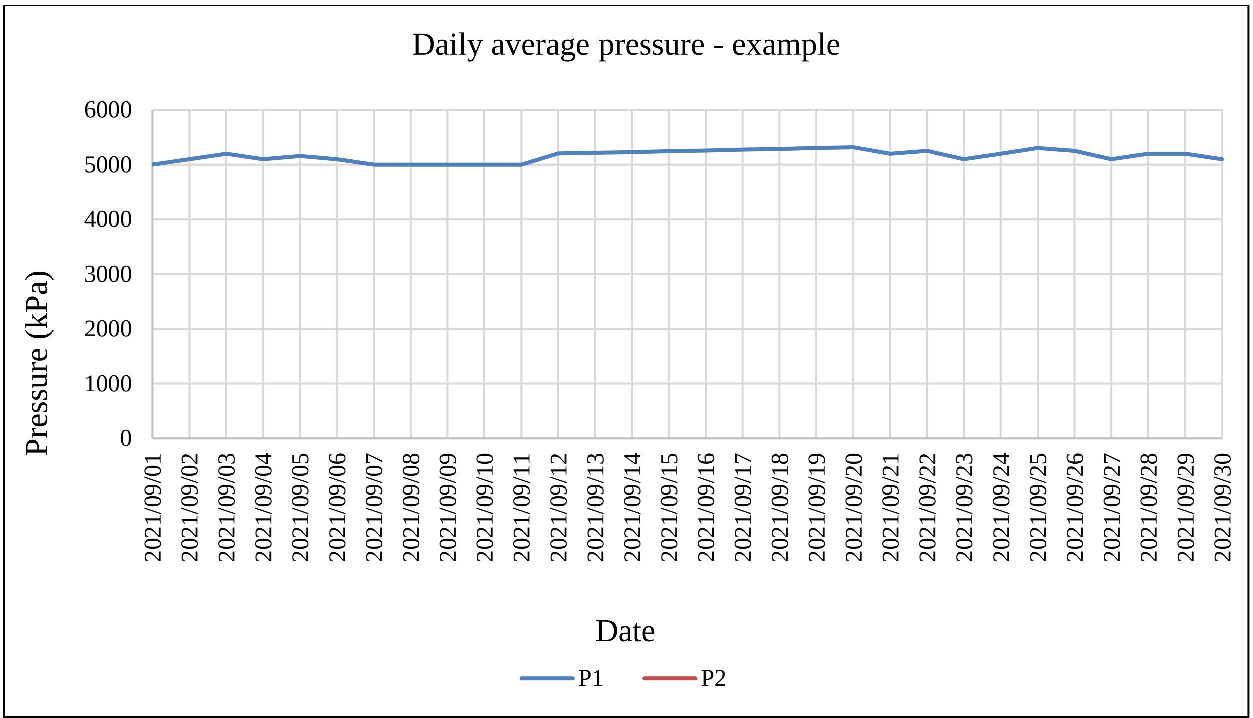


Figure 14: Example of completely missing data

The example in Figure 14 shows completely missing data for pump P2 in the reporting period. This kind of missing data can be identified in the initial data preparation step as it is more likely to occur indefinitely. If the cause is found to be beyond the author's control (e.g., damaged sensors), then the measurements are completely removed and efficiencies for the pumps concerned are not quantified.

2.5.2 Data quality: static data

Goosen's research revealed that static data are the biggest cause of erroneous data in mines [51]. Static data can occur in various forms [51], [52]:

- Values remain constant throughout – this can apply to all parameters involved in this study:
  - Running status – if it constantly remains 1 (i.e., "running"), it is regarded as static data because it is not possible for a pump to run indefinitely. The status will be compared with the rest of the parameters to verify that they correlate.
  - All other parameters – when a parameter remains constant regardless of a pump's running status, as demonstrated in the example shown in Figure 15. This type of static data is investigated by calculating the standard deviation for each parameter according to Eq.4. A standard deviation of zero for any parameter can only be accepted if the running status for that pump is 0 across the entire dataset.

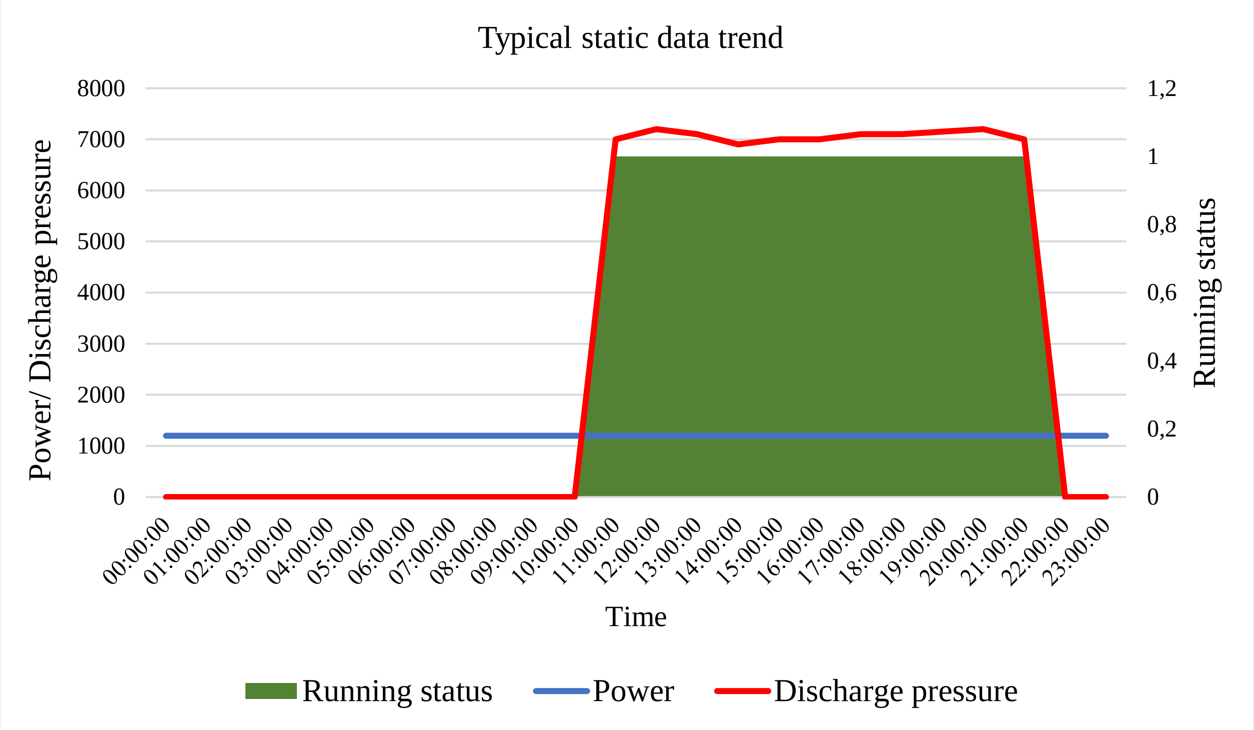


Figure 15: Example static data trend

Figure 15 shows an example of instances where power consumption values remain static although the running status and discharge pressure is changing. Such a scenario would suggest that the power reading is faulty as it is expected to change as a pump switches on and off.

- Values change with running status but remain static thereafter. This type of static data is investigated by considering data points that correspond with running statuses of 1.

As in the case of missing data, static data can also occur partially (for a finite period) or completely (indefinitely). Completely static parameters are discarded and parameter ignorability dictates whether efficiency can be calculated.

Standard deviation can be calculated to determine if there is any form of static data. Standard deviation measures how spread out a set of data points are. A large standard deviation indicates widely spread-out data, while a small standard deviation indicates that the data points are distributed closer to the mean [95]. Hence static data yield zero standard deviation. Standard deviation is calculated as follows [95]:

$$\boldsymbol{Standard\ deviation} = \sqrt{\frac{\sum(\boldsymbol{x}-\overline{\boldsymbol{x}})^2}{\boldsymbol{n}-\mathbf{1}}} \qquad \text{Eq.4}$$

where x represents a data point in the dataset, $\bar{x}$ is the mean, and n is the total number of data points in the dataset.

A data monitoring report is implemented to automatically detect static data in RTB. Parameters that are immediately flagged as completely static are discarded and parameter ignorability is applied. The data monitoring report then continues to be used in detecting partially static data.

The static data watchdog functionality in RTB is used to develop this report feature. The watchdog is based on the same principle of calculating standard deviation for data in a specified period. The feature consists of a threshold value which can be set by the user. The threshold is the number of data points that can be identical within a given period before the data is flagged as static. The challenge with the data presented in this case is that it can remain static for as long as a pump is off and that might raise false alarms. Therefore, the watchdog is only activated when the running status for a pump is 1. This is achieved by applying a filter to the data that is fed to the watchdog. The filter is implemented through a calculation as detailed below:

$$\boldsymbol{New\ data_{pumpx}} = \begin{cases} \boldsymbol{Raw\ data_{pumpx}} & \boldsymbol{if\ \ Status_{pumpx} = 1} \\ \boldsymbol{Null} & \boldsymbol{Otherwise} \end{cases} \qquad \text{Eq.5}$$

This is added to the data monitoring report which is triggered when the watchdog conditions are met.

2.5.3 Data quality: outliers

An outlier is a data point that is inconsistent with the rest of the data in a dataset [96], [97]. Outliers can be caused by sensor noise, equipment degradation or faulty data transmission [98]. There is a substantial amount of research on various outlier detection techniques (e.g., Wang *et al*. [97], Liu *et al*. [98]). After thoroughly considering those studies, the distance-based method was found to be applicable to this study. This method calculates the distance between data points by calculating the mean of all values and flagging all data points that highly deviate from the mean as outliers. This feature is only applicable to cases where a pump is running (i.e., running status = 1), because there cannot be a flow when a pump is off. Eq.6 shows the condition to be applied to eliminate outliers by setting boundaries (v_{min} and v_{max}), where v represents the value being filtered.

$$\boldsymbol{v_{min} \leq v \leq v_{max}} \qquad \text{Eq.6}$$

v_{min} is set to zero for all parameters. This is because a method of reliably determining these values could not be identified. The method of determining v_{max} is justified by the fact that a pump must operate within certain safety conditions as specified by the manufacturer. Therefore, v_{max} is set to those safety specifications, which are obtained from the pump nameplate as well as from the manufacturer’s website.

As mentioned in Section 1.1.4, suction pressure is dependent on the level of the water inside the dam from which a pump must pump. Therefore, the boundaries for suction pressure will be determined differently. Since suction pressure is directly proportional to the dam level, maximum suction pressure will be realised when a dam is full. Therefore, a test dataset is extracted from which instances where the dams reached maximum levels are identified. Suction pressures recorded at those instances are set as maximum boundaries.

Like missing and static data, outliers can also be partial (occur over a finite period) or complete (occur indefinitely). The first intervention for all outliers is to try to address the source of the outlier. If the source cannot be identified or resolved, then the same action taken with missing data is applied, depending on the ignorability of the parameter involved. If the source can be identified, measures are put in place to prevent new outliers from occurring.

2.5.4 Pump start-up and shutdown conditions

Equipment start-up and shutdown periods can often be characterised by inconsistent measurements [32], [33]. For example, in the case of the start-up period, the rotational speed of a pump increases rapidly from a stationary position while the flow-rate and pressure increase relatively slowly [99], [100], [101]. In both cases, a pump can experience surging whereby the flow-rate oscillates periodically [99].

To create a reliable dataset, start-up and shutdown data points must be discarded as they present inconsistent data. Pump start-ups and shutdowns are identified through the running status, where a status change from 0 to 1 indicates a start-up while 1 to 0 indicates a shutdown condition. Figure 16 shows a graphical representation of these conditions. The pump started up at t3 and shut down at t8.

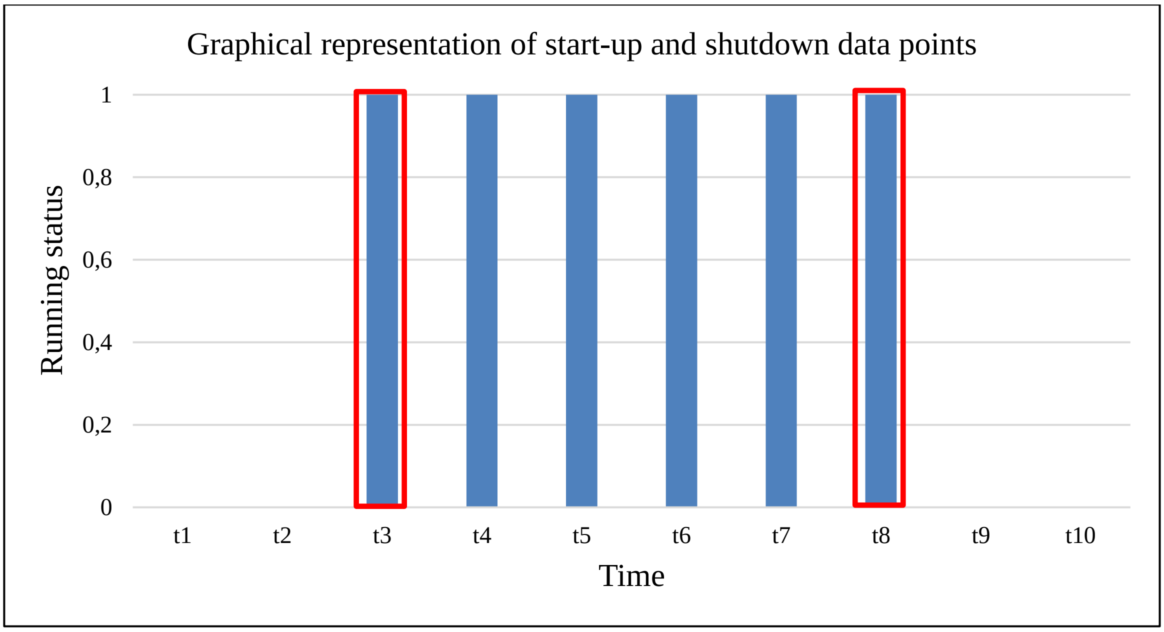


Figure 16: Graphical representation of start-up and shutdown data points

The highlighted bars in Figure 16 represent start-up and shutdown instances, respectively. Such data points are removed. Eq.7 shows how these data points are filtered out:

$$\boldsymbol{X(t_i)} = \begin{cases} \boldsymbol{Null} & \boldsymbol{if\ t_i \neq t_{i-1}} \\ \boldsymbol{X(t_i)} & \boldsymbol{otherwise} \end{cases} \qquad \text{Eq.7}$$

where: t_i represents a timestamp and $X(t_i)$ is the parameter value at a particular time stamp.

2.5.5 Data resolution

As mentioned in Section 1.2.2, data is logged through the SCADA from where it is sent to relational databases. Furthermore, it was mentioned that the frequency of logging is predefined depending on the application [47]. SCADA is responsible for aggregating real-time data according to the specified logging frequency [47]. SCADA data is then transmitted to a database through the process discussed in Section 1.2.1. Data can further be aggregated at the database level in two ways:

1. Data can be aggregated before being stored on the database to reduce volume.
2. Upon retrieving data from the database, further aggregation can occur to reduce the required computing power.

Although data aggregation helps in reducing data volumes as well as redundancy of raw data, it can negatively affect data accuracy [102], [103]. For example, if raw data consist of an outlier data point, the data point might get lost when aggregation takes place. Table 5 provides a hypothetical example where a single data spike affects the overall average of the dataset.

Table 5: Example of error caused by aggregation

Time	Value
t1	0
t2	0
t3	0
t4	10000
t5	0
t6	0
t7	0
t8	0
t9	0
t10	0
Average	**1000**

In the above example, if the average value is what is used for reporting, it might be interpreted as an outlier. However, if the data is used in its raw format, then it might be possible to identify the real outlier and avoid losing the rest of the data points. In the context of this study where missing data is prevalent, it is important to not waste any useful data points.

Mines typically install at least two pumps per discharge column. This implies that, if more than one pump is running per column, the flow that is recorded is a combined flow. To accurately allocate a discharge flow to a particular pump at a certain instance, that pump must be the only one running on that column at that instance. The following expression shows how this is achieved:

$$\eta_{pumpx}\,(\%) = \begin{cases} \eta & if\ (Status_{pumpx} = 1)\ and\ \left(\sum_{i=1}^{n} Status_{pumpi} = 1\right) \\ Null & Otherwise \end{cases} \qquad \text{Eq.8}$$

Interpretation: if the running status of $pumpx$ is 1 and the sum of running statuses for all the pumps (including $pumpx$) on the same column is 1, then calculate the efficiency of $pumpx$. η represents the efficiency (Eq. 1). This highlights the need for undiluted data, and Figure 17 shows why aggregated data can make this difficult to achieve:

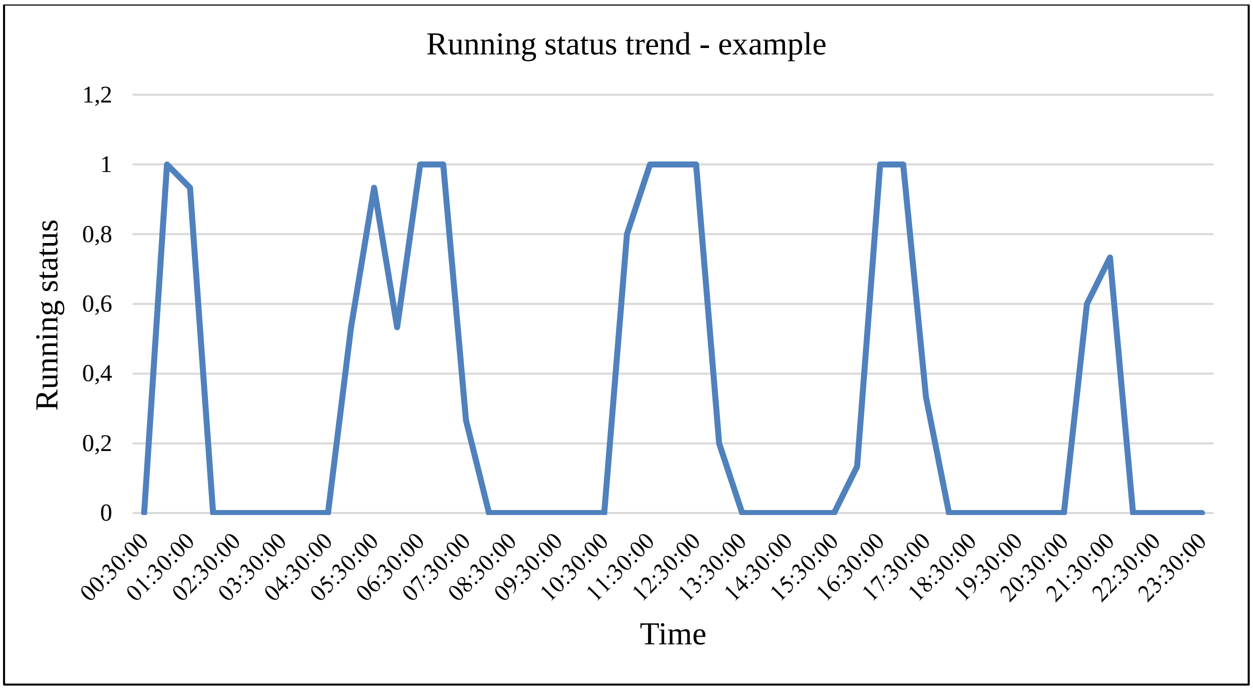


Figure 17: Running status vs discharge pressure with 10-minute resolution

Figure 17 is an example of data that was aggregated from 2-minute to 10-minute resolution. As can be observed, the running status at each interval can be any decimal between 0 and 1 as it consists of an average of on and off states (0's and 1's). To accurately execute the exercise of assigning flows for pumps that share a column, only running statuses of 1 must be used. This implies that some of the 1's that are diluted with 0's during aggregation will be lost. It is generally beneficial to have more data points as they can improve the accuracy of results [104]. The same is true in this study, where various data quality problems are prevalent. Therefore, it is important to use data in its most raw format to avoid any lost data points.

Due to the above reasons, the minimum data resolution used in this study is that which is logged by the SCADA on the mine. Therefore, no aggregation should happen at any point beyond the SCADA level.

2.6 Step 3: Development and automation

2.6.1 Implementation

After a case study has been identified and all required data have been prepared, efficiencies can be quantified. This step integrates all the data cleaning procedures developed in the previous section and quantifies the efficiencies. RTB is used to implement these processes. As mentioned in Section 2.5, RTB was chosen as it has functions that can be easily applied. However, any other data analysis tool such as Microsoft Excel can be used. In RTB, one can set calculations up so that they are computed per time stamp. The user does not have to worry about organising or sorting data, as RTB has that functionality built in. Thus, RTB was used to carry out all the data cleaning processes for this study, as well as the final automated report that will be discussed at the end of this subsection.

An RTB report can be automated and sent to specified recipients as shown in the following snippet:

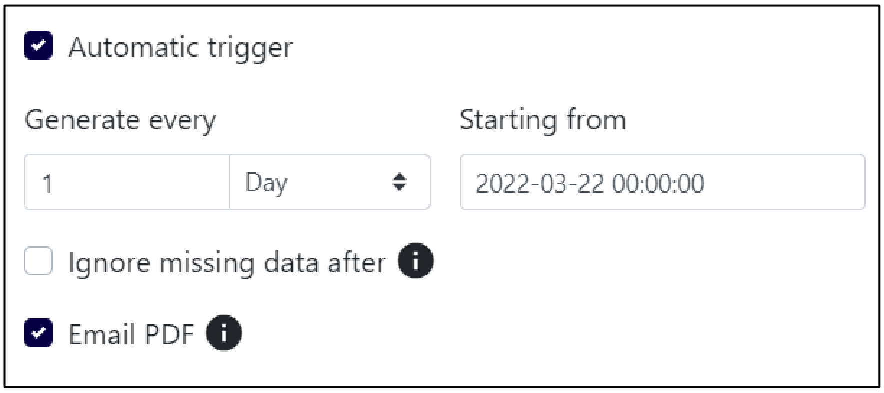


Figure 18: RTB automatic trigger function

A simple visual layout of how RTB is configured is shown in Figure 19.

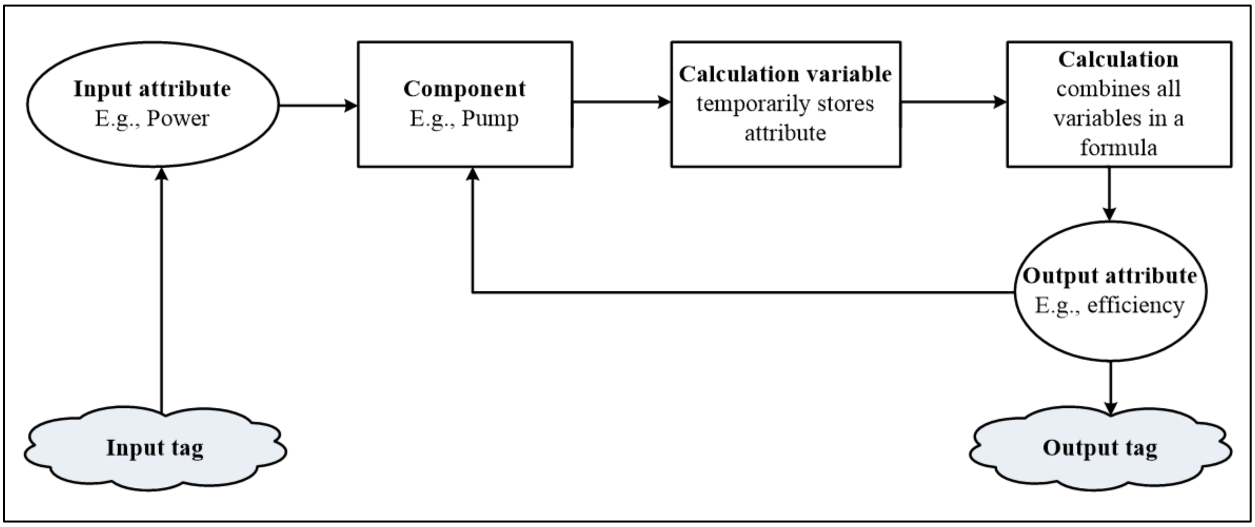


Figure 19: Layout of RTB calculation configurations

Each measurement is assigned to a tag which sits in the database. A tag is a unique identifier and is usually descriptive of the mine, equipment, parameter, and unit of measurement. For example, the tag name for the power consumption of pump 1 at Mine X can have the format: MineX_pump1_power_kW. Components are used as equipment identifiers. Each component consists of a list of attributes (parameters) e.g., power, discharge pressure, etc. Attributes are used to store tags and any other information specific to a parameter, such as limits, etc.

Calculation variables can also be created for each attribute. A variable 'tells' a calculation to which it is linked which component and attribute to use. One variable can be used by multiple calculations. Furthermore, a single calculation can have multiple variables as shown in the following snippet:

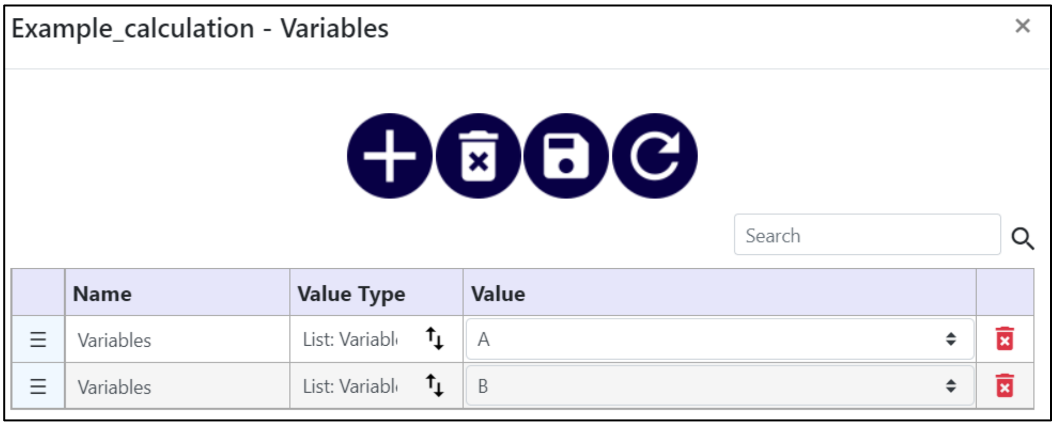
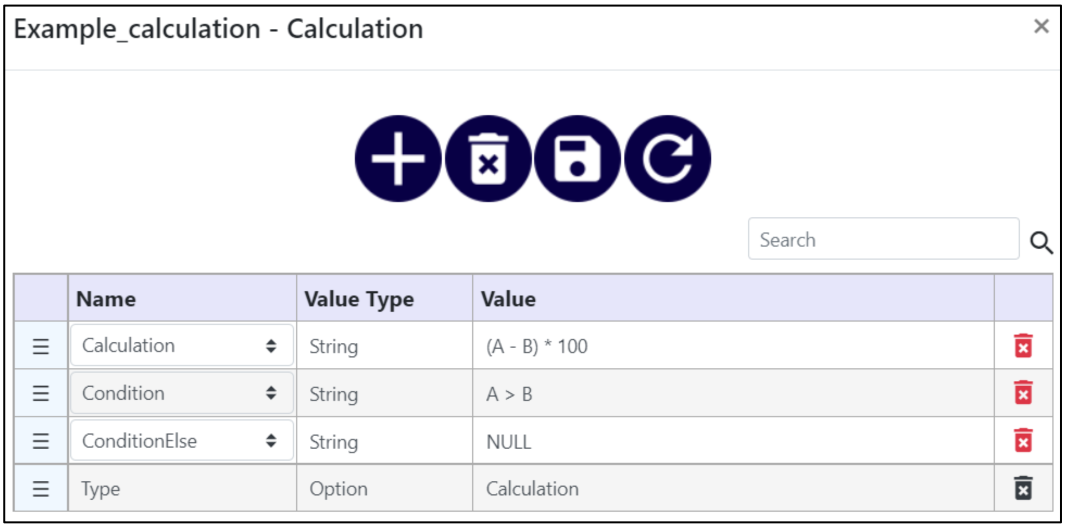


Figure 20: Example of variables linked to a calculation

Calculations can be defined using the linked variables. Calculations can also be accompanied by conditions, as shown in Figure 21.



	Name	Value Type	Value	
☰	Calculation	String	(A - B) * 100	
☰	Condition	String	A > B	
☰	ConditionElse	String	NULL	
☰	Type	Option	Calculation	

Figure 21: Example of an RTB calculation

The calculation in Figure 21 will only be done if the condition specified is satisfied, otherwise the result will be null. The condition feature will be used to automate all the data filtering conditions discussed in Section 2.5. For example, Eq.6 (outlier filtering) will be implemented as shown in the snippet on Figure 22.

	Name	Value Type	Value	
☰	Calculation	String	v	
☰	Condition	String	(v >= 0) AND (v <= vmax)	
☰	ConditionElse	String	NULL	
☰	Type	Option	Calculation	

Figure 22: Implementation of minimum and maximum boundaries in RTB

Calculated results are linked to an output attribute, which can be used as an input to another calculation, be displayed on a visual, or be used to export values to an output tag. Output tags are like input tags in how they are configured on the database. The only difference is that input tags get their data directly from SCADA and other external sources, while output tags are used to store calculation results. Output tags can also be used by other reports and systems that have access to the database.

Calculations are computed row by row in a Pandas DataFrame[3] according to the raw data resolution and hence calculation results generally have the same resolution as the raw data. To

[3] A Pandas DataFrame is a programmatic table that consists of labelled rows and columns

change the resolution of the results, a new data interval can be defined through a dataset that can be linked to the calculation. The dataset forces the output to be aggregated according to the interval specified. An example of a dataset is given in Figure 23.

	Name	Value Type	Value	
☰	Interval	Integer	1	
☰	IntervalType	Option	Month	
☰	Period	Integer	6	
☰	PeriodType	Option	Month	
☰	Type	Option	DataSet	

Figure 23: Example of a dataset in RTB

Dataset interpretation: monthly interval for a period of six months.

RTB also has a watchdog feature which is capable of monitoring missing and static data. A threshold can be used to specify how long missing or static data can occur until a flag is raised. Upon detecting one of those, the automatic trigger feature issues alerts. This feature is used to automate the data monitoring process.

RTB also allows the user to configure graphs, tables, and other data visualisation types which can be plotted in a PDF report. This feature is used to develop a PDF report containing all efficiencies and other important metrics to be sent to all relevant end users. As mentioned in Section 1.5, one of the objectives of this study is to automate the process of monitoring efficiencies. Therefore, the developed PDF report is automated using the automatic trigger feature.

2.6.2 Results and validation

The efficiencies obtained are presented, validated, and interpreted. Validation refers to assessment of the accuracy of the results obtained. This is an important exercise for any analysis process. Results can be validated by intuition [49] or through test data [83]. Intuitive validation in this study entails not accepting efficiency values that are negative or above 100%. This method helps identify obvious errors in results, in which case the data preparation stage is revisited. Validation through test data in this study entails manually gathering raw data through different measurement equipment and comparing the results produced by the manual data with the ones obtained through the developed process. In this case, the author went underground at the case study mines to collect data. The manually obtained data were compared with the data used by this study. The results were interpreted, and conclusions were drawn as to whether the results satisfy the study objectives.

2.7 Step 4: Reporting of findings to end users

Reporting refers to the process of creating and customising reports, as well as delivering these to relevant end users. The best method on how to effectively relay the information to end users was investigated. In this case the following mining personnel were consulted regarding the preferred and most effective reporting mechanism:

1. Instrumentation technician - responsible for overseeing all equipment and measurement instruments-related tasks. The instrumentation technician is the main point of reference for all tasks relating measurements [105].
2. Shaft engineer - responsible for planning mining operations, and designing and managing mining equipment, as well as supervising technicians and workers who operate the equipment[4].

These personnel were engaged through ad hoc meetings where various reporting options were presented to them. The best reporting mechanism, as well as the contents and structure of the report, was decided during these meetings. The available reporting mechanism is:

- Reporting documents – this can be PDF reports that are sent to end users periodically. In addition to pump efficiency, the report can also contain other important pump performance metrics as determined by all end users.
- A dashboard – refers to any data visualisation tool/page configured for the purpose of rapidly monitoring certain conditions [106]. One of the advantages about dashboards is that they can be automated and have the capability to self-update in real time[5]. A dashboard can be built with software such as Microsoft Power BI, Tableau Desktop, Google Data Studio, etc.

[4] "Mining Engineer Job Description, Career as a Mining Engineer, Salary, Employment." https://careers.stateuniversity.com/pages/69/Mining-Engineer.html (accessed Aug., 2021).
[5] "77 Open Source, Free and Top Dashboard Software." https://www.predictiveanalyticstoday.com/open-source-dashboard-software/ (accessed Nov., 2021).

2.8 Conclusion

This chapter presented the methodology through which the efficiencies of dewatering pumps were quantified for this study. The methodology consists of three major steps that were adopted from DA approaches found in literature.

Data visualisation plays a pivotal role in identifying and addressing data quality challenges. According to literature, the main data quality challenges faced in this study are missing data, static data, and outliers. Pump start-up and shutdown states, as well as aggregated input data, were also found to be another form of erroneous data. Filters that address these issues before calculating efficiencies were presented. The accuracies of the efficiencies were evaluated through a two-step process: validation by intuition and validation through test data. A report containing the efficiency values was developed and distributed to relevant end users.

CHAPTER 3

IMPLEMENTATION AND RESULTS

3.1 Preamble

The methodology developed in Chapter 2 was applied to two case study mines located in South Africa, hereafter to be referred to as Mine A and Mine B. A data monitoring system was implemented using the Reporting Toolbox software (RTB), which was discussed in detail in Section 2.6.1. RTB was also used to create automated pumping efficiency reports for both mines. Other software such as Excel can also be used. The reports were configured to generate monthly, although there is a functionality to generate a daily report when needed.

3.2 Step 1: Case study identification

Data availability dictated which mine(s) could be selected as potential case studies. To identify case studies, six mines were selected according to the criteria discussed in Section 2.4. A data availability audit was conducted on these mines. The audit was carried out in two steps:

- Identifying all existing pump instrumentation by conducting site visits and involving mine instrumentation technicians where necessary, and
- Verifying that the instruments are working, and that data are available on the SCADA.

From the six mines, only two mines were found to have enough measurement points to quantify efficiencies. The rest of the mines either had none of the required measurements or lacked measurements for parameters classified as non-ignorable parameters (Section 2.5.1).

The two mines with adequate data were selected as case studies. More information was gathered on the two mines to better understand the layouts of the dewatering systems. The information was obtained from instrumentation technicians, shaft engineers as well as doing underground visits where necessary.

Mine A

Mine A is a deep-level gold mine situated to the west of Johannesburg. The mine consists of five dewatering levels (L), namely, 115L, 100L, 75L, 52L and 29L. Table 6 summarises data availability audit results obtained for the mine.

Table 6: Data availability summary for Mine A

Pump		Discharge pressure	Suction pressure	Discharge flow	Power consumption	Running status
29L	P1	✓	✓	✓	✓	✓
	P2	✓	✓	✓	✓	✓
	P4	✓	✓	✓	✓	✓
	P5	✓	✓	✓	✓	✓
52L	P1	✓	✓	✓	✓	✓
	P2	✓	✓	✓	✓	✓
	P3	✓	✓	✓	✓	✓
	P4	✓	✓	✓	✓	✓
75L	P3	✓	✓	✓	✓	✓
	P4	✓	✓	✓	✓	✓
	P5	✓	✓	✓	✓	✓
	P6	✓	✓	✓	✓	✓
100L	P1	✓	✓	✓	✓	✓
	P3	✓	✓	✓	✓	✓
	P4	✓	✓	✓	✓	✓
	P5	✓	✓	✓	✓	✓
115L	P2	✓	✓	✗	✓	✓
	P3	✓	✓	✗	✓	✓
	P4	✓	✓	✗	✓	✓

Table 6 shows that most of the required data for efficiency calculations were available, although the 115L pumps were found to have dysfunctional discharge flow meters (to be addressed in Section 3.3.1). A visual layout of the system is shown in Figure 24.

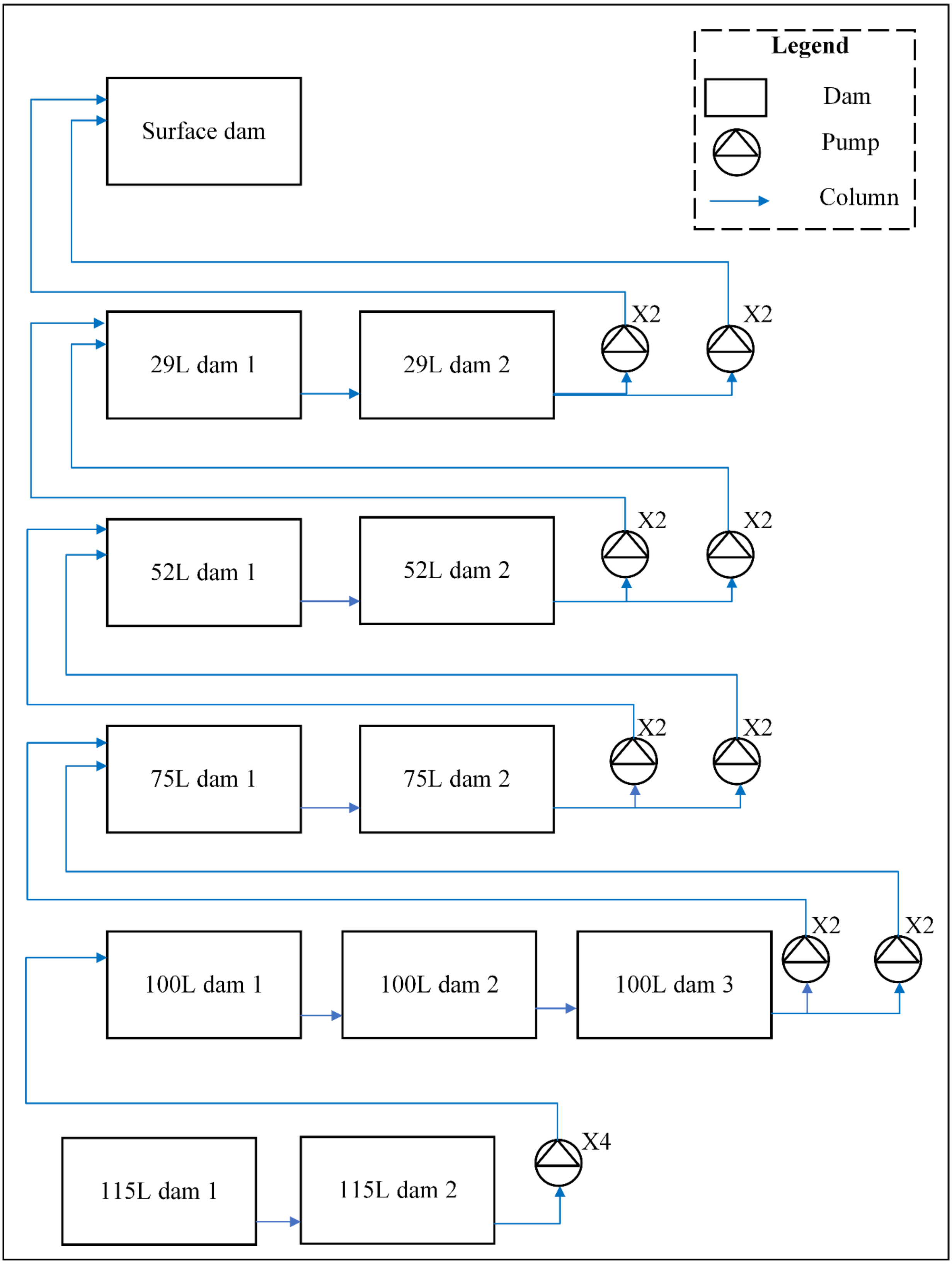


Figure 24: Simplified layout of Mine A's dewatering system

The pumping stations are set up as described below:

- There are four pumps connected in parallel on the 115L pumping station. The pumps are responsible for pumping water from the two dams on 115L to the 100L dams through a single column.
- On the 100L pumping station there are five pumps, although one has been decommissioned (pump P2). The station consists of two columns leading to the 75L dams. One column is shared by pumps P1 and P3 connected in parallel, and the other one by P4 and P5, also connected in parallel.
- The 75L pumping station consists of six pumps and two columns. However, two pumps have been decommissioned (P1 and P2). The remaining four pumps are responsible for pumping water from the 75L dams to 52L dams. One column is shared by pumps P3 and P5 and the other one by P4 and P6. Both sets of pumps are connected in parallel.
- The 52L pumping station consists of five pumps, but one has been decommissioned (P5). The four remaining pumps deliver water from 52L dams to 29L dams using two columns. Pumps P1 and P3 share one column, while P2 and P4 share the other. Both sets of pumps are connected in parallel.
- On the 29L pumping station there are five pumps, of which one has been decommissioned (P3). There are two columns leading to the surface dam. Pumps P1 and P4 share one column, while P2 and P5 share the other. Both sets of pumps are connected in parallel.

Information about each pump's manufacturer and type was obtained from its nameplate. The information was used to obtain pump specifications from the manufacturer's website. Table 7 shows a summary of the pump information obtained [107].

Table 7: Mine A's dewatering pump details

Pumping station	Pump type	Power (kW)	Discharge pressure (kPa)	Flow rate (l/s)
115L	HPH 54-25 7	2600	8000	175
100L	HPH 58-25 8	2600	9020	225
75L	HPH 50-20 9	1300	7700	120
52L	HPH 50-20 9	1300	7700	120
29L	HPH 50-20 9	1300	7700	120

Mine B

This is a deep-level gold mine located near Klerksdorp in the North-West province. It consists of three dewatering levels, namely 76L, 78L, 61L and 3000L. Table 8 presents the data availability summary obtained for this mine.

Table 8: Data availability summary for Mine B

Pump		Discharge pressure	Suction pressure	Discharge flow	Power consumption/ electric current	Running status
76L	P1	✓	✗	✓	✓	✓
	P2	✓	✗	✓	✓	✓
	P3	✓	✗	✓	✓	✓
78L	P1	✓	✗	✓	✓	✓
	P2	✓	✗	✓	✓	✓
	P3	✓	✗	✓	✓	✓
61L	P1	✓	✓	✓	✓	✓
	P2	✓	✓	✓	✓	✓
3000L	P1	✓	✓	✓	✓	✓
	P2	✓	✓	✓	✓	✓

As can be observed in Table 8, neither the 76L nor the 78L pumps are equipped with suction pressure meters (to be discussed in Section 3.3.1). Figure 25 shows a simplified layout of the mine's dewatering system.

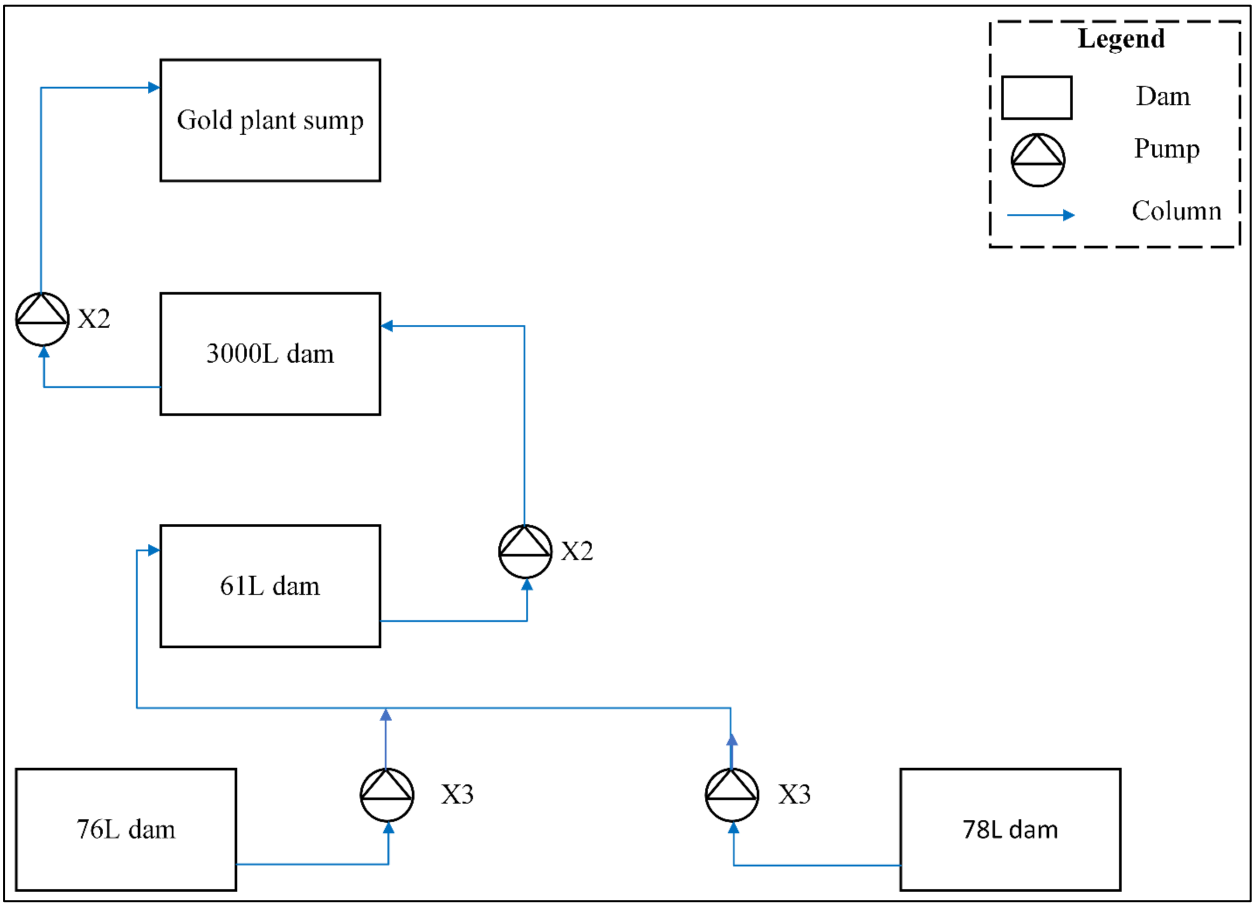


Figure 25: Simplified layout of Mine B's dewatering system

The pumping stations are set up as described below:

- The 78L pumping station consists of six pumps. Three of the pumps (P1, P2 and P3 – to be referred to as 78L pumps henceforth) are connected to the 78L dam, whereas the other three (P4, P5 and P6 – to be referred to as 76L pumps henceforth) are connected to the 76L dam. All six pumps are connected to one column which leads to the 61L dam. There is a pump control mechanism that is designed such that the 78L pumps and the 76L pumps cannot operate at the same time.
- There are two pumps connected in parallel on 61L. The pumps are responsible for pumping water from the 61L dam to the 3000L dam through a single column.
- The 3000L pumping station also consists of two pumps connected in parallel. These pumps are responsible for pumping water from the 3000L dam to a gold plant sump on surface through a single column.

Table 9 shows a summary of the pump information as obtained from the manufacturer's website as in the case of Mine A.

Table 9: Mine B's dewatering pump details

Pumping station	Pump type	Motor power rating (kW)	Discharge pressure (kPa)	Flow rate (l/s)
76L	HPH 32-17 5	1100	6300	125
78L	HPH 32-17 5	1100	6300	125
61L	HPH 54-25 10	2800	10000	175
3000L	HPH 54-25 10	2800	10000	175

3.3 Step 2: Data understanding

Data was gathered for case study mines. Both mines log data with 2-minute resolution. The gathered data was for a period of a month and consisted of 21600 data points for each parameter for each mine. This was done as a preliminary data understanding process, after which the methodology was applied on a continuous basis. The data visualisations for the two mines were done separately, but both followed the process discussed in Chapter 2.

3.3.1 Missing data

As mentioned in Section 2.5.1, sensitivity analysis (SA) provided a basis from which the ignorability of parameters was evaluated. Suction pressure was found to be ignorable. Therefore, as mentioned Section 2.5.1, in cases where suction pressure was found to be missing, efficiencies were calculated in its absence.

3.3.1.1 Missing data due to lack of instrumentation (completely missing data)

Mine A

As shown in Table 8, it was discovered that there is no functional instrumentation for the 115L column discharge flow. Discharge flow is non-ignorable and hence 115L pumps were not included in the efficiency evaluations. The rest of the pumps had all the measurements required.

Mine B

It was discovered that none of the pumps at the mine have power meters. However, the pumps are equipped with electric current meters and hence Eq.3 was used to calculate the power. For consistency, the upcoming sub-sections will only refer to calculated power values as raw data.

As discussed in Section 3.2, it was discovered that the 76L and 78L pumps do not have suction pressure meters. Therefore, the efficiencies for the 76L and 78L pumps will be quantified in the absence of suction pressure since the sensitivity analysis in Section 2.5.1 revealed that suction pressure is an insignificant parameter.

3.3.1.2 Missing data due to faulty data transmission (partially missing data)

This investigation was conducted on the datasets for both mines to understand the typical missing data patterns to implement measures to address the causes. Missing data gaps were found in the datasets. However, the gaps occurred at the same time stamps for all parameters and were found to have been caused by faults during data transmission e.g., server disconnecting from SCADA. This was found to be the case for both mines.

To combat these issues, a watchdog was set up in RTB to monitor missing data. The watchdog was used to build a report that serves as an alert. The report was set to automatically trigger every 10 minutes, provided there are missing data. The cause of the missing data can then be investigated, and action can be taken to resolve it. Graphs were also setup to show a two-minute profile for an hour for all tags. In the case that the report is triggered, one can simply view the graphs to see the instances during which the data loss occurred. The purpose of this exercise was to reduce the time it takes to detect and troubleshoot data loss. The report went live at the beginning of November 2021.

The data monitoring report reduced the time it took to address data loss, which ensures that there is adequate data available for more frequent and accurate efficiency reporting.

The overall data loss figures reduced since the implementation of the report. Figure 26 and Figure 27 show the overall data loss trends over a seven-month period for Mine A and Mine B, respectively. This highlights the impact that the intervention had on missing data recovery. As can be observed from Figure 26 and Figure 27, the number of missing data points started declining after implementation of the data monitoring report.

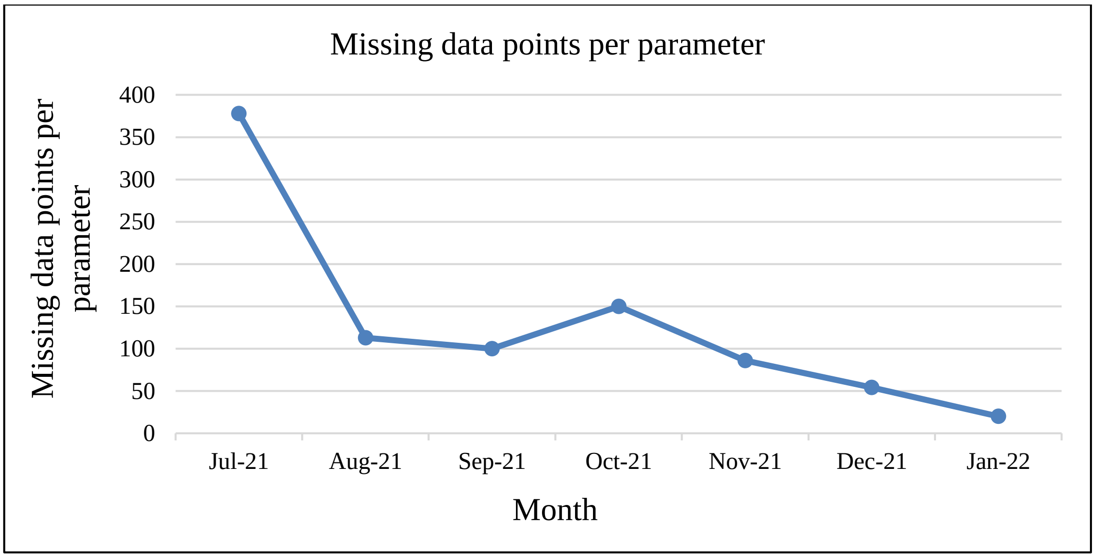


Figure 26: Missing data trends for Mine A

As can observed in Figure 26, the month of October had a higher missing data point count for Mine A. Most of the missing data occurred on a single day and was found to have been caused by a database configuration issue. The issue occurred on a Sunday and thus was only attended to the following day as database maintenance personnel were not on duty.

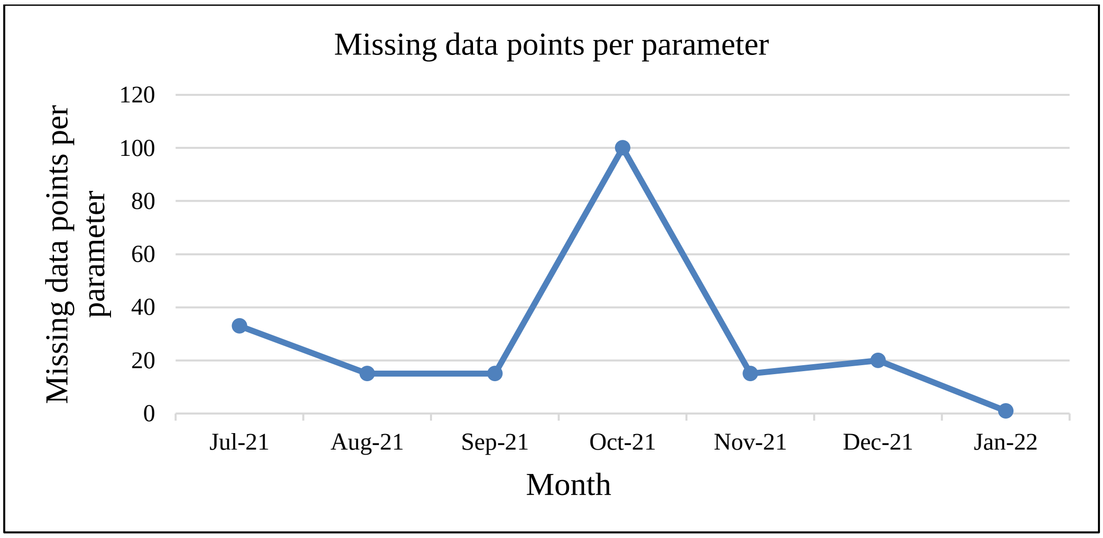


Figure 27: Missing data trends for Mine B

Like Mine A, Figure 27 shows that Mine B also experienced a higher missing data point count in October. The cause of the missing data was found to be similar to that of Mine A and was addressed similarly.

The automated data monitoring report can be used to avoid such issues in future as warnings will be issued at any time on any day.

3.3.2 Static data

As mentioned in Section 2.5.2, static data can occur partially (occur randomly for a finite period) or completely (occur indefinitely). This step involved activating a static data watchdog functionality in the data monitoring report, which was used to fulfil two tasks:

- Flagging completely static data – which were then discarded considering parameter ignorability.
- Automating the report so as continue monitoring partially static data.

3.3.2.1 Completely static data

Mine A

In the case of Mine A, there were no completely static data points in the dataset.

Mine B

At Mine B it was found that the electric current reading for 3000L pump P2, as well the 78L column flow reading had standard deviations of 0 throughout the dataset. Figure 28 highlights an example of instances where 78L pumps P1 and P3 were running, but the column flow remained 0 l/s. This was found to be the case across the entire dataset.

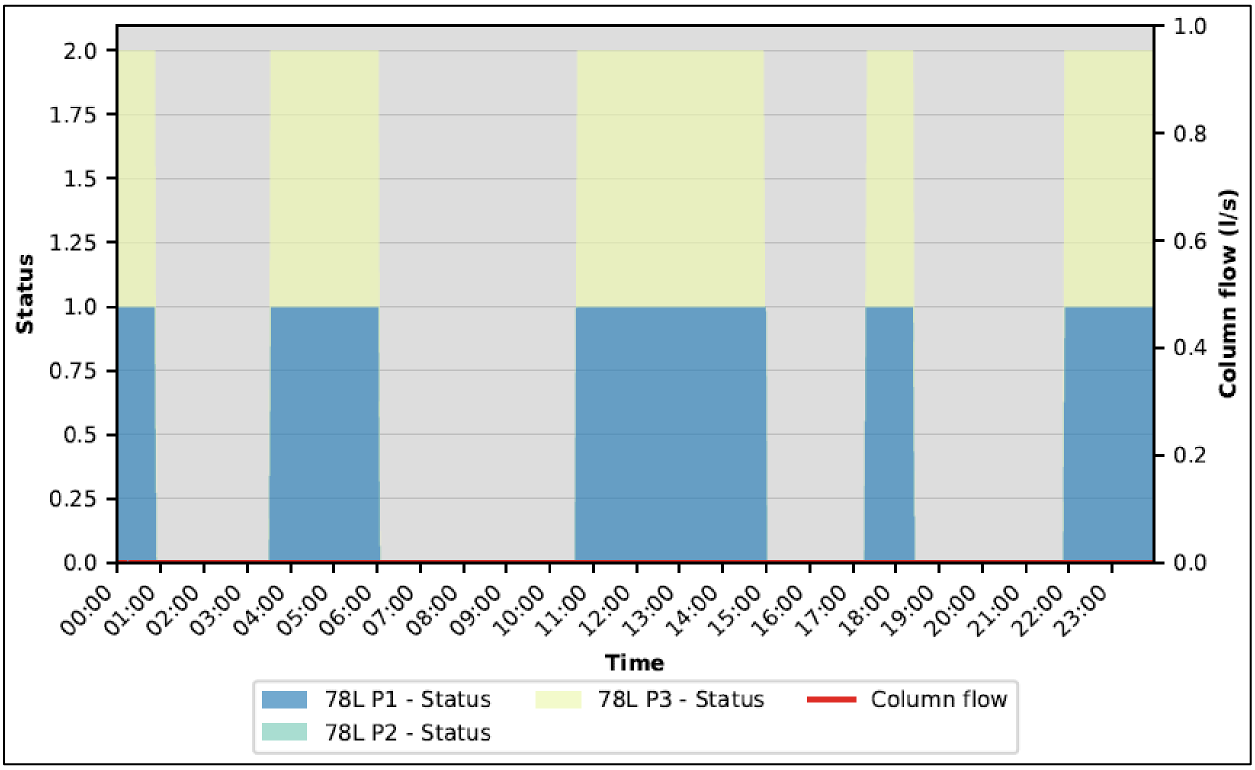


Figure 28: Comparison of 78L running statuses vs column flow

Further investigations into these revealed that all the aforementioned issues were caused by faulty instrumentation. This was confirmed by the mine's instrumentation technician. Therefore, it was not possible to calculate efficiencies for the affected pumps for now as discharge flow was identified as a non-ignorable parameter.

3.3.2.2 Partially static data

Unlike completely static data, partially static data do not necessitate complete removal of the parameters concerned. Partially static data can occur randomly and can be addressed as it occurs. Since implementation of the report there has not been a single case of partially static data for either mine.

3.3.3 Outliers

As mentioned in **S**ection 2.5.3, maximum boundaries for discharge pressure, flow and power were defined based on each piece of equipment's specifications as provided by the manufacturer. Suction pressure boundaries were determined based on maximum dam levels. These boundaries were only applied to cases where the running status is 1. Table 10 summarises the maximum boundaries that were set at Mine A.

Table 10: Maximum values set for parameters at Mine A

Pump	Suction pressure (kPa)	Discharge pressure (kPa)	Discharge flow (l/s)	Power (kW)
29L P1	67	7700	120	1300
29L P2	67	7700	120	1300
29L P4	70	7700	120	1300
29L P5	67	7700	120	1300
52L P1	56	7700	120	1300
52L P2	61	7700	120	1300
52L P3	58	7700	120	1300
52L P4	56	7700	120	1300
75L P3	183	7700	120	1300
75L P4	203	7700	120	1300
75L P5	181	7700	120	1300
75L P6	183	7700	120	1300
100L P1	555	9020	225	2600
100L P3	548	9020	225	2600
100L P4	546	9020	225	2600
100L P5	542	9020	225	2600

Table 10 shows that the suction pressures for pumps on the same level were similar, as expected (according to the definition of suction pressure provided in Section 1.1.4). The suction pressures on 100L were higher as compared to the rest of the pumps. The values are also higher than the value that was used for the SA (50 kPa). Appendix A shows that the SA still holds, even for such high suction pressure values.

Table 11 shows the maximum boundaries for pumps at Mine B.

Table 11: Maximum values set for parameters at mine B

Pump	Suction pressure (kPa)	Discharge pressure (kPa)	Discharge flow (l/s)	Power (kW)
78L P4	-	6300	125	1100
78L P5	-	6300	125	1100
78L P6	-	6300	125	1100
61L P1	74	10000	175	2800
61L P2	74	10000	175	2800
3000L P1	76	10000	175	2800

As can be observed from Table 11, the suction pressures for pumps on the same level were also similar at Mine B.

It can also be observed that, for both mines, parameters such as discharge pressure, discharge flow and power for pumps on the same level are identical, as so are the pumps.

Figure 29 shows an example of where the maximum boundary condition was practically applied.

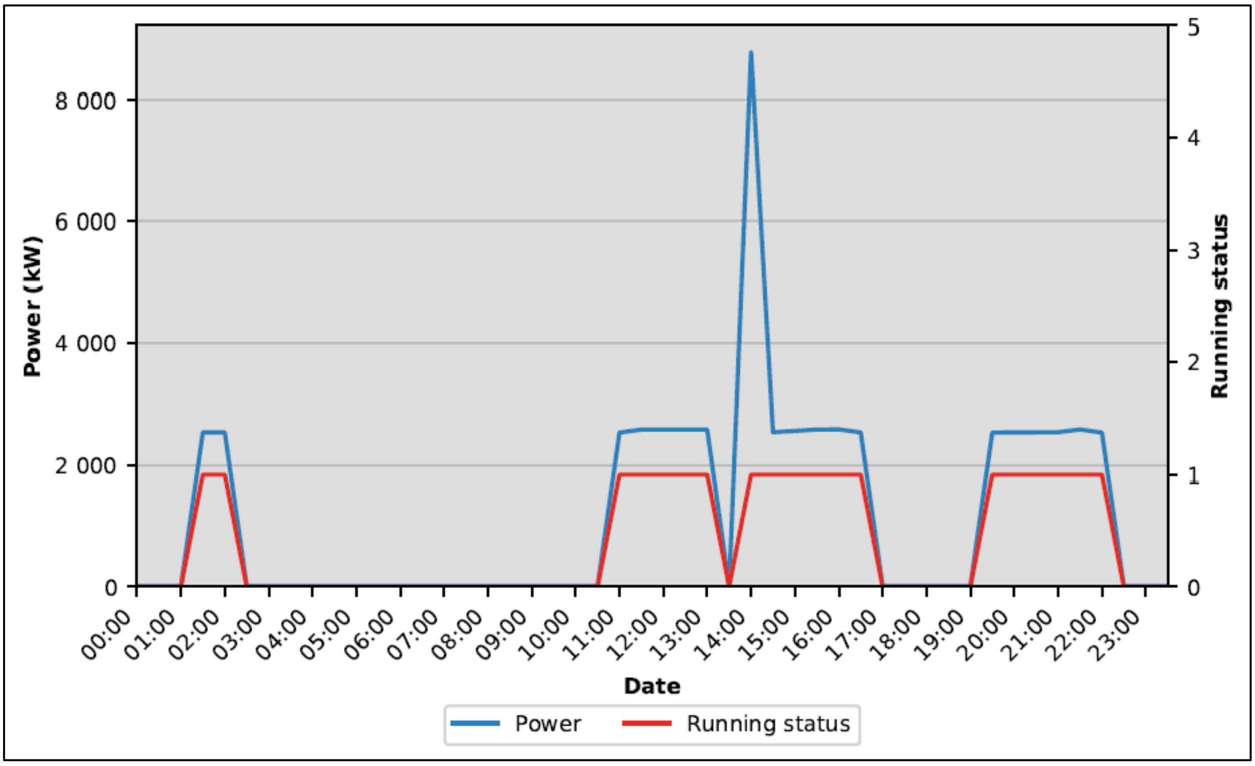


Figure 29: Example of outlier data points on 61L pump P2

The values reflected in Figure 29 are of the 61L pump P2's (Mine B) hourly maximum power consumption for a 24-hour period. As can be observed on the figure, the value realised at 14:00 was about 9000 kW, whereas the rest of the values that correspond to running statuses of 1 were averaging approximately 2500 kW, which is in line with the pump's rated power of 2800 kW as detailed in Table 9. The cause of this could not be linked to an issue in the data transmission process as it only happened to one parameter. Therefore, it can be attributed to an instrumental error.

3.3.4 Start-up and shutdown conditions

As mentioned in Chapter 2, this study focused on quantifying the efficiency when a pump is running at steady state. Hence, start-up and shutdown cases were removed from the data. Figure 30 shows an example of the visualisation that took place to identify those cases.

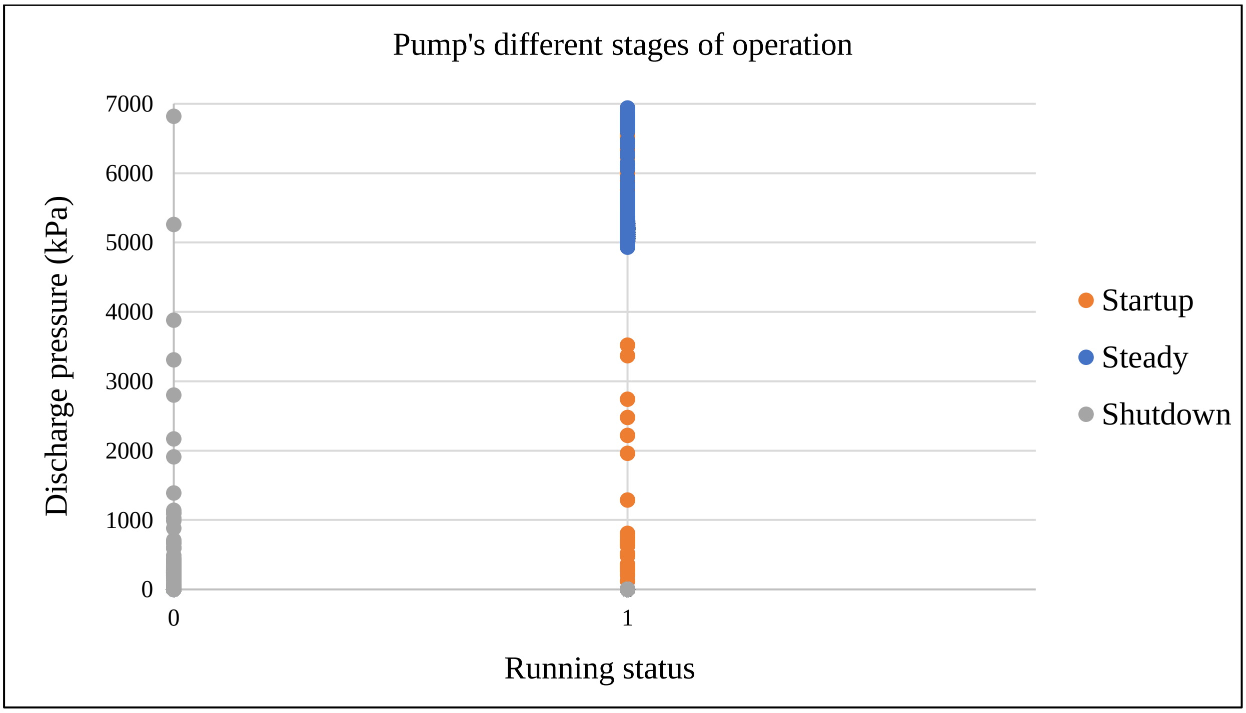


Figure 30: Example of different stages of pump operation

The blue dots in the scatter plot show the parameter values for when the system is at steady state. As expected, these values were close to the equipment's rated values. The orange dots show pump start-ups, with very low values, as expected [32], [33]. Lastly, the grey dots show pump shutdowns, whose values were also inconsistent [32], [33]. The rest of the data were visualised similarly to what is shown in Figure 30.

As can be observed in Figure 30, all the blue dots were found where the running status was 1 and resembled discharge pressures that are expected for a pump that is running. The orange dots were mostly concentrated between 0 kPa and approximately 3500 kPa. This implies that in the first two minutes (since two-minute resolution was used in this instance) of operation, the pump delivers low pressure. It is to be expected because the pump is still starting up and gradually building up pressure. The start-ups were filtered out by discarding values obtained when status changed from 0 to 1. The grey dots in Figure 30 are spread out randomly between 0 kPa and 7000 kPa. This implies that the pump's on-to-off transitions are also inconsistent. Both start-up and shutdown data points were filtered by applying Eq.7.

$$\boldsymbol{X(t_i)} = \begin{cases} \boldsymbol{Null} & \boldsymbol{if\ t_i \neq t_{i-1}} \\ \boldsymbol{X(t_i)} & \boldsymbol{otherwise} \end{cases} \qquad \text{Eq.7 (revisited)}$$

3.4 Step 3: Development and automation

3.4.1 Implementation

The development process began with the implementation of the methods presented in Section 3.3. As mentioned in Section 2.6.1, these methods were implemented in RTB – a system used to create and automate reports. Other software can also be used.

For a fully integrated solution, all equations developed in all the previous sections were incorporated into RTB. The following is a summary of the equations:

- Eq.3 – calculate power consumption from electric current and power factor (used for all pumps at Mine A with an assumed power factor of 0.9 [34]).
- Eq.5 – used to filter data according to running status for watchdog (both Mine A and B).
- Eq.7 – getting rid of pump start-up and shutdown data points (both Mine A and B).
- Eq.8 – modified pump efficiency formula that includes running status (both Mine A and B).

All these equations were integrated and implemented in RTB. The next sections provide the results obtained for both mines.

3.4.2 Results and validation

As explained in the previous sections, it was observed that each pump station is different in the number of pumps it has as well as the sizes of the pumps. Therefore, results are presented per pumping station. All results were compared with those that were obtained manually by an independent entity, to be referred to as Company X. Company X conducts efficiency tests through a manual method. The method involves physically placing sensors on pumps and recording instantaneous values. The values obtained by Company X were compared with the ones obtained by this study. The manual tests are completed monthly at both mines.

Mine A

As shown in Section 3.2, Mine A has five dewatering levels. The vertical distances between the adjacent levels range from 700 m to 3000 m [79]. At Mine A, it is very difficult to travel all the levels in a single day due to logistical reasons. For example, dewatering levels are seldom visited as there are no mining activities that take place, hence special travel arrangements are required. Due to how difficult it is to make these arrangements, the manual tests were typically carried out at an average rate of at most two levels per day.

From the 23rd to the 25th of November 2021, Company X took scheduled measurements at the mine, and the author went with to verify the manually obtained results. On the 23rd the tests were conducted on 75L followed by 100L and 115L on the 24th, and 52L and 29L on the 25th. However, as mentioned in Section 3.2, there is no reading for the column flow on 115L and hence none of the 115L pumps will be included in the calculations.

The following is a summary of the procedure that was followed for each pump on each level that was visited on the day of the site visit (29L and 52L):

- Only one pump was tested at a time.
- There are two columns on each of the two levels, and each column is shared by two pumps. Therefore, the control room operator was asked to only switch on the pump being tested.
- Each pump was allowed to run for two minutes before measurements were taken. This was to allow the system to reach steady state. The measurements were conducted as follows:
    - An ultrasonic flow meter was mounted on the discharge column to measure the discharge flow (as seen in Figure 31).

Figure 31: Portable flow meter setup

- A pressure gauge was mounted on the inlet and outlet pipes to measure the suction and discharge pressure, respectively (as seen in Figure 32).

Figure 32: Portable pressure gauge setup

- A portable power meter was used to take power consumption readings in the pump control room located near the pump stations (shown in Figure 33).

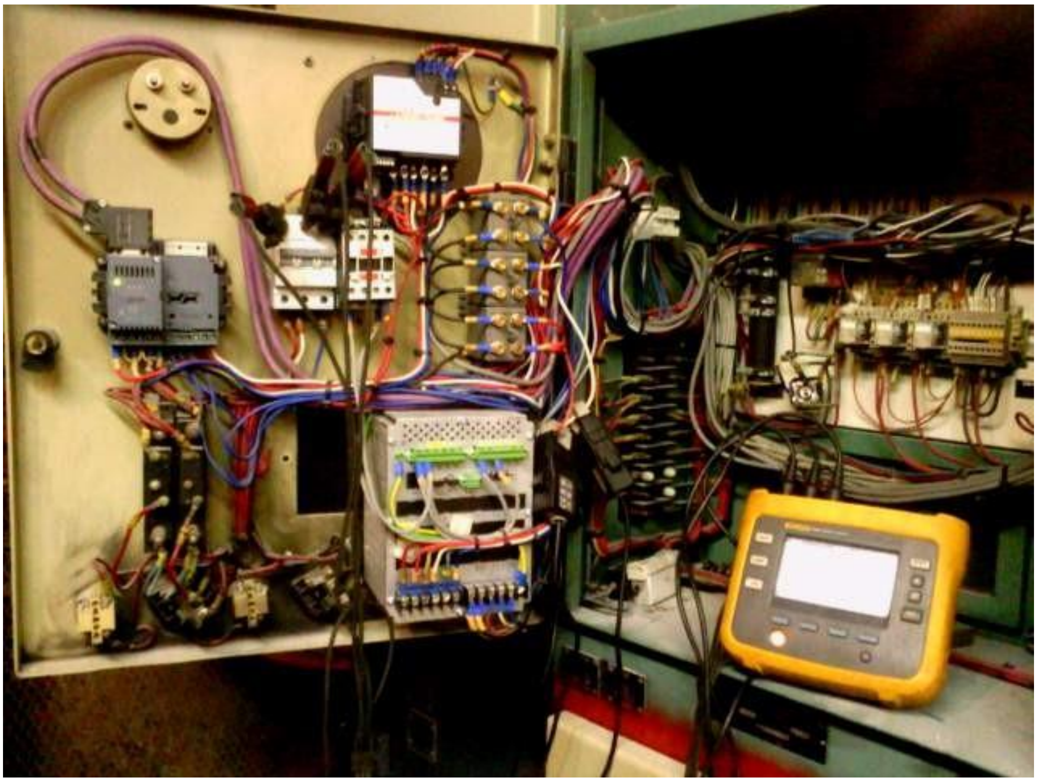


Figure 33: Portable Watt meter setup

The readings reflected on the portable meter were compared with those displayed on the installed meter in the control room (shown on Figure 34).

Figure 34: Underground control room readings

- All readings were verified against those reflected on the human machine interface (HMI) system located in the refuge chamber near the pump station (as shown on Figure 35).

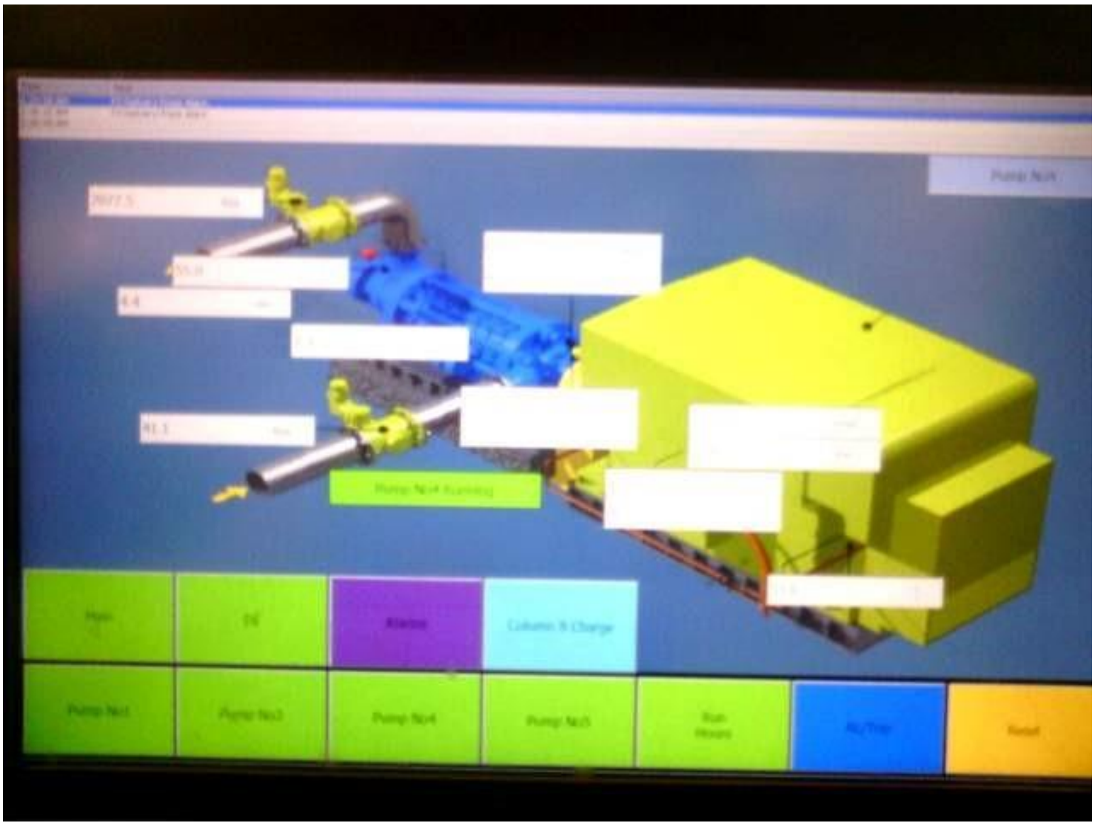


Figure 35: HMI display system

Efficiencies were consequently calculated using the obtained manual results. These efficiencies were compared with those obtained from the automated method. For consistent comparisons, the automated results obtained during the corresponding times were extracted. For example, on 29L, the manual tests were conducted from 08:00 to 09:00 and thus the same time period's efficiencies were compared. Table 12 presents the comparisons.

Table 12: Efficiencies comparisons at Mine A

Level	Pump	Manual efficiency (%)	Automated efficiency (%)
29L	P1	82	81
	P2	77	77
	P4	78	75
	P5	80	80
52L	P1	72	74
	P2	79	78
	P3	68	73
	P4	75	76
75L	P3	79	81
	P4	81	83
	P5	79	79
	P6	75	72
100L	P1	79	75
	P3	81	79
	P4	79	78
	P5	81	81

It can be observed from Table 12 that there were some slight differences between the two methods. These differences can be justified as follows:

- The manual test results represent the efficiencies of a pump at a particular point in time (instantaneous), whereas the automated results are obtained from aggregated two-minute data. With the two-minute data, it is not possible to pinpoint the exact time at which a SCADA data point was recorded. As discussed in Chapter 1, the dewatering system is dynamic. Therefore, results obtained at any two different instances may differ. Table 15 in Appendix A shows how each parameter varies over time. It was revealed that power consumption presents a significant variance. Therefore, this justifies the difference in results as power consumption presents a high sensitivity.

- There was a lag between the times at which the manual values for all different parameters were recorded. The time differences can lead to different efficiencies as the individual readings fluctuate as mentioned above.

The sensitivity analysis proved how pump efficiency can be affected by a change in any of the non-ignorable parameters' values, and thus the importance of having accurate data. This, as well as the asynchronous data collection times between the manual and automated results, justifies the differences observed. Thus, the absolute efficiency values may not be consistent between the two methods. However, the manual results provide a benchmark for the kind of values that can be expected from the automated results. For example, if the discrepancy between a manually vs automatically obtained efficiency was more than 10%, then that would raise a question about either method. However, looking at the values presented for Mine A (Table 12), the results obtained can be trusted as the differences are small enough (less than 5% in each case) and can be neglected. Therefore, the developed method correctly quantified efficiencies at Mine A.

Mine B

As discussed in Section 3.3, the following data issues were encountered at Mine B:

- All 76L and 78L pumps do not have suction pressure readings.
- 78L column has a static discharge flow reading.
- 3000L pump P2 has a static power consumption reading.

As illustrated by the sensitivity analysis, suction pressure can be neglected from the calculation without having a significant effect on the result. However, the sensitivity analysis revealed that discharge flow and power consumption are critical and hence the efficiencies of the 78L pumps as well as that of 3000L pump P2 cannot be calculated. That implies that, for Mine B, efficiencies will only be quantified for:

- 76L pumps P1, P2 and P3
- 61L pumps P1 and P2
- 3000L pump P1

The results are presented in the following section. The automated results were also compared to the manual test results which were obtained on the 15th of November 2021. The same procedure as in Mine A was followed to obtain the results. Table 13 presents the results obtained by the two methods.

Table 13: Efficiencies comparisons at Mine B

Level	Pump	Manual efficiency (%)	Automated efficiency (%)
78L	78L P4	79	77
	78L P5	78	80
	78L P6	78	78
61L	61L P1	83	80
	61L P2	82	78
3000L	3000L P1	77	73

As in the case of Mine A, the results obtained at Mine B were consistent between the two methods. Based on the arguments presented regarding the results at Mine A, such as time discrepancies involved in the data acquisition processes involved in the two methods, the results at Mine B were also expected to differ to a certain degree., The results presented in Table 13 only have small differences, which can be accepted. Therefore, it can be concluded that the method was successful in quantifying efficiencies at Mine B.

3.5 Step 4: Reporting of findings to end users

After several consultations and meetings with various end users (mine senior engineers, shaft engineers, instrumentation technicians, etc.) a monthly pumping efficiency report was proposed. Although the report is sent monthly, the efficiencies are still calculated daily. The benefit of calculating daily efficiencies is to ensure that issues are identified timeously, in which case the mine will be notified. The monthly report consists of monthly pump efficiencies for a period of six months. The reasoning behind the monthly period was as follows:

1. The fact that the efficiencies are dependent on individual pump operation as discussed in previous sections. Therefore, each pump is more likely to have run individually by the time the report is distributed.
2. Monthly reporting can provide an overview of the pump's life cycle over a longer period.
3. The report allows incorporation of financial figures, most of which can only be quantified monthly. Those figures are presented in Appendix A

The pumping efficiency report consists of average monthly pump efficiencies for each pump. The report consists of two sections. The first section comprises historic pump efficiencies while the second is an overview of the monthly pump scheduling as well overall energy consumption.

Mine A: historic efficiencies

The section is broken up into pumping stations, as was Section 3.4.2. As mentioned, this section gives average monthly efficiencies for a period of six months.

The historic efficiencies obtained on 29L are shown in Figure 36.

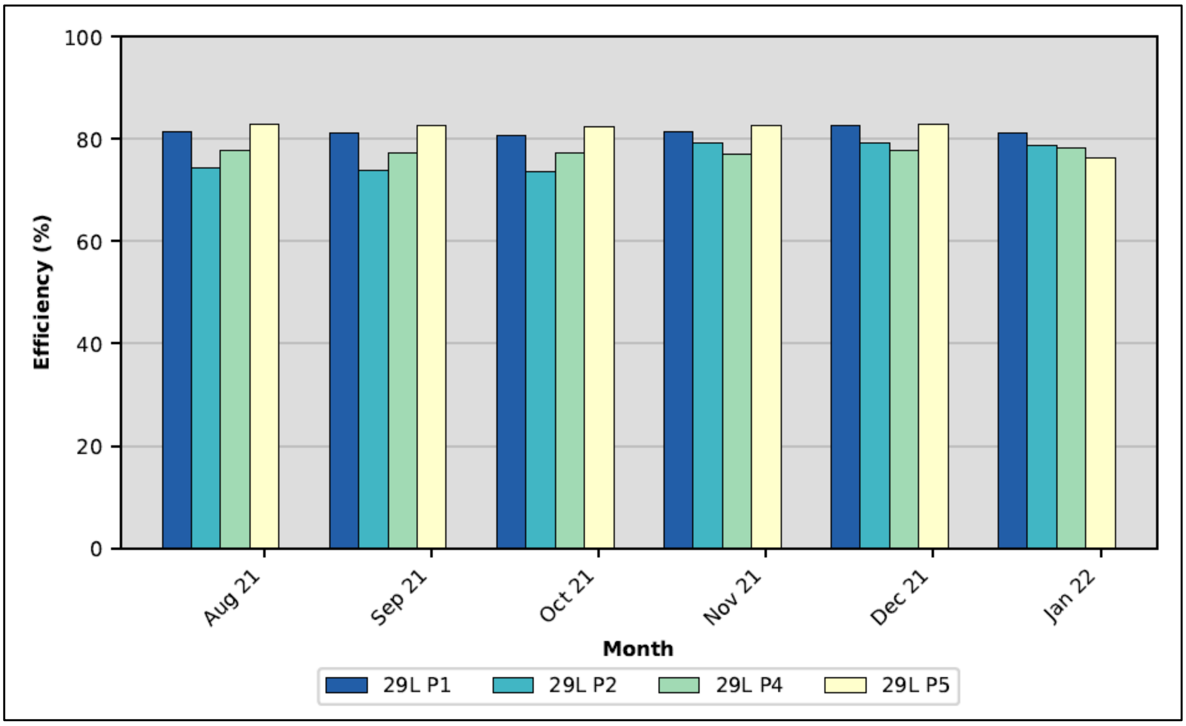


Figure 36: 29L history pump efficiencies

As shown in Figure 36, the pumps all had similar efficiencies for this reporting period. Pump P5 was the most efficient pump up until Jan 2022 where its efficiency dropped significantly, while the rest remained relatively constant. An investigation into this revealed that the discharge pressure for pump P5 was low in January 2022 compared to previous months. This is shown in Figure 37.

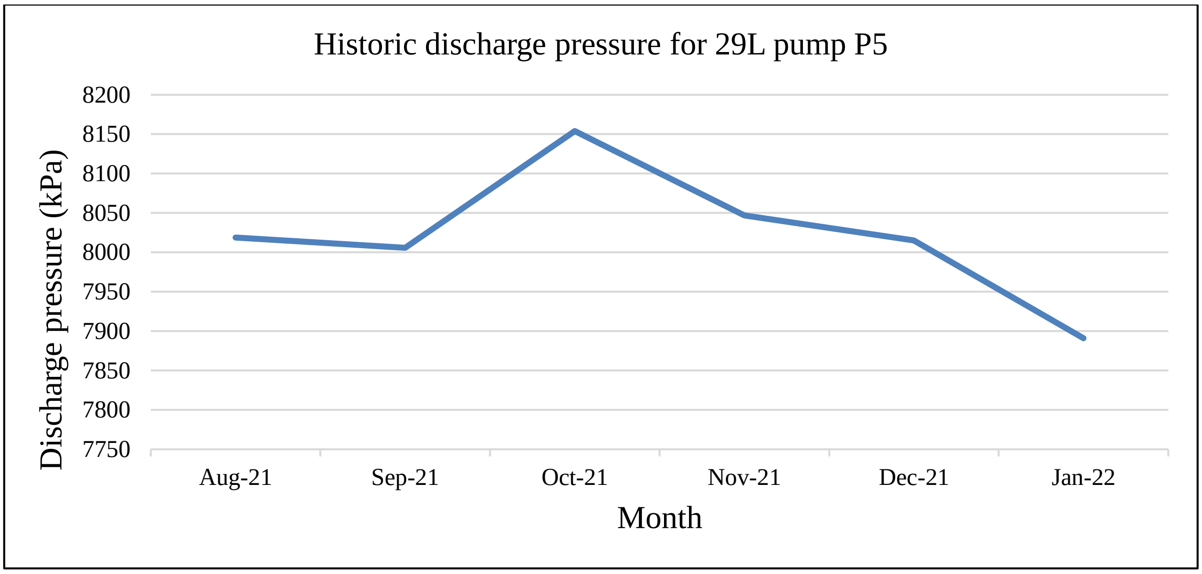


Figure 37: 29L pump P5's historic monthly average pressures while running

This occurred after the pump had not been operated for a while (from 25 December 2021 until 16 January 2022). Figure 38 shows the total hours for which the pump was running in January 2022. It can be observed that the pump only started running from the 18th of January 2022.

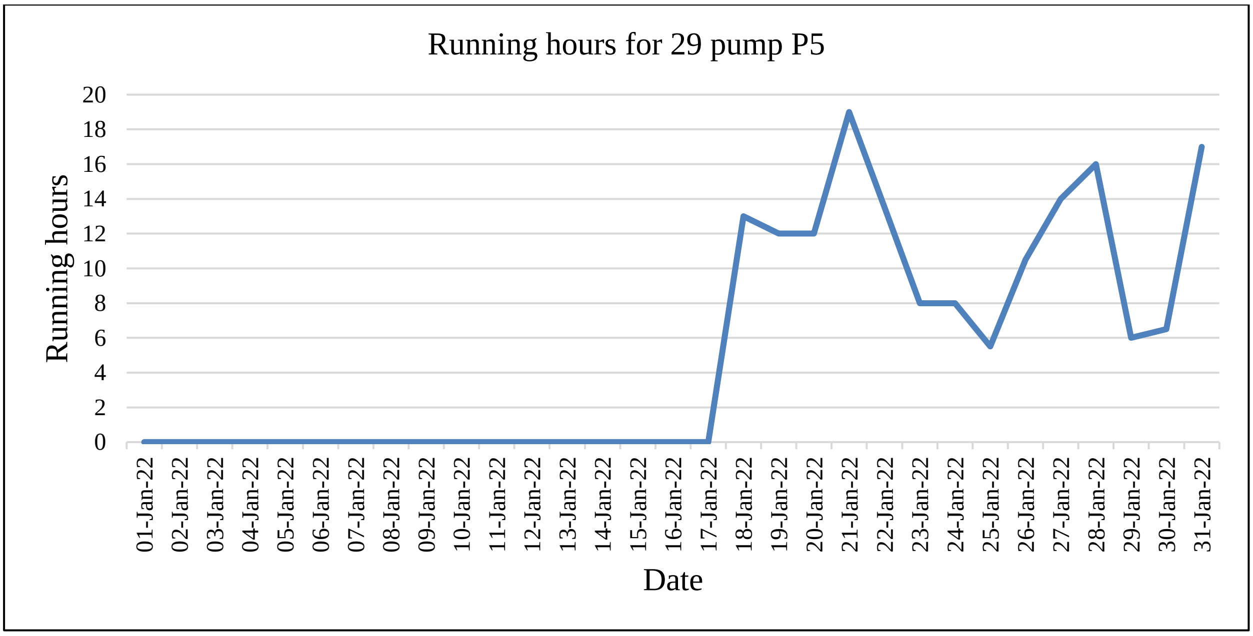


Figure 38: 29L pump P5's running hours in January 2022

As discussed in Chapter 1, the water that goes through these pumps contains mud [6]. If a pump is not operated for a long time and there happens to be some remaining mud inside it, it can dry up, thus causing blockage and a drop in efficiency. This seemed to have been the case for the first two days after the pump resumed activity, from where the efficiency went back to normal.

Figure 39 presents the historic efficiencies obtained on 52L.

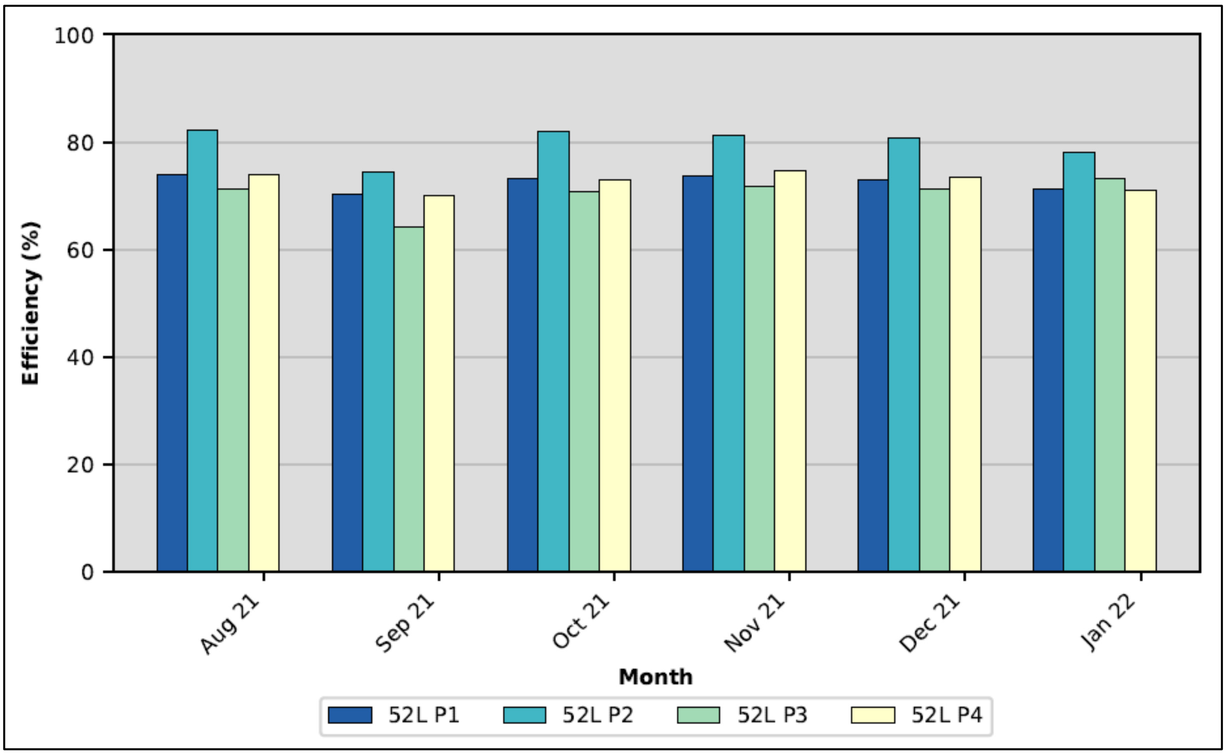


Figure 39: 52L historic pump efficiencies

On 52L, pump P2 proved to be the most efficient pump. The rest of the pumps' efficiencies seemed to fluctuate throughout the reporting period. In this case, no issues were found which could have led to the fluctuations. This can, therefore, be attributed to the dynamic nature of the system.

The historic efficiencies obtained on 75L are presented in Figure 40.

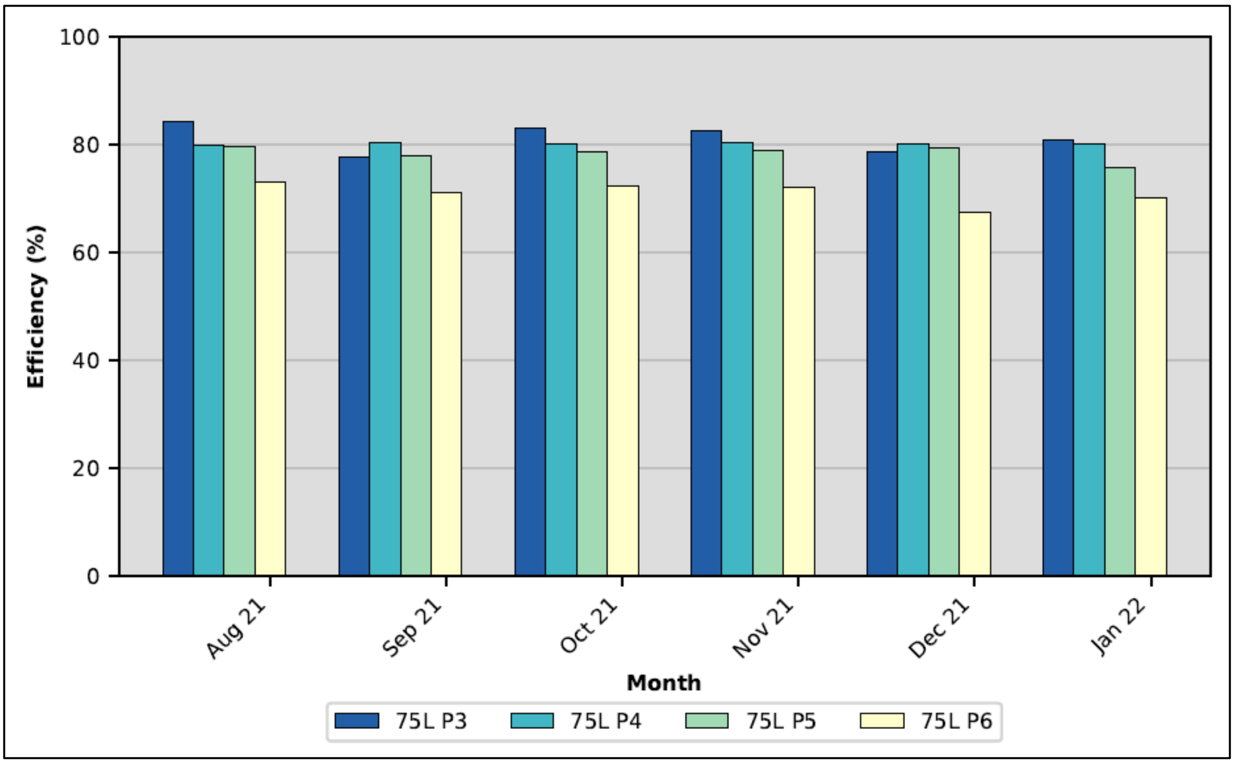


Figure 40: 75L historic efficiencies

On this level, the highest efficiency seemed to generally alternate between pumps P3 and P4. Furthermore, the efficiencies for all pumps also fluctuated throughout the reporting period, although the efficiency rankings remained consistent.

Figure 41 presents the efficiencies obtained on 100L.

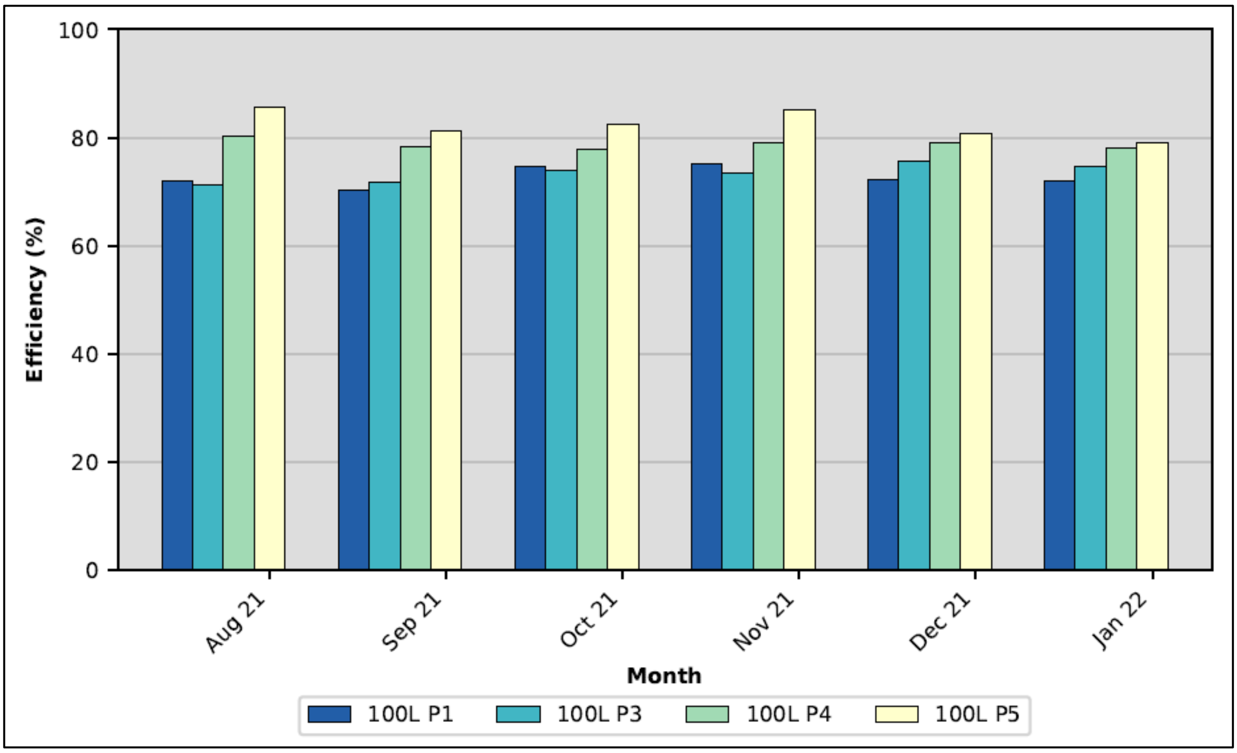


Figure 41: 100L historic pump efficiencies

As can be observed in Figure 41, 100L pump P5 proved to be the most efficient pump throughout, followed by pump P4. Pumps P1 and P2 alternated between third and fourth position respectively, in terms of efficiency rankings.

Mine B: historic efficiencies

This section presents the historic monthly average efficiencies for Mine B. These form part of the monthly pumping efficiency report for the mine.

Figure 42 shows the historic efficiencies obtained on 76L.

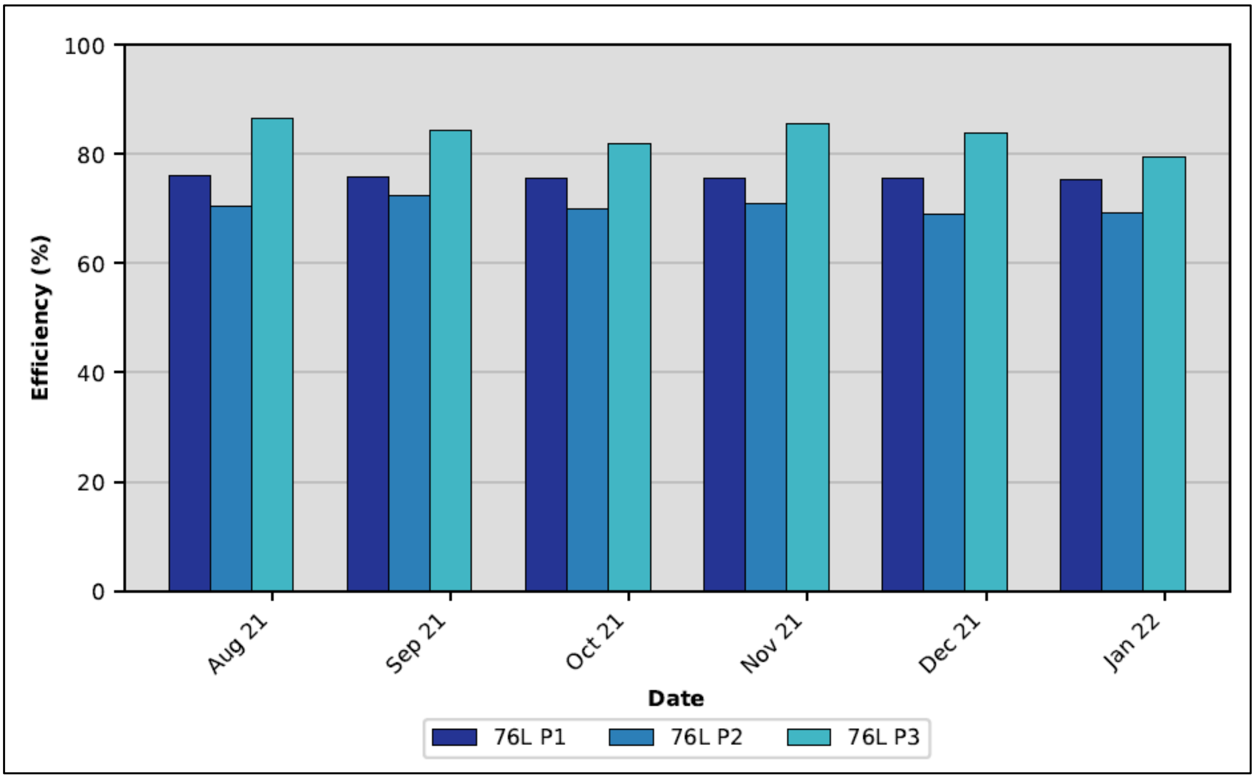


Figure 42: 76L historic monthly pump efficiencies

As can be observed from Figure 42, pump P1 was consistently the most efficient pump amongst the three, followed by P3, and then P2.

Figure 43 shows the historic efficiencies on 61L.

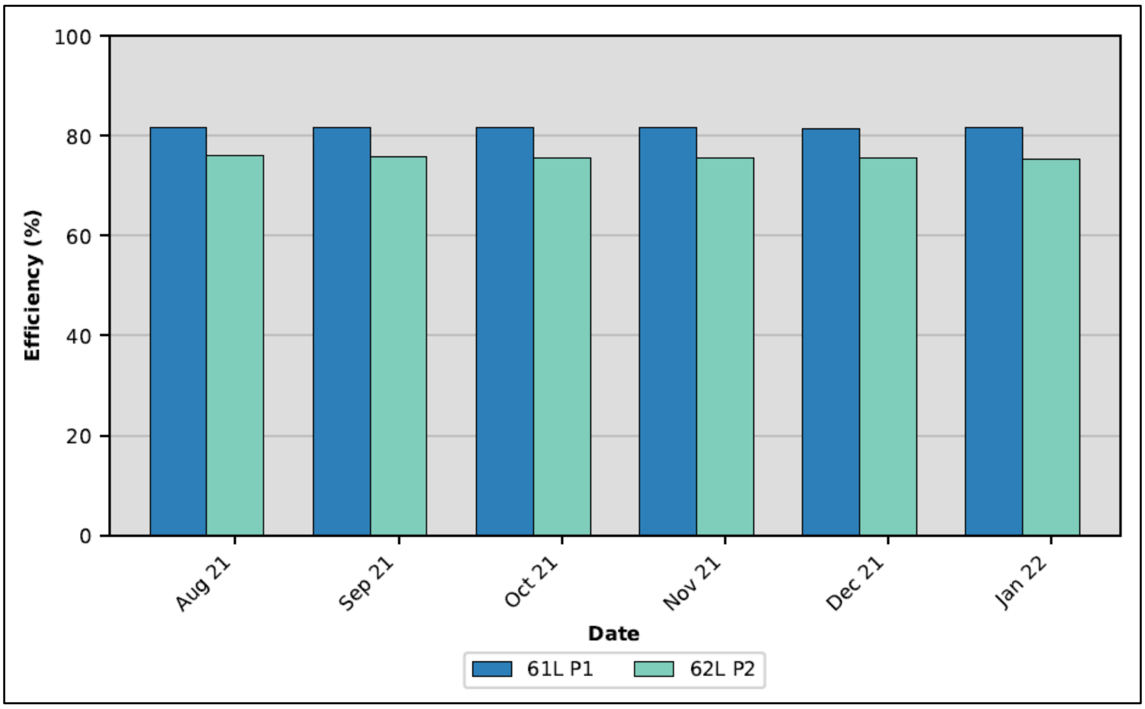


Figure 43: 61L historic monthly pump efficiencies

As can be observed from Figure 43, the 61L pump P1 was consistently the most efficient amongst the two pumps on 61L. Furthermore, the efficiencies of both pumps were almost constant throughout the six-month period.

Figure 44 shows the historic efficiencies obtained on 3000L.

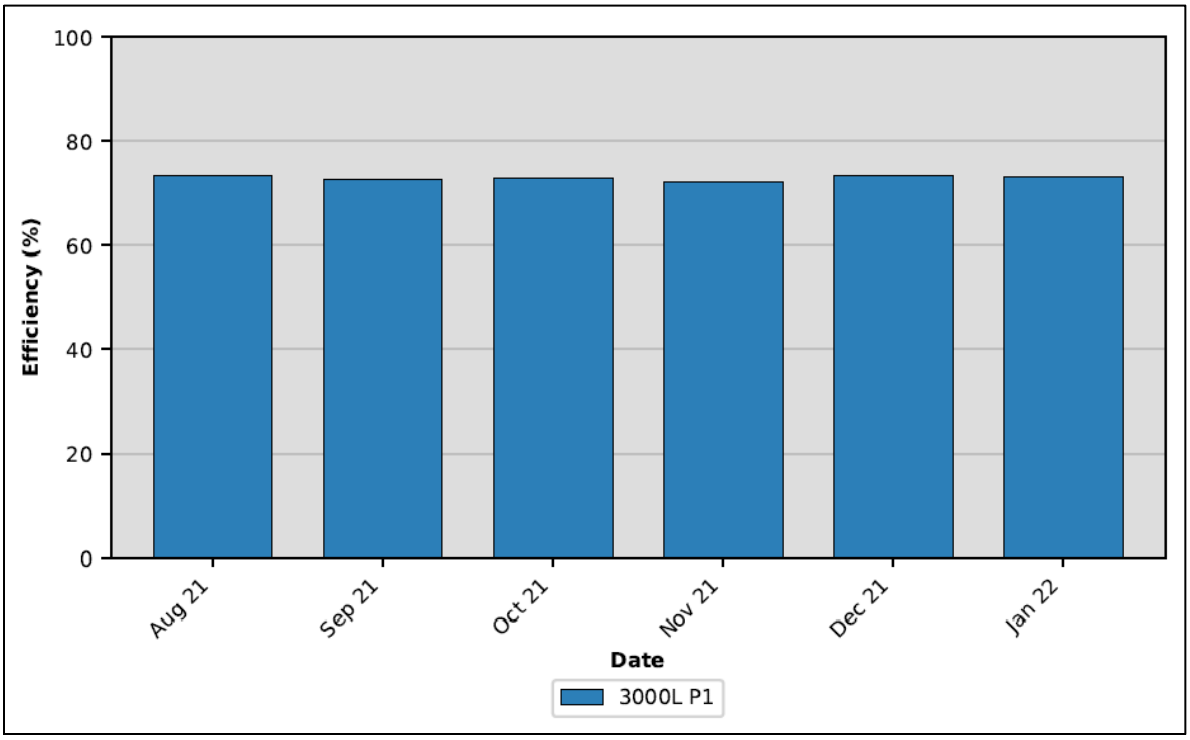


Figure 44: 3000L pump P1's historic monthly pump efficiencies

As mentioned in Section 3.3.1, on 3000L, only pump P1 had all the necessary measurements required for efficiency calculation, hence only its efficiencies are presented. Its efficiency proved to be consistent, with December 2021 and January 2020 seeing slightly higher efficiencies.

3.6 Impact of study

This study has the following potential benefits:

1. Alleviate the cost of hiring external contractors. As mentioned in Section 1.1.4, labour and external contractors constitute about 31% of a mine's operating cost. This study has proven the feasibility and accuracy of the automated method, which requires minimal human intervention.
2. Time and resource reallocation. The mine provides foremen to accompany external contractors underground as per mine safety rules. Automating the pump efficiency process will allow the foremen to be allocated other work that can enhance the mine's productivity.
3. Adoption of the developed data quality framework. The data understanding process brought about a flexible framework through which data quality challenges were addressed. As discussed in Chapter 1, data challenges are not only unique to pump-related measurements but

are prevalent to other measurements as well. Thus, the framework can be adopted and utilised to address any applicable data quality problem.

4. Appendix A provided a breakdown of the energy figures that form part of the report discussed in the previous section. This part of the report is used by high-level managers to oversee the energy consumptions of the pumps. The energy figures at Mine A revealed the need for an energy optimisation mechanism as there is currently no load shifting mechanism at the mine.

3.7 Conclusion

In this chapter, two case studies were identified through which the methodology was tested and implemented. An automated data quality monitoring system was implemented in RTB. Pump efficiencies were quantified using the clean data. For validation, the efficiencies were compared with manually obtained results. Pump efficiency reports were developed for both case studies, containing pump efficiencies as well as energy consumption figures. This chapter was concluded with a discussion of the potential impact of this study.

CHAPTER 4

CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

Deep-level mines consist of dewatering systems which are responsible for the removal of underground water originating from mine services as well as fissure water. A dewatering system consists of a series of dewatering dams and pumps. Dewatering pumps are responsible for pumping water from lower-level dams to upper-level dams through columns. The water is moved until it reaches the mine's surface where it can be processed and recirculated.

Pumps are some of the biggest consumers of electricity in deep-level mines. Due to harsh conditions under which dewatering pumps operate, their efficiencies deteriorate rapidly. Pump efficiency was found to be an important metric for pump condition monitoring. Mines hire external contractors to conduct scheduled manual pump efficiency audits. However, the dynamic nature of the dewatering pump efficiencies necessitates the need for an automated method of quantifying efficiencies, which can help avoid human error as well as provide timeous feedback.

Previous studies revealed various data challenges encountered in deep-level mines which negatively impact data quality. Research revealed a gap in literature pertaining to quantifying dewatering pump efficiencies in deep-level mines, considering the data challenges faced by the industry.

This study proposed an automated method through which dewatering pump efficiencies can be quantified. The study objectives were set out as follows:

1. Address the following data challenges through a DA approach:
    - Missing data because of lack of instrumentation or communication issues.
    - Static data challenges because of instrumental errors or communication issues.
    - Outlier data points
2. Develop an automated system that can quantify efficiencies continuously.
3. The system must be applicable to any centrifugal pump in a deep-level mine.

A methodology was developed (Chapter 2) based on previous DA methods that were found to be closely related to this study. The methodology was evaluated by identifying case studies on which the proposed solution could be tested, developing a data quality monitoring system for the required data, and implementing the solution in RTB for automated reporting purposes, through which objectives 1 and 2 were addressed.

The developed methodology was tested on two case studies. The case study mines belong to a deep-level gold mining group in South Africa. The dewatering systems of the two case study mines differ in terms of complexity and size. Mine A is deeper and therefore consists of more dewatering levels and pumps. By successfully applying the methodology on these case studies, the third objective was met.

Results from both case studies highlighted the importance of having a data quality monitoring system. The author validated the automatically quantified efficiencies by comparing them with efficiencies obtained through manual measurements by an impartial efficiency measurement contractor.

The automated solution requires minimal human intervention and can thus help the mine reduce the cost associated with hiring external contractors to conduct efficiency audits. The solution can also help enhance labour productivity as foremen can be allocated to other work as opposed to accompanying external contractors to dewatering levels.

The developed solution met all the study objectives: the data quality monitoring system addressed the data challenges and implementation of the automated pump efficiency quantification method was successful.

4.2 Recommendations for further research

There were assumptions made throughout this study, which led to some limitations. These limitations provide opportunity for further research. The following are recommendations for future research related to this study:

1. The results obtained at Mine A revealed that there is no load shifting mechanism on dewatering pumps, and hence the pumps operate in a constant pattern regardless of TOU period. The results also revealed that the pumping electricity cost realised during peak times was very high. The solution to this could be to implement a load shifting mechanism whereby most of the pumping is scheduled for non-peak times or use efficiencies to allocate pumps to operate during peak times. Therefore, this presents an opportunity for a study that can research the feasibility of implementing either method.
2. This study did not investigate other pump performance metrics such as temperature, vibration, etc. Future studies can investigate and incorporate these metrics into the efficiency monitoring to form a complete pump condition monitoring system
3. The data analysis and monitoring system developed in this study can be expanded to suit other applications within the mining industry, such as for determining compressor efficiency or developing a holistic data quality management framework for other mining equipment. Future studies can investigate the applicability of the system to be adopted for such applications.

REFERENCES

[1] W. Schoeman, N. Oosthuizen, and J. van Rensburg, "Optimum water distribution between pumping stations of multiple mine shafts," in *2013 Proceedings of the 10th Industrial and Commercial Use of Energy Conference*, Aug. 2013, pp. 1-4.

[2] J. Vosloo, L. Liebenberg, and D. Velleman, "Case study: Energy savings for a deep-mine water reticulation system," *Applied Energy*, vol. 92, pp. 328-335, Apr. 2012, doi: 10.1016/j.apenergy.2011.10.024.

[3] Z. Peijie and L. Baozhu, "The application of TOPSIS method to deep mine water environmental quality assessment," in *2011 International Conference on Electrical and Control Engineering*, Sept. 2011, pp. 1802-1806, doi: 10.1109/ICECENG.2011.6058116.

[4] P. Oberholzer, "Best practices for automation and control of mine dewatering systems," M.Eng dissertation, North-West University, Potchefstroom, 2015.

[5] J. Stols, H. Groenewald, and J. van Rensburg, "Quantifying the effects of system constraint improvements on the electricity cost of dewatering pumps," in *2016 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Aug. 2016, pp. 123-127.

[6] S. Taljaard, "Comparative study of mine dewatering control systems," M.Eng dissertation, North-West University, Potchefstroom, 2018.

[7] B. Pascoe, J. van Rensburg, and R. Joubert, "Dewatering processes to sustain savings when a critical system fails," in *2018 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Aug. 2018, pp. 1-5.

[8] W. Schoeman, J. van Rensburg, and G. Bolt, "Cost-effective methods for automisation of a mine pumping systems to realise energy cost savings," in *2011 Proceedings of the 8th Conference on the Industrial and Commercial Use of Energy*, Aug. 2011, pp. 111-115.

[9] A. Appuhamillage and D. Senadhira, "Optimization of energy utilization for dewatering system in Bogala Graphite mine, Aruggammana, Sri Lanka," M.S dissertation, University of Sri Lanka, Nugegoda, 2014.

[10] A. Romero, D. Millar, M. Carvalho, J. M. Maestre, and E. F. Camacho, "A comparison of the economic benefits of centralized and distributed model predictive control strategies for optimal and sub-optimal mine dewatering system designs," *Applied Thermal Engineering*, vol. 90, pp. 1172-1183, Nov. 2015, doi: 10.1016/j.applthermaleng.2015.01.031.

[11] H. Grobbelaar, "Maintenance procedures on DSM pumping projects to improve sustainability," M.Eng dissertation, North-West University, Potchefstroom, 2015.

[12] C. Cilliers, "Cost savings on mine dewatering pumps by reducing preparation- and comeback loads," M.Eng dissertation, North-West University, Potchefstroom, 2014.

[13] J. P. de Jager, H. J. Groenewald, and C. Cilliers, "Investigating the effect of pump availability on load shift performance," in *2016 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Aug. 2016, pp. 98-103.

[14] S. Walker, "The pump-makers' perpective," *Engineering & Mining Journal*, vol. 214, 11, pp. 50-54, Nov. 2013.

[15] B. Jafarzadeh, A. Hajari, M. M. Alishahi, and M. H. Akbari, "The flow simulation of a low-specific-speed high-speed centrifugal pump," *Applied Mathematical Modelling*, vol. 35, no. 1, pp. 242-249, Jan. 2011, doi: 10.1016/j.apm.2010.05.021.

[16] L. Jayamaha, *Energy-Efficient Industrial Systems: Evaluation and Implementation*, 1st ed. New York: McGraw-Hill Education, 2016.

[17] J. Hall, "Process pump control," *Chemical Engineering*, vol. 117, no. 12, pp. 30-33, 2010.

[18] J. K. Igor, P. M. Joseph, C. Paul, and C. H. Charles, *Pump Handbook*, 4th ed. New York: McGraw-Hill Education, 2008.

[19] A. D. Build, "The mathematics of pumping water," *STEM learning*, 2019. Royal Academy of Engineering.

[20] P. Sorin and H. Liviu Valer, "Reducing energy consumption by upgrading pumping stations in water distribution systems," *Environmental engineering and management journal*, vol. 12, Sept. 2013, doi: 10.30638/eemj.2013.090.

[21] W. Schoeman, J. van Rensburg, and G. Bolt, "Automation of a mine pumping system to realise energy cost savings," ed: Energize, pp. 36-38.

[22] R. Fockema, "A customisable data analysis interface for an online electrical energy information system," M.Eng dissertation, North-West University, Potchefstroom, 2014.

[23] T. Masuku, "A target-based performance tracker for mine energy monitoring," M.Eng dissertation, North-West University, Potchefstroom, 2020.

[24] Eskom, *Tariffs & Charges Booklet*. Eskom, 2021.

[25] Y. Kim *et al.*, "Smart day-ahead pump scheduling scheme for electricity cost optimization in a sewage treatment plant," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2019, pp. 565-567, doi: 10.1109/ICTC46691.2019.8939710.

[26] H. Groenewald, A. Stols, and J. van Rensburg, "The value of maintaining a DSM load shifting project on mine dewatering pumps," in *2013 Proceedings of the 10th Industrial and Commercial Use of Energy Conference*, Aug. 2013, pp. 1-4.

[27] T. Smith, H. Joubert, and J. van Rensburg, "Automated control of mine dewatering pumps to reduce electricity cost," in *2015 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Aug. 2015, pp. 62-69, doi: 10.1109/ICUE.2015.7280248.

[28] M. Šavar, H. Kozmar, and I. Sutlović, "Improving centrifugal pump efficiency by impeller trimming," *Desalination*, vol. 249, no. 2, pp. 654-659, Dec. 2009, doi: 10.1016/j.desal.2008.11.018.

[29] M. H. Shojaeefard, B. Salimian Rizi, A. Khalkhali, and M. Tahani, "A new method to calculate centrifugal pump performance parameters for industrial oils," *Journal of Applied Fluid Mechanics*, vol. 8, no. 4, pp. 673-681, 2015.

[30] Hazen, "Hydraulic considerations in pumping system design," in *VWEA Wastewater Operations Education Conference*, Roanoke Virginia, July 2016: Virginia Water Environment Association.

[31] J. Chaurette, "Centrifugal pump systems." Canada: Fluide Design Inc, 2005.

[32] U.S. Department of Energy, "Determining electric motor load and efficiency," in "Motor Challenge," U.S. Department of Energy [Online]. Available: https://www.energy.gov (accessed: 15 Dec. 2021).

[33] M. Saravanan and D. Winston, "Single parameter fault identification technique for DC motor through wavelet analysis and fuzzy logic," *Journal of Electrical Engineering and Technology*, vol. 8, no. 5, pp. 1049-1055, 2013.

[34] A. Ukil, R. Bloch, and A. Andenna, "Estimation of induction motor operating power factor from measured current and manufacturer data," *IEEE Transactions on Energy Conversion*, vol. 26, no. 2, pp. 699-706, 2011, doi: 10.1109/TEC.2010.2098408.

[35] J. Stols, "Implementing a remote condition monitoring system for South African gold mines," M.Eng dissertation, North-West University, Potchefstroom, 2019.

[36] L. Wen-Guang, "Effects of viscosity of fluids on centrifugal pump performance and flow pattern in the impeller," *International Journal of Heat and Fluid Flow*, vol. 21, no. 2, pp. 207-212, Apr. 2000, doi: 10.1016/S0142-727X(99)00078-8.

[37] A. R. Al-Obaidi and H. Towsyfyan, "An experimental study on vibration signatures for detecting incipient cavitation in centrifugal pumps based on envelope spectrum analysis,"

Journal of Applied Fluid Mechanics, vol. 12, no. 6, pp. 2057-2067, Nov. 2019, doi: 10.29252/jafm.12.06.29901.

[38] U.S. Department of Energy, *Improving pumping system performance*, 2nd ed. Berkeley, California: Lawrence Berkeley National Laboratory, 2006.

[39] M. D. Anis, "Identifying a mathematical model to optimize pump maintenance planning decisions - a case of irrigation asset management in K.S.A," in *2018 Condition Monitoring and Diagnosis (CMD)*, Sept. 2018, pp. 1-6, doi: 10.1109/CMD.2018.8535792.

[40] V. Vodovozov, I. Bakman, Z. Raud, and T. Lehtla, "High-efficiency predictive control of pumping," in *2016 15th Biennial Baltic Electronics Conference (BEC)*, Oct. 2016, pp. 203-206, doi: 10.1109/BEC.2016.7743764.

[41] T. Augustyn, "Energy efficiency and savings in pumping systems — The holistic approach," in *2012 Southern African Energy Efficiency Convention (SAEEC)*, Nov. 2012, pp. 1-7, doi: 10.1109/SAEEC.2012.6408587.

[42] Statista. "South Africa's mining industry distribution of expenses 2021." https://www.statista.com (accessed 10 March 2022).

[43] E. Brzychczy, P. Gackowiec, and M. Liebetrau, "Data analytic approaches for mining process improvement - machinery utilization use case," *Resources*, vol. 9, p. 17, Jul. 2020, doi: 10.3390/resources9020017.

[44] A. Lakshmi, B. Sangeetha, A. Naveenkumar, B. Ganesh, and N. Bharathi, "Experimental validation of PID based cascade control system through SCADA-PLC-OPC interface," in *2012 International Conference on Computer Communication and Informatics*, Jan. 2012, pp. 1-4, doi: 10.1109/ICCCI.2012.6158893.

[45] K. Chakraborty, P. De, and I. Roy, *Industrial applications of programmable logic controllers and SCADA*. Hamburg: Anchor Academic Publishing, 2016.

[46] S. Dutta, A. Sarkar, K. Samanta, R. Das, and A. Ghosh, "Supervision of control valve characteristics using PLC and creation of HMI by SCADA," in *2014 First International Conference on Automation, Control, Energy and Systems (ACES)*, Feb. 2014, pp. 1-5, doi: 10.1109/ACES.2014.6808025.

[47] A. Daneels and W. Salter, "What is SCADA?," *International Conference on Accelerator and Large Experimental Physics Control Systems*, pp. 339-343, 1999.

[48] V. Baggiolini *et al.*, "Results of the OPC evaluation done within the JCOP for the control of the LHC experiments," *International Conference on Accelerator and Large Experimental Physics Control Systems*, pp. 511-513, 1999.

[49] J. H. Hellerstein, "Quantitative data cleaning for large databases," EECS Computer Science Division, UC Berkeley, 2008. [Online]. Available: http://db.cs.berkeley.edu (accessed 10 Nov. 2021).

[50] P. J. J. Van Wyk, G. A. Oosthuizen, and P. T. T. Conradie, "Evaluation of communication systems for deep level mining monitoring applications," in *ISEM 2011 Proceedings*, Stellenbosch, South Africa, Sept. 2011 2011.

[51] A. Goosen, "A system to quantify industrial data quality," M.Eng dissertation, North-West University, 2019.

[52] J. De Meyer, "Validating the integrity of single source condition monitoring data," M.Eng dissertation, North-West University, 2020.

[53] H. Sike, J. Xuejing, and G. Huifen, "Optimization of the number of multiple pumps running simultaneously in open cycle cooling water system in power plant," *Energy Procedia*, vol. 17, pp. 1161-1168, Jan. 2012, doi: 10.1016/j.egypro.2012.02.222.

[54] C. Lettner, R. Stumptner, W. Fragner, F. Rauchenzauner, and L. Ehrlinger, "DaQL 2.0: Measure data quality based on entity models," *Procedia Computer Science*, vol. 180, pp. 772-777, Jan. 2021, doi: 10.1016/j.procs.2021.01.327.

[55] R. Mahanti, *Data Quality*. Milwaukee, Wisconsin: Quality Press, 2018.

[56] K. Tim and S. Julian, *Managing Data Quality : A Practical Guide*. Swindon, United Kingdom: BCS, The Chartered Institute for IT, 2020.

[57] R. Vaziri, M. Mohsenzadeh, and J. Habibi, "Measuring data quality with weighted metrics," *Total Quality Management & Business Excellence*, Article vol. 30, no. 5/6, pp. 708-720, 2019, doi: 10.1080/14783363.2017.1332954.

[58] F. Sidi, P. H. S. Panahy, L. S. Affendey, M. A. Jabar, H. Ibrahim, and A. Mustapha, "Data quality: A survey of data quality dimensions," in *2012 International Conference on Information Retrieval & Knowledge Management*, Mar. 2012, pp. 300-304, doi: 10.1109/InfRKM.2012.6204995.

[59] C. Batini, D. Barone, F. Cabitza, and S. Grega, "A data quality methodology for heterogeneous data," *International Journal of Database Management Systems*, vol. 3, pp. 60-79, 2011.

[60] R. Mahanti, "Data quality and data quality dimensions," *Software Quality Professional*, Article vol. 22, no. 1, pp. 4-8, 2019.

[61] S. C. A. Ramasamy, "Big data quality dimensions: a systematic literature review," *Journal of Information Systems and Technology Management*, vol. 17, 2020, doi: 10.4301/S1807-1775202017003.

[62] T. Ikbal, S. Mohamed Adel, B. Chafik, and D. Rachida, "Big data quality framework: a holistic approach to continuous quality management," *Journal of Big Data*, article vol. 8, no. 1, pp. 1-41, Jan. 2021, doi: 10.1186/s40537-021-00468-0.

[63] S. Moossavizadeh, "A new approach to measure believability dimension of data quality," *Growing Science*, vol. 2, no. 7, p. 5, 2012.

[64] B. Bohn, J. Olson, B. Gopaluni, and B. Stoeber, "Sensing concept for practical performance-monitoring of centrifugal pumps," in *2019 IEEE SENSORS*, Oct. 2019, pp. 1-4, doi: 10.1109/SENSORS43011.2019.8956559.

[65] M. E. Qazizada, V. Sviatskii, and P. Bozek, "Analysis performance characteristics of centrifugal pumps," *MM Science Journal*, Article pp. 1151-1159, 2016 2016, doi: 10.17973/MMSJ.2016_10_201691.

[66] M. E. Qazizada and E. Pivarčiová, "Reliability of parallel and serial centrifugal pumps for dewatering in mining process," *Acta Montanistica Slovaca*, Article vol. 23, no. 2, pp. 141-152, 2018.

[67] R. Huang, Z. Zhang, W. Zhang, J. Mou, P. Zhou, and Y. Wang, "Energy performance prediction of the centrifugal pumps by using a hybrid neural network," *Energy*, Article vol. 213, Dec. 2020, doi: 10.1016/j.energy.2020.119005.

[68] Y. Wang and X. Huo, "Multiobjective optimization design and performance prediction of centrifugal pump based on orthogonal test," *Advances in Materials Science & Engineering*, pp. 1-10, Feb. 2018, doi: 10.1155/2018/6218178.

[69] A. Kara Omar, A. Khaldi, and A. Ladouani, "Prediction of centrifugal pump performance using energy loss analysis," *Australian Journal of Mechanical Engineering*, Article vol. 15, no. 3, pp. 210-221, 2017, doi: 10.1080/14484846.2016.1252567.

[70] A. Martinus, J. H. C. Pretorius, and A. Wessels, "Predictive maintenance as a means to improve the availability of centrifugal slurry pump at Ergo City Deep Plant," *Proceedings of the International Conference on Industrial Engineering & Operations Management*, Article pp. 1204-1215, May. 2019.

[71] M. Tan, Y. Lu, X. Wu, H. Liu, and X. Tian, "Investigation on performance of a centrifugal pump with multi-malfunction," *Journal of Low Frequency Noise, Vibration & Active Control*, Article vol. 40, no. 2, pp. 740-752, 2021, doi: 10.1177/1461348420942349.

[72] S. Aiamsamang, T. Thepmanee, P. Julsereewong, and A. Julsereewong, "LabVIEW-based system for testing centrifugal pump efficiency," in *2017 56th Annual Conference*

of the Society of Instrument and Control Engineers of Japan (SICE), Sept. 2017, pp. 1378-1383, doi: 10.23919/SICE.2017.8105718.

[73] X. Ping, F. Yang, H. Zhang, J. Zhang, W. Zhang, and G. Song, "Introducing machine learning and hybrid algorithm for prediction and optimization of multistage centrifugal pump in an ORC system," *Energy*, vol. 222, p. 120007, May. 2021, doi: 10.1016/j.energy.2021.120007.

[74] W. Li, X. Jiang, Q. Pang, L. Zhou, and W. Wang, "Numerical simulation and performance analysis of a four-stage centrifugal pump," *Advances in Mechanical Engineering*, vol. 8, no. 10, p. 1687814016673756, 2016, doi: 10.1177/1687814016673756.

[75] M. A. El-Naggar, "A one-dimensional flow analysis for the prediction of centrifugal pump performance characteristics," *International Journal of Rotating Machinery*, pp. 1-19, 2013, doi: 10.1155/2013/473512.

[76] M. E. Matlakala and D. V. Von Kallon, "Investigating the efficiency of centrifugal pumps via FEA," *Proceedings of the International Conference on Industrial Engineering & Operations Management*, pp. 4535-4548, Jul. 2021.

[77] J. Jacobs, M. Mathews, and M. Kleingeld, "Failure prediction of mine de-watering pumps," *Journal of Failure Analysis and Prevention*, vol. 18, pp. 927–938, 2018.

[78] B. Yuxing, K. Fanyu, Z. Fei, W. Jiaqiong, X. Bin, and H. Qianlan, "Experiments and numerical simulation of performances and internal flow for high-speed rescue pump with variable speeds," *Mathematical Problems in Engineering*, article vol. 2018, 01/01/ 2018, doi: 10.1155/2018/9168694.

[79] W. Conradie, "Reconfiguring mine water reticulation systems for cost savings," M.Eng dissertation, North-West University, 2018.

[80] J. Chen, Y. Wang, Q. Jiang, and J. Tang, "Study of data analysis model based on Big Data technology," *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, pp. 1-6, 2016.

[81] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526-534, Aug. 2021, doi: 10.1016/j.procs.2021.01.199.

[82] A. Azevedo and M. Santos, "KDD, SEMMA and CRISP-DM: a parallel overview," in *IADIS European Conf. Data Mining*, 2008.

[83] M. H. U. Rehman, V. I. C. Chang, A. Batool, and T. Y. Wah, "Big data reduction framework for value creation in sustainable enterprises," *Int. J. Inf. Manag.*, vol. 36, pp. 917-928, 2016.

[84] K. L. Sainani, "Dealing with missing data," *PM&R*, vol. 7, no. 9, pp. 990-994, Sept. 2015, doi: 10.1016/j.pmrj.2015.07.011.

[85] W. H. Domonique, "Robust statistical methods for handling missing data," vol. 78.

[86] Z. Ma and G. Chen, "Bayesian methods for dealing with missing data problems," *Journal of the Korean Statistical Society*, Review Article vol. 47, no. 3, pp. 297-313, Sept. 2018, doi: 10.1016/j.jkss.2018.03.002.

[87] J. Hox, "A review of current software for handling missing data," vol. 62, pp. 123-138.

[88] C. Pichery, *Sensitivity Analysis*, 3rd ed. (Encyclopedia of Toxicology). Oxford: Academic Press, 2014, pp. 236-237.

[89] T. Lenhart, K. Eckhardt, N. Fohrer, and H. G. Frede, "Comparison of two different approaches of sensitivity analysis," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 27, no. 9, pp. 645-654, Jan. 2002, doi: 10.1016/S1474-7065(02)00049-9.

[90] A. Saltelli, "Sensitivity analysis for importance assessment," *Risk Analysis: An International Journal*, vol. 22, no. 3, pp. 579-590, Jun. 2002.

[91] H. Christopher Frey and S. R. Patil, "Identification and review of sensitivity analysis methods," *Risk Analysis: An International Journal*, vol. 22, no. 3, pp. 553-578, 2002.

[92] H. Frey, A. Mokhtari, and T. Danish, "Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models " Dept. Civ. Cons. and Env Eng., North Carolina State University, Raleigh, SC, Jan. 2003.

[93] J. Little, "Missing data assumptions," *Annual Review of Statistics and Its Application*, vol. 8, no. 1, pp. 89-107, 2021/03/07 2021, doi: 10.1146/annurev-statistics-040720-031104.

[94] N. Mittag, "Imputations: benefits, risks and a method for missing data."

[95] H. M. Safhi, B. Frikh, and B. Ouhbi, "Assessing reliability of Big Data Knowledge Discovery process," *Procedia Computer Science*, vol. 148, pp. 30-36, Jan. 2019, doi: 10.1016/j.procs.2019.01.005.

[96] I. Ben-Gal, "Outlier Detection," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach Eds. Boston, MA: Springer US, 2005, pp. 131-146.

[97] H. Wang, M. J. Bah, and M. Hammad, "Progress in outlier detection techniques: a survey," *IEEE Access*, vol. 7, pp. 107964-108000, Aug. 2019, doi: 10.1109/ACCESS.2019.2932769.

[98] H. Liu, S. Shah, and W. Jiang, "On-line outlier detection and data cleaning," *Computers & Chemical Engineering*, vol. 28, no. 9, pp. 1635-1647, Aug. 2004, doi: 10.1016/j.compchemeng.2004.01.009.

[99] H. Tsukamoto and H. Ohashi, "Transient characteristics of a centrifugal pump during starting period," (in English), *Journal of Fluids Engineering, Transactions of the ASME*, Article vol. 104, no. 1, pp. 6-13, 1982, doi: 10.1115/1.3240859.

[100] Y. Zhang, Z. Zhu, Y. Jin, B. Cui, Y. Li, and H. Dou, "Experimental study on a centrifugal pump with an open impeller during startup period," *Journal of Thermal Science*, Original Paper vol. 22, no. 1, pp. 1-6, Jan. 2013, doi: 10.1007/s11630-013-0584-3.

[101] I. Chalghoum, S. Elaoud, M. Akrout, and E. H. Taieb, "Transient behavior of a centrifugal pump during starting period," *Applied Acoustics*, vol. 109, pp. 82-89, Aug. 2016, doi: 10.1016/j.apacoust.2016.02.007.

[102] H. Yousefi, M. H. Yeganeh, N. Alinaghipour, and A. Movaghar, "Structure-free real-time data aggregation in wireless sensor networks," *Computer Communications*, vol. 35, no. 9, pp. 1132-1140, May 2012, doi: 10.1016/j.comcom.2011.11.007.

[103] T. Du, Z. Qu, Q. Guo, and S. Qu, "A high efficient and real time data aggregation scheme for wsns," *International journal of distributed sensor networks*, vol. 11, no. 6, p. 261381, 2015.

[104] E. J. de Fortuny, D. Martens, and F. Provost, "Predictive modeling with Big Data: is bigger really better?," *Big Data*, vol. 1, no. 4, pp. 215-226, 2013, doi: 10.1089/big.2013.0037.

[105] J. G. Pretorius, M. J. Mathews, P. Maré, M. Kleingeld, and J. van Rensburg, "Implementing a DIKW model on a deep mine cooling system," *International Journal of Mining Science and Technology*, vol. 29, no. 2, pp. 319-326, Mar. 2019, doi: 10.1016/j.ijmst.2018.07.004.

[106] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher, "What do we talk about when we talk about dashboards?," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 682-692, 2019, doi: 10.1109/TVCG.2018.2864903.

[107] D. P. Tech. "Products." https://donnleepumptech.co.za (accessed 15 Feb. 2022).

APPENDIX A

1. Sensitivity analysis

In Section 3.3.3, it was discovered that the suction pressures for 100L pumps at Mine A were much higher than the base suction pressure that was used to support the sensitivity analysis (SA). This section provides the results obtained performing the same SA but using a suction pressure of 500 kPa (in the range similar to those recorded for 100L pumps). The rest of the base case values remained the same as in Table 4.

Table 14: Base case values with 500kPa suction pressure

Suction pressure [kPa]	Discharge pressure [kPa]	Discharge flow [l/s]	Power [kW]	Efficiency [%] (calculated using Eq. 1)
500	7000	130	1120	75

Table 14 shows the base case values with a new suction pressure as well as updated efficiency. The results for each step change are presented in Figure 45.

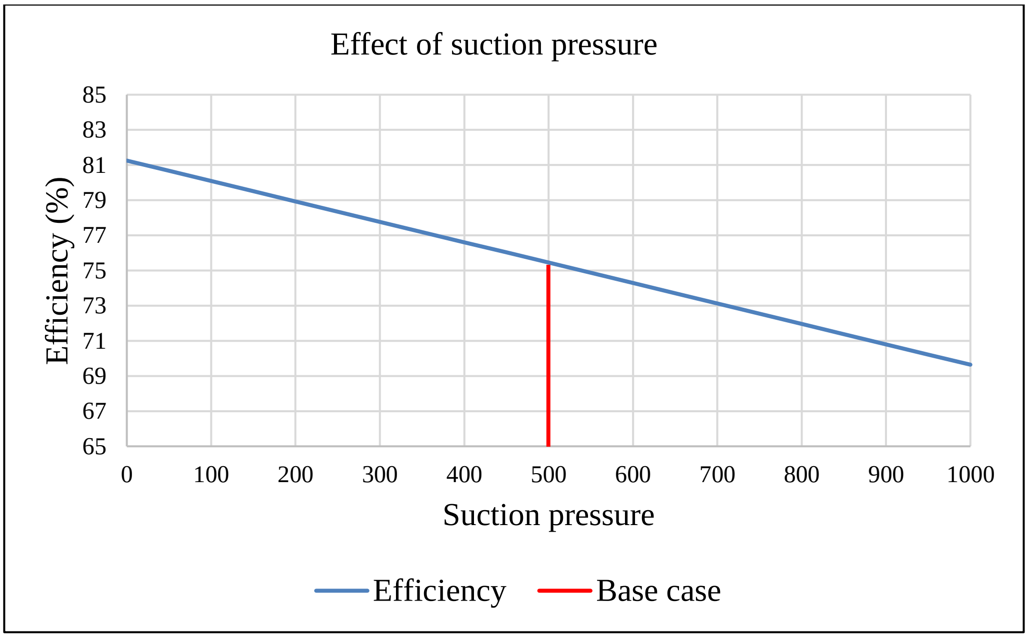


Figure 45: Sensitivity analysis with suction pressure of 500 kPa

From Figure 45, it can be observed that the differences in efficiencies are within 6% of the base efficiency. This difference is still insignificant, and thus the results of the sensitivity analysis still hold.

2. Raw data

Table 15 shows a snapshot of the raw SCADA data produced from 29L pump P1 at Mine A.

Table 15: Example of how each parameter varies with time

Time	Discharge pressure (kPa)	Suction pressure (kPa)	Flow (l/s)	Power consumption (kW)
00:00:00	7106.48	57.87	140.29	1236.72
00:02:00	7103.59	57.58	140.13	1235.48
00:04:00	7106.48	57.00	139.97	1228.07
00:06:00	7106.48	56.71	141.41	1240.64
00:08:00	7109.38	56.71	141.09	1238.24
00:10:00	7094.91	56.42	137.07	1201.72
00:12:00	7100.69	56.42	137.07	1200.18
00:14:00	7100.69	56.13	136.75	1208.12
00:16:00	7103.59	55.56	136.10	1202.44
00:18:00	7103.59	55.27	136.59	1198.69
00:20:00	7100.69	54.98	136.75	1201.79
00:22:00	7103.59	54.69	137.87	1219.79
00:24:00	7100.69	54.11	137.87	1219.87
00:26:00	7103.59	53.82	139.80	1237.80
00:28:00	7103.59	53.53	140.13	1240.51
00:30:00	7109.38	53.24	142.22	1249.48
00:32:00	7100.69	53.24	140.29	1230.16
00:34:00	7103.59	53.24	140.13	1233.96
00:36:00	7103.59	52.66	138.52	1220.92
00:38:00	7094.91	52.37	137.71	1214.97
00:40:00	7094.91	52.37	136.59	1211.13
00:42:00	7100.69	52.37	135.78	1201.22
00:44:00	7094.91	52.37	136.91	1202.79
00:46:00	7094.91	52.08	138.20	1215.97
00:48:00	7089.12	52.08	136.91	1204.31
00:50:00	7100.69	52.08	138.20	1216.18
00:52:00	7094.91	52.08	135.94	1202.51
00:54:00	7092.01	52.08	137.39	1199.80
00:56:00	7092.01	52.08	138.36	1206.26
00:58:00	7092.01	51.79	137.39	1203.44
01:00:00	7094.91	51.79	137.87	1210.25

From Table 15, it can be observed that each parameter's raw data varies with time, with power consumption having the most significant variance among the rest. The rest of the pumps at both Mine A and Mine B were also found to exhibit a similar trend.

3. Energy and financial figures

In addition to efficiencies, the reports discussed in Section 3.5 include a financial section. The section consists of total monthly running hours vs energy consumption per pump as well an estimated pumping energy cost. These figures are quantified according to the three different TOU periods.

Mine A

Table 16 presents the results obtained for the total running hours and power consumption at Mine A in December 2021.

Table 16: Total running hours vs energy consumption per pump for each TOU period

Pump	Peak TOU		Standard TOU		Off-peak TOU	
	Hours	MWh	Hours	MWh	Hours	MWh
29L P1	59	71	183	222	242	293
29L P2	61	62	183	186	198	200
29L P4	52	61	187	222	207	245
29L P5	50	62	179	223	213	265
52L P1	54	201	170	627	211	782
52L P2	61	166	205	554	268	724
52L P3	54	75	158	221	144	202
52L P4	62	208	196	656	222	742
75L P3	82	162	214	425	274	544
75L P4	46	131	156	450	183	481
75L P5	45	127	151	430	155	441
75L P6	66	109	190	313	239	392
100L P1	35	100	112	319	118	335
100L P3	12	34	53	153	84	243
100L P4	55	155	136	383	116	327
100L P5	39	104	104	279	162	437
Total	**833**	**1825**	**2577**	**5662**	**3036**	**6654**

As can be observed in Table 16, the running hours are evenly distributed amongst all pumps on the same level. Furthermore, the running hours imply that there is no load shifting practice on pumping. This is because the total peak running hours for each pumping station were just above 200 and there are approximately 200 peak hours in a 31-day calendar month. This suggests that, on average, at least one pump was operated during every peak time instance.

The energy consumption was quantified using the following equation:

$$\boldsymbol{Consumption_{TOUx}(kWh) = Average\ power(kWh) * Total\ hours_{TOUx}} \qquad \text{Eq. 9}$$

As can be observed in Table 16, there is less consumption during peak periods, but that is simply because there are only 5 peak hours in a normal day, compared to eleven and eight standard and off-peak periods, respectively.

To quantify the total electricity cost, the electricity tariffs were used as shown in Eq. 10 below.

$$\boldsymbol{Electricity\ cost(R) = Power\ consumption(kWh) * Tariff(R/kWh)} \qquad \text{Eq. 10}$$

Using the respective tariffs and the consumption provided in Table 16, the following results are obtained:

$$\boldsymbol{Peak\ cost(R) = 1825 * 1000 * 4.1942 \approx R\,7.7million}$$

$$\boldsymbol{Standard\ cost(R) = 5662 * 1000 * 1.2761 \approx R\,7.2million}$$

$$\boldsymbol{Offpeak\ cost(R) = 6654 * 1000 * 0.6968 \approx R\,4.6million}$$

This leads to a total pumping electricity cost of approximately R20 million per month. It is worth noting that this cost does not include the 115L pumps, as they do not form part of the report for reasons mentioned in Section 3.3.1. As expected, the highest electricity cost was incurred during peak TOU. This highlights the need for intervention either in the form of load shifting and/or efficiency-based pump prioritisation. However, that is not within the scope of this study, and is thus recommended for future research.

Mine B

Table 17 presents the results obtained for the total running hours and power consumptions at Mine B in December 2021.

Table 17: Total running hours vs energy consumption per pump for each TOU period

Pump	Peak		Standard		Off-peak	
	Hours	MWh	Hours	MWh	Hours	MWh
76L P1	4	3	48	34	72	51
76L P2	4	4	67	56	75	62
76L P3	7	5	86	61	90	64
61L P1	2	2	27	34	45	57
61L P2	8	16	98	212	105	226
3000L P1	2	4	34	73	64	137
Total	**27**	**34**	**360**	**469**	**451**	**597**

The peak running hours shown in Table 17 indicate that load shifting is practiced at Mine B. For example, the total number of peak operating hours on 76L were 20 out of the available 200 peak hours in the month. This section of the report is for the engineering manager at the mine to verify that the pump operators effectively practice load shifting.